Implementation and Efficacy Study of Raz-Plus for Students from Kindergarten to Fifth Grade

Authors:

Hsiang-Yeh Ho, Ph.D.

Mathias Smrekar, M.A.

Publication Date: March 25, 2020

Published By:

McREL International

4601 DTC Boulevard, Suite 500

Denver, CO 80237-2596

Recommended citation: Ho, H., Smrekar, M. (2018). *Implementation and efficacy study of Raz-Plus for students from kindergarten to fifth grade*. Denver, CO: McREL International.

This study evaluates the efficacy of a commercial product, Raz-Plus, developed by Learning A–Z. The authors of this paper are employed by McREL International, a private 501(c)(3) nonprofit corporation specializing in research and evaluation services, which was contracted by Learning A–Z to design and carry out the study. None of the researchers receive commission on sales of the products.

Abstract

The current study examined the effect of a web-based literacy program, Raz-Plus, that serves as a resource for blended learning instruction in K–5 classrooms. Following a 13-week implementation period at three rural elementary schools, comparisons of the treatment (21 teachers and 249 students) and control (18 teachers and 198 students) groups revealed greater outcomes in the treatment group for overall literacy achievement, student interest in academic reading, and student interest in recreational reading, as compared to the control group. Core dimensions of implementation fidelity were measured through classroom observations and surveys with participating teachers; analyses revealed no significant associations between fidelity measures and student outcomes. This study provides evidence of the efficacy of Raz-Plus as a tool for supporting literacy development of elementary students. Additional research is warranted to determine the elements of implementation that support optimal student outcomes.

Key words

Cluster randomized controlled trial, educational technology program, literacy achievement, elementary

Implementation and Efficacy Study of Raz-Plus for Students from Kindergarten to Fifth Grade

Literacy development in elementary school is a key predictor of long-term academic and professional success (Fiester, 2013). Students who do not meet standards for reading proficiency in third grade fail to graduate from high school at a rate four times higher than that of proficient third-grade readers (Hernandez, 2011). The long-term implications of early-life reading are wide-ranging; age-appropriate literacy development in elementary school sets the foundation for continued academic success and a positive life trajectory (Kern & Friedman, 2008). However, according to the National Assessment of Educational Progress (NAEP), just 37% of fourth-grade students in the United States scored proficient on reading skills in 2017; average achievement scores on the NAEP have not improved since 2007 (U.S. Department of Education, 2017).

Educators are continuously searching for effective methods to facilitate students' development as readers. One approach that has gained considerable traction in recent years is the integration of technology into the classroom (Horn & Fisher, 2017; Powell et al., 2015). This increase in technology integration can be attributed to widespread internet availability, improvements in the capabilities and affordability of technology, and legislative efforts to prepare students for 21st century careers (Horn & Staker 2011; U.S. Department of Education 2015; Wilkes et al., 2016). One increasingly popular method of technology integration in K–12 classrooms is blended learning initiatives that combine traditional teacher-led instruction (face-to-face) with technology-enhanced learning activities (online component) that are often completed individually by students (Horn & Fisher, 2017; Horn & Staker, 2011). The teacher-led component (face-to-face instruction) and student individual work (online) component are connected to provide an integrated learning experience. Additionally, the increasing popularity

of blended learning in K–12 classrooms stems from a larger pedagogical trend toward individualized learning in modern education (Horn & Fisher, 2017). Through individual online work, students learn on their own time and pace; hence, blended learning creates an individualized learning pathway for each individual student—a form of individualized differentiation not easily facilitated by teachers during whole-class instruction (Horn & Fisher, 2017; Horn & Staker, 2012; Powell et al., 2015; Prescott et al., 2017).

Blended learning literacy programs often incorporate a variety of technology features that facilitate effective instruction, such as easy access to materials and resources that are aligned with learning standards and student learning, adaptive content that is scaffolded to students' individual needs based on user performance (EdSurge, 2016; Kazakoff et al., 2018), built-in formative assessments to monitor student progress (EdSurge, 2016; Horn & Staker, 2012), and real-time data reports to help teachers provide differentiated support to students (Freeland, 2015; Horn & Staker, 2011; Powell et al., 2015). Many add gaming and incentivization features to drive student interest, engagement, and motivation (Area et al., 2015; Evans et al., 2014). These features are beneficial from a pedagogical standpoint; the literature is rife with studies linking reading interest and engagement with literacy achievement (Baroody & Diamond, 2017; Guthrie et al., 2007; Kush et al., 2005; Parsons et al., 2015). Studies of elementary students have demonstrated a link between interest in reading and reading comprehension (Guthrie et al., 2007) and have shown that student reading attitudes in elementary school are predictive of reading achievement in later grades (Kush et al., 2005).

To date, research on blended learning has largely centered around high school and post-secondary educational contexts (Kazakoff et al., 2018; Prescott et al., 2017; Shechter et al., 2017). While the literature generally supports the benefits of technology integration with older

students, there is a little evidence to support its effectiveness in elementary classrooms (Kazakoff et al., 2018; Means et al., 2013; Prescott et al., 2017). Adding to this uncertainty, the literature largely lacks information on the ways teachers are utilizing blended learning literacy programs, and how teachers' methods of implementation affect student outcomes (Schechter et al., 2017). The present study aims to fill these gaps by examining both the efficacy of an elementary blended learning literacy program and the ways in which teachers are utilizing it to support students' literacy development.

Literature Review

While blended learning literacy programs are increasingly incorporated into English Language Arts (ELA) instruction as a means of facilitating students' literacy development (Schechter et al., 2017), there is little research into the efficacy of these programs on student reading outcomes in general education classrooms (Kazakoff et al., 2018; Means et al., 2013; Prescott et al., 2017). The few studies that have examined the effects of such programs in elementary settings are encouraging. Prescott and colleagues (2017) studied 722 students ages kindergarten to fifth grade in a Title 1 urban school who participated in a blended learning literacy program. Students demonstrated significant improvement from the beginning to the end of the school year on a standardized reading assessment, and student progress on the online component of the blended learning program was significantly associated with student gains on the reading assessment. Wilkes and colleagues (2016) conducted a randomized control trial to examine the effects of a blended learning approach to literacy instruction with a sample of 74 second graders across three classrooms in a Title I school. Following a 16-week implementation period, students receiving the blended learning intervention demonstrated greater mean gains on a measure of reading proficiency than students who received a traditional ELA curriculum.

While these studies are encouraging, given their use of small samples from single school sites, their rigor and generalizability are limited. Further research that can provide stronger evidence of program efficacy, such as randomized control trials involving multiple implementation sites, are warranted (What Works Clearinghouse, 2013).

The literature is likewise lacking investigation into the fidelity of implementation for such programs and the ways in which teachers' implementation mediates literacy outcomes for students (Schweighofer & Ebner, 2015). The importance of assessing implementation fidelity of an intervention has been widely documented in the program evaluation literature (Carroll et al., 2007; Century et al., 2010; Fixsen et al., 2005; O'Donnell, 2008; Zvoch, 2008). There is consensus in implementation science that variation in implementation of a program under investigation (i.e., lack of fidelity) poses a threat to both internal and external validity of the research study (O'Donnell, 2008). Without a full understanding of the extent to which the program is implemented as intended—which is the definition of fidelity (Dane & Schneider, 1998)—researchers are unable to separate a design failure from an implementation failure when null findings are observed (Century et al., 2010; Zvoch, 2009).

Beginning with Dane and Schneider's (1998) seminal paper on implementation fidelity, the literature includes varied frameworks for conceptualizing and measuring program implementation fidelity (Century et al., 2010; Dane & Schneider, 1998). Core dimensions of implementation fidelity include adherence (the extent to which program components are delivered as intended), exposure (frequency of program delivery), dosage (amount of program delivery), quality (qualitative measures of implementors' delivery), and participant responsiveness (extent of participant participation and engagement), among others (Century et al., 2010; Dane & Schneider, 1998). Given the increasing prevalence of blended learning literacy

programs in elementary settings, additional research is warranted to help determine the extent to which these programs are beneficial for elementary students' literacy development, how teachers implement them, and how implementation affects student outcomes.

Program Used in Current Study

The current study examined the benefits of a web-based literacy program, Raz-Plus, developed by Learning A–Z, that serves as a resource for blended learning instruction in K–5 classrooms. Raz-Plus comprises a collection of online literacy resources for teachers and individualized content for students at varying levels of literacy development. Raz-Plus incorporates several key evidence-based features that support teacher instruction and student learning within the context of blended learning.

First, it provides an extensive collection of virtual reading resources and materials, including leveled books, that are developmentally appropriate and standards-aligned, for teachers to incorporate into their daily instruction. The reading resources and materials are accessible in both online (projectable, eBooks, mobile) and offline (printable) formats, providing teachers flexibility in selecting the materials and formats that best fit their instructional needs. In addition to providing resources for teacher-led instruction, Raz-Plus offers online learning activities that correspond with the teacher-led instructional resources and incorporate gaming and incentivization features that engage students in learning. As the resources are designed to meet the needs of students at varying levels of literacy development and are available for both online and in-person use, teachers can use Raz-Plus in whole-group, small-group, and individual work formats, depending on student and classroom needs. Therefore, by design, Raz-Plus can serve as a resource for teachers implementing blended learning models in common instructional formats (e.g., station rotation, lab rotation, individual rotation).

Second, Raz-Plus is designed to serve students at every reading level by accelerating targeted reading skills for struggling readers and increasing progress for on-level and advanced readers. Specifically, Raz-Plus helps teachers provide individualized instruction that meets the learning needs of all students by allowing teachers to assign leveled books and/or corresponding level-based virtual activities to groups of students or individual students depending on reading levels and learning needs. Teachers can also assign online quizzes to students to collect real-time data on student progress and use it to provide individualized supports.

Lastly, students have access to all Raz-Plus reading materials in class and at home. Students can choose what, when, where, and at what pace they read outside of school. Many of the materials incorporate gaming features to keep students motivated to develop their literacy skills. Through the provision of literacy resources on a wide array of topics, students are able to select reading materials of interest to read for pleasure. Individual completion of the online resources provides students with a degree of control over their learning experience and allows them to progress through individualized content at a pace that is comfortable.

Current Study and Research Question

As of 2018, there were 64,000 active Raz-Plus accounts being used by elementary teachers in 28,773 schools throughout the United States (Learning A–Z, Tucson, AZ). To date, there are no systematic experiments published of Raz-Plus outcomes. Several small-scale studies have assessed the effects of specific components within Raz-Plus, particularly of Raz-Kids, the online platform through which students access the literacy resources and engage in learning activities. Findings are promising: Students using Raz-Kids have demonstrated increases in reading motivation and confidence (Gülşen & Mede, 2016), reading speed and accuracy (Hendrickson, 2014), attitudes toward reading (Carroll, 2013), and a variety of measures of

comprehension and reading proficiency (Carroll, 2013; Hampton et al., 2015; Hendrickson, 2014; Marchland, 2015; Resendez & Azin, 2014). However, these studies only investigated the effects of the electronic books and quizzes that students experience online, without the teacher-delivered components that take place during small- or whole-group instruction, and their findings are somewhat limited by small samples consisting primarily of special populations of students (e.g., students with learning disabilities or English language learners). Additional research is needed to examine the efficacy of Raz-Plus as an integrated program consisting of teacher-led instruction as well as independent student practice.

The primary goal of the current study was to advance the literature on Raz-Plus by investigating its effects on student literacy outcomes in K–5 general education settings. A secondary goal was to gather information on how teachers utilize the program as a supplemental resource to support their existing ELA curriculum, and how teachers' implementation affects student outcomes. Three research questions were addressed:

1. To what extent did Raz-Plus affect the students' literacy outcomes?
2. To what extent did teachers implement the program with fidelity?
3. To what extent was the fidelity of implementation associated with student reading outcomes?

Methods

Study design

The study was a two-level multisite randomized control trial with 39 K–5 teachers from three rural elementary schools in two school districts in the southeastern United States. No participating teachers had prior experience with Raz-Plus. Participating teachers were randomly assigned to treatment or control groups within school and grade-level strata. This design was

chosen to ensure equal representation of teachers across grade levels and intervention sites in both treatment and control conditions. The treatment and control conditions included 21 and 18 teachers, respectively. The study occurred during the spring semester of the 2017–18 school year, during which a total of 662 students were enrolled in the participating classrooms. Prior to the study, parents were given the opportunity to opt their child out of the study. As a result, 10 students (five in each condition) were removed from the study. These students received access to Raz-Plus in the same way as their peers, but their data were excluded from all analyses. All students from the participating schools received free or reduced-price lunch (an indicator of poverty status), and the schools primarily served students from racial/ethnic minority backgrounds (80–96%), predominantly African-American. The population of English learners within these schools is small, ranging from 0% to 0.8%. The population of students in special education ranged from 15% to 18%. Student enrollment across the participating schools ranged from 246 to 401.

Treatment teachers, on average, were slightly older, had more education, and had more years of teaching experience than did control teachers. However, the difference was not statistically significant. Specifically, slightly more than half of the treatment teachers (52%) and slightly less than half of the control teachers (46%) were 45 years and older. About 45% and 21% of the treatment and control teachers, respectively, had a master's degree. About 29% and 30% of the treatment and control teachers, respectively, had 0–1 years of teaching experience. Fifty-five percent and 43% of the treatment and control teachers, respectively, had more than 11 years of teaching experience.

Procedure

The 21 teachers assigned to the treatment condition used Raz-Plus as a supplemental resource in ELA instruction during the 13-week spring semester of the 2017–18 school year. To support use and implementation of the program, two professional development (PD) sessions were delivered to the treatment teachers. The first, delivered onsite prior to the study's commencement, provided an overview of Raz-Plus; the developer's recommended curriculum plan to support the integration and adoption of Raz-Plus in the existing curriculum; the key functions and features of the Raz-Plus online portals for teachers and students; and expectations for program implementation. The second session was delivered in-person or via webinar, per teachers' availability, during weeks 5 and 6 of implementation. This PD focused on data-driven instruction using the Raz-Plus portal, and teachers had opportunities to ask questions about implementation. Additionally, the treatment teachers received unlimited access to pre-recorded online PD sessions and technical support from Learning A–Z, a service provided to all Raz-Plus subscribers.

During the implementation period, treatment teachers had access to all Raz-Plus resources, materials, and reports. Students had access to the program in and outside of the class. Teachers had the autonomy to determine how to incorporate the program into their instruction and how their students utilized the program, with only one explicit expectation: Teachers were instructed to incorporate Raz-Plus into their literacy instruction and/or student activities in class for at least 150 minutes weekly. As a support to help teachers adopt and integrate the program in the current curriculum, Learning A–Z developed optional curriculum plans available to all teachers in the treatment group. These grade-based curriculum plans included standards-based lessons centered around a weekly key objective (e.g., *Author's Point of View*) and recommended

Raz-Plus resources to support each day's lesson. Specifically, from Mondays to Thursdays, teachers would provide instruction using Raz-Plus materials in whole-group, small-group, or individual work settings. On Fridays, students would access lesson-related text for independent practice through their online portal. Teachers were encouraged to integrate the recommended curriculum plan whenever possible.

The 18 teachers in the control condition were instructed to conduct "business-as-usual" literacy instruction with the curriculum they had been using since the beginning of the school year. The only difference between the treatment and control groups was that treatment teachers integrated Raz-Plus into the existing curriculum, but control teachers did not. Therefore, students taught by the treatment teachers used Raz-Plus online, but students taught by the control teachers did not. After the study, all control-group teachers were offered the program and the same professional development their colleagues in the treatment group received.

Measures

Student Literacy Outcomes

The current study examined two specific literacy outcomes—student interest in reading and overall literacy achievement—as they are key predictors of reading achievement in later grades (Kush et al., 2005). The methods of data collection and measures of literacy outcomes are detailed below.

Student interest in reading. The Elementary Reading Attitude Survey (ERAS; McKenna and Kear, 1990; 1999) was administered online with students from grades 3–5 to assess student interest in reading. ERAS was administered twice: once before the implementation period and once after. The ERAS consists of 20 statements that assess two reading attitude subscales: academic reading (10 items) and recreational reading (10 items). Each item is displayed with

four cartoon faces expressing various emotions on a four-point scale from *very happy* to *very upset*. Students were instructed to select the emotion that best reflected their own feelings about each statement. The higher the number, the more positive the attitude. Subscale mean scores were used in the analyses. Alpha reliability of student interest in the recreational reading subscale was 0.84 and 0.87 at pretest and posttest, respectively. Alpha reliability of student interest in the educational reading subscale was 0.85 and 0.89 at pretest and posttest, respectively.

Student reading assessment. The Renaissance STAR Early Literacy assessment (SEL) for grades K–2 (Renaissance Learning, 2016a) and the STAR Reading assessment (SR) for grades 3–5 (Renaissance Learning, 2016b) were used to assess student reading achievement at baseline and at the end of the program implementation. SEL and SR are norm- and criterion-referenced assessments that are computer-adaptive, adjusting automatically to reflect the skill level of a student. SEL is intended for students who are in the early reading stages of literacy development, designed explicitly for students in grades K–2, focusing on foundational skills of word knowledge in seven literacy domains: alphabetic principal, concept of word, visual discrimination, phonemic awareness, phonics, structural analysis, and vocabulary. Split-half reliabilities are high, with reported coefficient alphas ranging from 0.75–0.83 (Renaissance Learning, 2016a). SR focuses on factors that predict comprehension in five domains: word knowledge and skills, comprehension strategies and constructing meaning, analyzing literary text, understanding author's craft, and analyzing argument and evaluating text. Split-half reliabilities for this assessment are also high, with reported coefficient alphas of 0.89 for grades 3–5 (Renaissance Learning, 2016b). Both SEL and SR were revised in 2012 with additions to the test bank and alignment with U.S. Common Core State Standards (CCSS). The tests show high

correlations (concurrent validity) with other widely used literacy assessments, including several statewide reading assessments (Renaissance Learning, 2016a; 2016b). The National Center on Response to Intervention rates the STAR assessments highly as tools for monitoring progress (American Institutes for Research 2009a) and screening (American Institutes for Research 2009b).

In the present study, SR and SEL were administered twice: once before program implementation and once after. The assessments required students to respond to 25 multiple-choice questions of varying difficulty. SEL and SR assessments are particularly well-suited for short-term interventions because they can be administered multiple times within a brief timeframe; the assessment system keeps track of the questions presented to each student across test sessions and will not ask the same question more than once in any 90-day period.

SEL and SR assessment outputs provide standard scaled scores. Additionally, to allow for vertical scaling across grade levels, the assessment outputs also provide unified scores (vertically scaled across two assessments and across grade levels) that allow researchers to monitor student progress over time and across grade levels. Hence, unified scores were used for analyses in the current study. Additionally, according to the SEL technical manual, students from kindergarten to second grade are recommended to take the SEL. However, students in these grades receiving a scaled score of 775 or higher are encouraged to take the SR test to ensure adequate alignment between students' performance levels and the assessment questions, and to better measure students' progress over time. For the current study, this procedure was followed at both pretest and posttest: Students in grades K–2 who scored 775 or higher on the SEL were assessed a second time using the SR.

SEL and SR assessment outputs also provide percentile rank benchmarks. Students at or below the 10th percentile are in need of urgent literacy intervention (severely at-risk); students who are between the 11th and 25th percentiles are candidates for literacy intervention (at-risk); students who are between the 26th and 40th percentiles are "on watch"; and students who are above the 40th percentile are considered to be at grade level or above for literacy development (Renaissance, 2016a; 2016b). For the impact models used in analyses, student risk status at pretest was used as one of the key covariates.

Implementation Data

Fidelity measures. For this study, four aspects of implementation fidelity were assessed (Dane & Schneider, 1998; Century et al., 2010): exposure (frequency of program use), dosage (amount of program delivery), quality (qualitative measures of implementors' delivery), and participant responsiveness (extent of participant participation and engagement). It should be noted that adherence measure was not assessed in the current study given the nature of the implementation design: Teachers were given the autonomy to decide what to use and how to integrate the program in their daily instruction. Methods used to assess exposure, dosage, quality, and participant responsiveness are described in the next section.

Exposure and dosage measures. Participating teachers completed surveys monthly during the implementation period. Individual items on the teacher survey served as measures of program exposure (*How often did you use Raz-Plus materials for instruction and/or student activities?*) and program dosage (*Approximately how many minutes per week did you use Raz-Plus materials for instruction and/or student activities?).* Response options for the exposure item were: *never*, *occasionally* (*1–2 times per week*), *frequently* (*3–5 times per week*), and *daily*. Response options for the dosage item were: *less than 30 minutes*, *30–60 minutes*, *60–90 minutes*, *90–120 minutes*,

120–150 minutes, and *more than 150 minutes*. Scores for exposure and dosage were averaged across the three survey administrations and were used in the analyses.

Quality measure. Implementation quality was assessed via classroom observations (Century et al., 2010). A classroom observation protocol was developed by the researchers to evaluate the quality of teachers' instruction when using Raz-Plus materials. The observation focused on three dimensions of teacher instruction (Carlisle et al., 2011; Kelcey & Carlisle, 2013): organization (pedagogical structure; 4 items), delivering literacy content (directing knowledge and skill acquisition; 4 items), and supporting student learning (i.e., fostering student engagement and self-understanding; 5 items). Each observation item was rated on a four-point scale: *absent*, *unsatisfactory quality*, *somewhat satisfactory*, and *satisfactory quality*. Two researchers were trained to conduct the observations. They conducted the initial three classroom observations together to assess inter-rate reliability; the average exact and adjacent inter-rater reliability scores were 88% and 100%, respectively. After establishing this high inter-rater reliability, the observers conducted the remaining observations individually. The observers were able to schedule and conduct observations with 17 of the 21 treatment teachers (81%). All 17 teachers were observed twice, except for one teacher whose schedule only permitted one observation. Observations occurred for 30 minutes during the classroom's normal literacy block. All observations were conducted during weeks 10 and 11 of the implementation period. In the current study, alpha reliability was 0.69, 0.93, and 0.77 for organizing, delivering literacy content, and supporting student learning, respectively. A composite score was calculated by averaging scores from the three dimensions of instruction and used in the analyses.

Participant responsiveness. Whereas quality measures teachers' instructional practice, participant responsiveness assesses student participation during instruction (Century et al., 2010).

In the present study, participant responsiveness data were collected through classroom observations as described earlier. Five items guided by the literature on student motivation and engagement were developed to measure students' level of engagement and learning during the lesson (Archer & Hughes, 2011; Lane & Harris, 2015). Each item was measured on a 4-point scale: *absent*, *unsatisfactory quality*, s*omewhat satisfactory quality*, and s*atisfactory quality*. An example item is "*Student questions and comments often determined the focus and direction of classroom discourse.*" Alpha reliability for the scale was 0.52. A scale mean was computed and used in the analyses.

Program usage data. Program usage data were collected through teacher interviews, teacher surveys, classroom observations, and Raz-Plus program data. Findings from these data sources were reported in the Results section.

Teacher interview. All treatment teachers were invited to participate in a 30-minute interview over the phone or in person depending on teacher availability. Sixteen of the 21 treatment teachers (76%) were interviewed. Through interviews, teachers shared their experiences with Raz-Plus implementation, suggestions for additional resources to support implementation, and their perceptions of outcomes for participating students. Thematic analysis was conducted to identify the themes that emerged from the interview data.

Teacher survey. The teacher survey directed respondents to indicate how often they used the teacher portal to assign reading resources to students, review student performance and reports, and use student performance data to assign appropriate reading resources. Teachers were also asked to report how often they used specific program materials, including suggested curriculum, ancillaries (e.g., worksheets, vocabulary cards, discussion cards, word work activities, assessments), projectable books, and printable books. These survey items were rated

on a four-point scale: *never*, *occasionally (1–2 times per week*), *frequently* (*3–5 times per week*), and *daily*.

Classroom observations. As part of the classroom observations, researchers took notes to describe teachers' use of the recommended curriculum plan. Based on the field notes, researchers then gave a score to describe the extent to which teachers adhere to the recommended curriculum plan using the following response categories: (1) *Non-compliance: Did not use Raz-Plus during the observation,* (2) *Some compliance: Used Raz-Plus resources, but the lesson was not aligned with the recommended curriculum plan for the day,* (3) *High Compliance: Used some of the suggested Raz-Plus resources for the day,* and (4) *Absolute Compliance: Used all of the suggested Raz-Plus resources for the day*. Observers also documented the types of resources used during the observation and types of instructional grouping strategies when Raz-Plus was in use during the observation.

Raz-Plus program data. The Raz-Plus program collects usage data from teachers and students every time teachers and students interact with the program. The types of data collected include number of teacher and student logins; number of materials and resources downloaded by teachers and students; number of custom assignments assigned by teachers; and number of listens, reads, and quizzes completed by students. These data were descriptively analyzed and discussed in the Results section.

Covariates

Covariates included in all analytic models were student gender, minority status, at-risk status at baseline (determined by STAR percentile rank benchmarks), grade level (five dummy variables were created using kindergarten as the reference group), and intervention site (two dummy variables were created to account for between-school variations).

Data analysis

Attrition analyses were conducted to determine whether attrition rates differed by intervention status. Guidelines provided by What Works Clearinghouse (2013) were used to determine attrition levels and appropriately address missing data. A comparison of baseline characteristics of the treatment and control groups (student demographics and student pretest scores) were examined to ascertain baseline equivalence.

To address research question 1, multilevel modeling analyses, accounting for the nested structure of the data (student nested within classrooms), were conducted to examine program impact on student literacy outcomes. In preliminary analysis, researchers found that student at-risk status was the strongest predictor of student posttest outcome. When centering methods were employed in the model (group mean-centered student level at-risk status and grand mean-centered classroom level at-risk status) it increased the statistical power and explained additional variances on posttest measures by parsing out the effect of at-risk status within and between classrooms (Bryk & Raudenbush, 1992; Hofmann & Gavin, 1998). Hence, the centering methods, as described, were employed in the final impact analysis models. Effect size (ES) is an estimate that represents the magnitude of the program impact on student outcomes. Hedge's *g* was calculated using impact estimate divided by the pooled student and classroom level variance components (What Works Clearinghouse 2013).

To answer the second research question—*To what extent did teachers implement the program with fidelity?*—descriptive analyses were conducted on four dimensions of fidelity: exposure, dosage, quality, and participant responsiveness. Additional program usage data collected through teacher surveys, teacher interviews, and classroom observations were analyzed to provide contextual understanding of how teachers utilized the program in the classroom.

To address the third research question—*To what extent was the fidelity of implementation associated with student reading outcomes?*—multilevel modeling analyses were employed to account for the nested structure of the data. All covariates used in the impact analysis models were also included in the analyses.

All descriptive analyses were conducted using IBM SPSS Statistics for Windows, version 25.0. All multilevel modeling analyses were conducted using Hierarchical Linear and Nonlinear Modeling (HLM) software for Windows, version 7.

Results

Samples used for analyses varied by analysis type. Details regarding study samples and findings from the analyses are reported in the following sections.

Raz-Plus impact on student literacy achievement

Sample. Attrition occurred at both classroom level and student level. At the classroom level, one control teacher resigned during the implementation period, and one treatment teacher was unable to administer the STAR assessment with their class at posttest. Both classrooms were removed from analyses, resulting in an overall attrition rate of 5%, with 5% and 6% attrition rates in the treatment and control conditions, respectively. At the student level, 585 students took STAR assessments at pretest (n_t = 313, n_c = 272). Of those, 464 students took STAR assessments at posttest (n_t = 253, n_c = 211). The overall attrition rate at the student level is 21%, with 19% and 22% attrition in the treatment and control conditions, respectively. According to WWC standards, both classroom-level and student-level attrition rates are low (What Works Clearinghouse, 2013). However, to avoid a ceiling effect, teachers in grades K–2 were instructed to administer the SR with any student scoring 775 and higher on the SEL. Appropriate follow-up administration did not occur with all students. Students that scored 775 or higher on the SEL but

were not assessed using SR were treated as missing. This approach resulted in an overall attrition rate of 21% (565 students took the appropriate STAR assessment at pretest, and 447 of these students took the appropriate STAR assessment again at posttest), with 18% among the treatment group (54 out of 303 students did not take the appropriate STAR assessment either at pretest or posttest) and 24% among the control group (64 out of 262 students did not take the appropriate STAR assessment either at pretest or posttest). Still, according to WWC standards, these attrition rates are considered low under the conservative boundary (What Works Clearinghouse, 2013).

Attrition analysis further confirms there were no statistical differences between treatment and control students in terms of gender, minority status, at-risk status, and literacy skills at baseline. Attrition would therefore not be expected to affect results. Complete case analysis with no regression adjustment was performed to analyze the impact of Raz-Plus on student literacy achievement as measured by STAR. Table 1 displays the baseline characteristics for the analysis sample. Overall, treatment and control students who were included in the analysis were similar at baseline in terms of gender, minority status, at-risk status, and literacy achievement.

Table 1. Student Baseline Characteristics: Sample for Impact Analysis on Literacy Achievement

	Whole sample (n = 447)	Treatment (n = 249)	Control (n = 198)	p-value
Male (%)	51%	52%	49%	0.295
Minority students (%)	88%	88%	88%	0.502
At-risk status (%)	51%	50%	52%	0.356
STAR unified score (mean, SD)	868.34 (116.30)	867.59 (111.85)	869.29 (121.93)	0.878

Primary analysis findings. Table 2 displays the results from the multilevel model predicting student literacy achievement. Student minority status, at-risk status, intervention site, and grade level significantly predicted student literacy achievement at posttest. After accounting for covariates, Raz-Plus participation significantly predicted student literacy achievement at posttest (p = 0.044). This provides support for the efficacy of Raz-Plus on student literacy outcome.

Table 2. Multilevel Model Predicting Student Literacy Achievement

	Estimate	Standard error
Fixed effects		
Intercept	735.40***	14.17
School A	-1.99	8.55
School B	19.15**	6.92
1st grade	68.95***	13.13
2nd grade	160.10***	13.57
3rd grade	188.97***	13.78
4th grade	230.91***	13.49
5th grade	243.07***	17.86
Gender [a]	-0.41	5.75
Minority status [b]	-19.06**	6.49
Individual at-risk status (group mean-centered) [c]	-68.17***	6.02
Classroom level at-risk status (grand mean-centered)	-119.68***	18.42
Group [d]	14.85*	6.40
Variance Components		

	Estimate	Standard error
Level 1 residual	3299.32	
Level 2 residual	135.89	

** p <0.05. ** p < 0.01. *** p < 0.001.*

[a] *Gender was a binary variable: male (1) versus female (0).*

[b] *Minority status was a binary variable: minority (1) versus non-minority (0).*

[c] *Individual at-risk status was a binary variable: at-risk (1) versus not at-risk (0).*

[d] *Group was a binary variable: treatment group (1) versus control group (0).*

Raz-Plus impact on student interest in reading

Sample. Students from third to fifth grade (n = 301, n_t = 150, n_c = 151) completed the ERAS online before and after the intervention. The response rate was 77% (n = 232, n_t = 107, n_c = 125) and 66% (n = 199, n_t = 100, n_c = 99) at pretest and posttest, respectively. Altogether, 53% of students (n = 159, n_t = 78, n_c = 81) completed both pretest and posttest surveys. The overall attrition rate was 47%, with 48% and 46% in treatment condition and control condition, respectively. Taking overall and differential attrition rates into account, this is considered low attrition (What Works Clearinghouse, 2013). Attrition analysis revealed that students who were excluded from the analysis due to missing data were similar to the students retained in the analysis in terms of gender, minority status, at-risk status, interest in academic reading, and interest in recreational reading at baseline. Complete case analysis with no regression adjustment was performed to analyze the impact of Raz-Plus on student interest in reading as measured by ERAS. Table 3 shows the baseline characteristics for the analysis sample. Overall, treatment and control students included in the analysis were similar at baseline, except for student interest in

recreational reading. To account for this baseline difference, pretest data were included in the analytical model.

Table 3. Student Baseline Characteristics: Sample for the Impact Analysis on Interest in Reading

	Whole sample (n = 159)	Treatment (n = 78)	Control (n = 81)	p-value
Male (%)	50%	47%	53%	0.633
Minority students (%)	89%	89%	89%	1.000
At-risk status (%)	54%	49%	58%	0.329
STAR unified score (mean, SD)	957.42 (67.26)	965.31 (72.59)	949.64 (61.04)	0.154
Interest in academic reading (mean, SD)	2.74 (0.60)	2.78 (0.53)	2.71 (0.65)	0.476
Interest in recreational reading (mean, SD)	2.69 (0.59)	2.79 (0.52)	2.60 (0.64)	0.042

Primary analysis finding. Table 4 presents the results from multilevel models predicting student interest in academic reading and recreational reading. For both models, student interest in reading measured at pretest was the strongest predictor of student interest in reading at posttest ($p < 0.001$). After accounting for covariates, student participation in Raz-Plus is a significant predictor of student interest in academic reading ($p = 0.038$) and student interest in recreational reading ($p = 0.048$). These findings suggest that Raz-Plus increased student interest in both academic reading and recreational reading.

Table 4. Multilevel Model Predicting Student Interest in Reading

	Academic reading		Recreational reading	
	β	SE	β	SE
Fixed effects				
Intercept	2.33***	0.11	2.30***	0.20
School B [a]	0.44**	0.14	0.33	0.16
4th grade	0.09	0.17	0.11	0.17
5th grade	0.13	0.15	0.08	0.19
Gender [b]	0.03	0.07	-0.05	0.10
Minority status [c]	0.04	0.09	-0.05	0.15
Individual at-risk status (group mean-centered) [d]	0.02	0.07	0.13	0.10
Interest in academic reading (pretest)	0.54***	0.09	--	--
Interest in recreational reading (pretest)	--	--	0.56***	0.08
Classroom level at-risk status (grand mean-centered) [e]	-0.06	0.39	-0.20	0.54
Group	0.43**	0.11	0.37*	0.16
Variance Components				
Level 1 residual	0.33		0.307	
Level 2 residual	0.05*		0.02	

** p <0.05. ** p < 0.01. *** p < 0.001.*

[a] *After attrition, the study sample only contained classrooms from Site A and Site C.*

[b] *Gender was a binary variable: male (1) versus female (0).*

[c] *Minority status was a binary variable: minority (1) versus non-minority (0).*

[d] *Individual at-risk status was a binary variable: at-risk (1) versus not at-risk (0).*

[e] *Group was a binary variable: treatment group (1) versus control group (0).*

Table 5 displays the effect sizes (ES) for the outcomes of interest. The effect size for student literacy achievement was 0.14 (p = 0.415), which is considered small (Cohen, 1988; Wolf, 1986). However, given the brevity of the intervention period (13 weeks), we argue that the magnitude of difference between the treatment and control students cannot be ignored. The most promising findings from the study are that Raz-Plus participation significantly increased students' interest in reading, and effect sizes were large for both academic reading (ES = 0.63, p = 0.027) and recreational reading (ES = 0.57, p = 0.045). It may take more time to change student achievement outcomes, yet we expect that these high effect sizes in interest in reading outcomes are precursors to larger changes in achievement with sustained use of the program over a longer period of time.

Table 5. Effect Size Estimations

	Impact estimate	Student level SD		ES [a]	p [b]
		Treatment	Control		
STAR posttest	14.85	98.29	113.97	0.14	0.415
Interest in academic reading	0.43	0.59	0.75	0.63	0.027
Interest in recreational reading	0.37	0.57	0.71	0.57	0.045

[a] ES was calculated using Hedge's g formula (What Works Clearinghouse, 2013).

[b] p value (clustering-corrected statistical significance) is calculated based on Hedges (2005) recommended by What Works Clearinghouse (2013).

Analysis of implementation fidelity and program usage

Sample. All treatment teachers (*n* = 21) provided data on exposure and dosage through teacher survey. Data on quality and participant responsiveness were collected via classroom observation with 17 teachers.

Implementation fidelity. Table 7 displays the mean and score range of four fidelity measures. Overall, 14% (3 out of 21) of treatment teachers complied with the implementation usage requirement of at least 150 minutes weekly. Most used the program frequently (3–5 times weekly). On average, teachers used the program approximately 60–90 minutes weekly. In terms of quality and participant responsiveness, on average, teachers' instructional practices were somewhat satisfactory, and students' engagement during the Raz-Plus instruction was also somewhat satisfactory. Frequencies of each measure's response categories are reported in Table 6. Notably, implementation varied across teachers on all four measures of fidelity.

Table 6. Descriptive Statistics of Fidelity Measures

	n	Min	Max	M	SD
Exposure	**21**	**1.00**	**4.00**	**2.71**	**0.78**
Never (M <=1.49)	1 (4.8%)				
Occasionally (1-2 times per week) (M = 1.50-2.49)	6 (28.6%)				
Frequently (3-5 times per week) (M = 2.50-3.49)	11 (52.4%)				
Daily (M >= 3.50)	3 (14.3%)				
Dosage (per week)	**20**	**1.00**	**4.75**	**3.18**	**1.11**
Less than 30 minutes (M <= 1.49)	1 (5.0%)				
30-60 minutes (M = 1.50-2.49)	4 (20.0%)				
60-90 minutes (M = 2.50-3.49)	4 (20.0%)				

	n	Min	Max	M	SD
90-120 minutes (M = 3.50-4.49)	8 (40.0%)				
120-150 minutes (M = 4.50-5.49)	3 (15.0%)				
More than 150 minutes (M >= 5.59)	0 (0.0%)				
Quality	**17**	**1.13**	**3.81**	**2.66**	**0.84**
Absent (M <= 1.49)	2 (11.8%)				
Unsatisfactory (M = 1.50-2.49)	4 (23.5%)				
Somewhat satisfactory (M = 2.50-3.49)	9 (52.9%)				
Satisfactory (M >= 3.50)	2 (11.8%)				
Participant Responsiveness	**17**	**1.67**	**3.83**	**2.92**	**0.65**
Absent (M <= 1.49)	0 (0.0%)				
Unsatisfactory (M = 1.50-2.49)	5 (29.4%)				
Somewhat satisfactory (M = 2.50-3.49)	8 (47.1%)				
Satisfactory (M >= 3.50)	4 (23.5%)				

Program usage data. In interviews and surveys, teachers overall reported a positive experience with Raz-Plus content and materials; most agreed that the program is of high quality (M = 4.11, SD = 0.67), useful (M = 4.24, SD = 0.78), and relevant (M = 4.21, SD = 0.76). When asked about their experience with the program, teachers reported challenges incorporating the recommended curriculum plan and using the program materials in instruction and student activities at least 150 minutes per week. Most challenges stemmed from the timing of the intervention (beginning half-way through the school year), need for more training to understand and implement the recommended curriculum plan, and other academic initiatives requiring

teachers' attention. Despite these implementation challenges, teachers reported observing improvements in literacy development, engagement, and interest in reading among students.

Classroom observation data further supports teacher interview data in terms of the challenges in integrating the recommended curriculum. Overall, 15% of the observed sessions did not utilize any Raz-Plus resources or materials during the observations (non-compliance), 82% somewhat complied with the recommended curriculum (some compliance or high compliance), and 3% absolutely complied with the recommended curriculum (absolute compliance). In most cases, the implementation timeline was either behind or ahead of the recommended plan (e.g., using suggested resources from Week 10 during Week 11).

In surveys, teachers reported the types of resources they utilized and how they used them to support instruction and student learning. On average, teachers *occasionally (1–2 times per week)* used the suggested curriculum (M = 2.04, SD = 0.89), projectable books (M = 1.90, SD = 0.93), print books (M = 1.80, SD = 0.76), and ancillaries (M = 2.11, SD = 0.87). They also *occasionally (1–2 times per week)* used the online portal to assign reading resources to students (M = 2.38, SD = 0.79), reviewed student performance and reports (M = 2.27, SD = 0.81), and used student data to assign appropriate resources (M = 2.10, SD = 0.92). Classroom observations further provided information regarding the instructional context when these resources were used. Specifically, eBooks were the most frequently used Raz-Plus resource during the observations, followed by projectable books, ancillaries, and printable books. Students read eBooks during small group time or individual work time. When printable books were used, they were used mostly in whole group or small group settings, and sometimes used in paired and individual activities. When projectable books were used, they were mostly used in whole group activities, while on some occasions, they were used in individual work time. When ancillaries were used,

they were mostly used in small group and individual settings. Taken together, during ELA block time, teachers often implemented different grouping strategies and utilized different types of Raz-Plus materials to facilitate student learning.

Teachers also reported on how frequently they used the teacher portal to assign reading resources, review student performance and reports, and use student data to assign appropriate resources to students. These practices are essential to providing differentiated instruction for individual students. Results showed that teachers, on average, *occasionally* (1–2 times per week) conducted these practices to support student learning.

Raz-Plus program data provides an overall snapshot of the extent of program use on both the teacher side and the student side during the 13-week implementation period (see Table 7). On average, teachers logged into their account 4–5 times a week and used 7–8 resources per week, but they rarely used the custom assignment function. Students, on average, logged into their account 2–3 times a week, downloaded 11–12 resources per week, completed 1–2 lessons and reads each week, and took 2–3 quizzes per week. Variation in program use is evident based on the range of the usage data (Min, Max) presented in Table 7.

Table 7. Raz-Plus Program Data During the 13-Week Implementation Period

Activities	n	Min	Max	M	SD
Teacher Usage During the 13-Week Implementation Period					
# Teacher Logins	21	4	177	59.90	51.56
# Materials and Resources Downloads by Teachers	21	0	601	95.48	138.53
# Custom Assignments	21	0	21	1.62	4.72
Student Usage During the 13-Week Implementation Period					

Activities	n	Min	Max	M	SD
# Student Logins	315	1	251	29.83	39.46
# Materials and Resources Downloads by Students	313	1	773	149.14	145.27
# Listens	312	0	224	24.84	37.01
# Reads	312	0	184	24.69	26.31
# Quizzes	312	0	244	38.13	45.41

Correlational analysis between fidelity measures and student outcomes

The sample used for the correlational analyses of implementation fidelity and student outcomes included 185 students from whom literacy skill data was collected at pretest and posttest, from 15 classrooms where complete fidelity data was collected. Multilevel model analyses were conducted to examine the associations between fidelity measures and student literacy outcome. Covariates, including student gender, minority status, at-risk status, grade level, and intervention site, were included in the model. Student at-risk status and grade level were significant predictors of student literacy achievement ($p < 0.05$). After accounting for covariates, none of the fidelity measures were significant predictors of student literacy outcomes.

Discussion

The primary goal of the present study was to examine the efficacy of a blended learning literacy program on student literacy development in K–5 classrooms. Controlling for baseline skills and demographics, and accounting for nesting within classrooms, results indicated that the program increased students' overall literacy achievement, interest in academic reading, and interest in recreational reading. These findings contribute to the literature by providing empirical

evidence of the benefits of a blended learning literacy program for elementary students in general education settings.

Although the effect size estimates revealed a small magnitude of impact on students' overall reading skills (ES = 0.14), it was quite large on students' interest in academic reading and interest in recreational reading (ES = 0.57-0.63). The small effect size for overall literacy achievement should be viewed in light of the brevity of the intervention (13 weeks) and the wide variation in level of program use and implementation. That is, based on the implementation data, the average usage of 3–5 times and 60–90 minutes per week was below the dosage recommended by the researchers. Additionally, implementation quality, participant responsiveness, and program usage by teachers varied between classrooms; and program use by students varied within and between classrooms. Most importantly, teachers did not take full advantage of what the program has to offer—teachers seemed to underuse the features that support differentiated learning, such as reviewing student data and using data to inform instructional decision. The program effect might have been larger if teachers had used the program with greater fidelity (Schechter et al., 2017). The current study aimed to address this consideration by examining correlations between fidelity measures and student literacy achievement. However, measuring fidelity with a program designed to provide flexibility and freedom on the users' end is challenging. None of the four dimensions of fidelity assessed—exposure, dosage, quality, and participant responsiveness—was statistically associated with student literacy development.

One potential explanation for the lack of statistically significant correlations is lack of power; missing data and using half of the sample (treatment group only) resulted in small samples to analyze. Another possible explanation is that teachers with more struggling readers in

their classroom used the program more often. Or, conversely, these teachers may struggle to successfully implement the program with a high number of students requiring individualized support, and who are less likely to remain engaged during the instruction. Exploratory analyses were conducted to examine the correlations between classroom percentages of struggling readers at pretest, and four fidelity measures. Results supported these speculations. After accounting for grade level and intervention site, the classroom percentage of struggling readers at baseline was marginally related to exposure (frequency of program use; $\beta = 1.98$, $SE = 1.06$, $p = 0.086$) and significantly related to dosage (amount of program use; $\beta = 4.39$, $SE = 1.39$, $p = 0.009$). Conversely, the percentage of struggling readers within the classroom showed a statistically significant negative association with quality of instructional delivery ($\beta = -2.57$, $SE = 0.91$, $p = 0.023$) and a marginal negative association with participant responsiveness ($\beta = -2.18$, $SE = 0.99$, $p = 0.06$). Regardless, after accounting for baseline covariates, including student at-risk status and classroom percentage of struggling readers, fidelity measures were not significant predictors of student literacy achievement.

Perhaps the most important question becomes: What does optimal implementation look like for a blended learning literacy program that intentionally lacks strict implementation guidelines because it emphasizes flexibility for users? Additional research is warranted to develop an implementation model that supports optimal student outcomes, taking into account both the program's flexibility and its lack of standardization.

The strengths of the current study include a rigorous large-scale randomized control trial design across multiple implementation sites, measurement of multiple student literacy outcomes, and a rich collection of qualitative and quantitative implementation data that contribute to the interpretation of results (Horner et al., 2006; O'Donnell, 2008; Zvoch, 2009). Yet, there are

several notable limitations. First, the primary goal of the study was to examine program efficacy within a routine condition; therefore, teachers got standard implementation training from the developer. Teachers said in interviews that more training would have been helpful, something that future studies may consider providing. Second, all participating sites were rural schools that mainly served low-income minority students. Future studies may consider including urban schools with more-diverse student populations to increase the generalizability of the findings. Third, starting the study in the middle of the school year brought implementation challenges to teachers whose curriculum and classroom routine were well established, likely explaining why many failed to meet the recommended 150 minutes of weekly usage.

National data continues to show that that U.S. students' literacy achievement has not improved over the past decade (U.S. Department of Education, 2017). The current study provides evidence to support the benefits of a blended learning literacy program and its influence on student literacy achievement. Despite the aforementioned limitations, the current study is one of the very few that have applied a rigorous research design to investigate the efficacy of a blended learning literacy program in elementary general education. There is great need for evidence-based programs that support student literacy achievement in elementary school; we believe this study contributes to the educational literature in that aspect. Further study of Raz-Plus, particularly regarding implementation best practices, will benefit current Raz-Plus users, as well as educators seeking a new approach to effectively support their students' literacy achievement.

References

American Institutes for Research, National Center on Response to Intervention. (2009a). *Progress monitoring tools* (*STAR Reading*). Retrieved from http://www.intensiveintervention.org/chart/progress-monitoring

American Institutes for Research, National Center on Response to Intervention. (2009b). *Screening tools chart* (*STAR Reading*). Retrieved from http://www.rti4success.org/resources/tools-charts/screening-tools-chart

Archer, A., & Hughes, C. (2011). *Explicit instruction: Effective and efficient teaching.* New York, NY: Guilford Press.

Area, M., González, C., & Mora, C. (2015). Beyond textbooks: Educational digitals[1] texts and gamification of learning materials. In J. Rodríguez, E. Bruillard, & M. Horsley (Eds.), *Digital textbooks: What's new?* (pp. 213–230). Santiago de Compostela, Spain: IARTEM.

Baroody, A., & Diamond, K. (2012). Links among home literacy environment, literacy interest, and emergent literacy skills in preschoolers at risk for reading difficulties. *Topics in Early Childhood Special Education*, *32*(2), 78–87.

Bryk, A., & Raudenbush, S. (1992). *Hierarchical linear models: Applications and data analysis methods.* Newbury Park, CA: Sage.

Carlisle, J., Kelcey, B., Berebitsky, D., & Phelps, G. (2011). Embracing the complexity of instruction: A study of the effects of teachers' instruction on students' reading comprehension. *Scientific Studies of Reading*, *15*(5), 409–439.

[1] The published work includes this typo.

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science*, *2*(40).

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 Education*, *31*(2), 199–218.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*, 2nd ed. Hillsdale, NJ: Erlbaum.

Dane, A., & Schneider, B. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review*, *18*(1), 23–45.

De Jong, D., Grundmeyer, T., & Anderson, C. (2018). Comparative study of elementary and secondary teacher perceptions of mobile technology in classrooms. *International Journal of Mobile and Blended Learning*, *10*(1), 12–33.

EdSurge. (2016). *Decoding Adaptive*. London: Pearson. Retrieved from http://d3e7x39d4i7wbe.cloudfront.net/static_assets/PearsonDecodingAdaptiveWeb.pdf

Evans, M., Hawkins, J., & McCrary, P. (2014). Blended learning for early learners. *Principal*, *94*(2), 26–29.

Fixsen, D., Naoom, S., Blasé, K., Friedman, R., & Wallace, F. (2005). *Implementation research: A synthesis of the literature* (FMHI Publication No. 231). Tampa: University of South Florida Louis de la Parte Florida Mental Health Institute.

Fiester, L. (2013). *Early warning confirmed: A research update on third grade reading*. Baltimore, MD: Annie E. Casey Foundation.

Freeland, J. (2015). *Three false dichotomies in blended learning*. San Mateo, CA: Clayton Christensen Institute.

Gülşen, E., & Mede, E. (2016). Efficacy of multi-level extensive reading in young learners' reading motivation. *International Online Journal of Education and Teaching*, *4*(4), 290–315.

Guthrie, J., Wagner, A., Wigfield, A., Tonks, S., Humenick, N., & Littles, E. (2007). Reading motivation and reading comprehension growth in the later elementary years. *Contemporary Educational Psychology*, *32*(3), 282–313.

Hampton, D., Morrison, W., Rizza, M., & Osborn, J. (2015). A volunteer-tutoring program in reading: Examining the growth in reading achievement of elementary grade students with SLD in a tutoring-based intervention. *The Ohio Journal of Teacher Education*, *29*(2), 5–23.

Hendrickson, B. (2014). The effectiveness of electronic books in the primary classroom. *The Eagle Feather*. Retrieved from https://eaglefeather.honors.unt.edu/2014/article/309#.WraUeGa-IWc

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

Hofmann, D., & Gavin, M. (1998). Centering decisions in hierarchical linear models: Implications for research in organizations. *Journal of Management*, *24*(5), 623–641.

Horn, M., & Fisher, J. (2017). New faces of blended learning. *Educational Leadership*, *74*(6), 59–63.

Horn, M., & Staker, H. (2011). *The rise of K–12 blended learning*. The Innosight Institute. Retrieved from https://www.christenseninstitute.org/wp-content/uploads/2013/04/The-rise-of-K-12-blended-learning.pdf

Horner, S., Rew, L., & Torres, R. (2006). Enhancing intervention fidelity: A means of strengthening study impact. *Journal for Specialists in Pediatric Nursing*, *11*(2), 80–89.

Kazakoff, E., Macaruso, P., & Hook, P. (2018). Efficacy of a blended learning approach to elementary school reading instruction for students who are English Learners. *Educational Technology Research and Development*, *66*(2), 429–449.

Kelcey, B., & Carlisle, J. (2013). Learning about teachers' literacy instruction from classroom observations. *Reading Research Quarterly*, *48*(3), 30–317.

Kern, M., & Friedman, H. (2008). Early educational milestones as predictors of lifelong academic achievement, midlife adjustment, and longevity. *Journal of Applied Developmental Psychology*, *30*(4), 419–430.

Kush, J., Watkins, M., & Brookhart, S. (2005). The temporal-interactive influence of reading achievement and reading attitude. *Educational Research and Evaluation*, *11*(1), 29–44.

Lane E., & Harris, S. (2015). A new tool for measuring student behavioral engagement in large university classes. *Research and Teaching*, *44*(5), 83–91.

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

McKenna, M., & Kear, D. (1999). Garfield revisited: Unlimited extension of permission to copy the ERAS. *The Reading Teacher*, *53*, 244.

Means, B., Toyama, Y., Murphy, R., & Baki, M. (2013). The effectiveness of online and blended learning: A meta-analysis of the empirical literature, *Teachers College Record*, *115*(3), 1–47.

O'Donnell, C. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K–12 curriculum intervention research. *Review of Educational Research*, *78*(1), 33–84.

Parsons, S., Malloy, J., Parsons, A., & Burrowbridge, S. C. (2015). Students' engagement in literacy tasks. *The Reading Teacher*, *69*, 223–231.

Powell, A., Watson, J., Staley, P., Patrick, S., Horn, M., Fetzer, L., . . . Verma, S. (2015). *Blended learning: The evolution of online and face-to-face education from 2008-2015*. International Association for K–12 Online Learning. Retrieved from https://files.eric.ed.gov/fulltext/ED560788.pdf

Prescott, J., Bundschuh, K., Kazakoff, E., & Macaruso, P. (2018). Elementary school-wide implementation of a blended learning program for reading intervention. *The Journal of Educational Research*, *111*(4), 497–506.

Renaissance Learning. (2016a). *STAR Early Literacy technical manual.* Madison, WI: Renaissance Learning, Inc. Retrieved from https://resources.renlearnrp.com/UK/Manuals/SEL/SELRPUKTechnicalManual.pdf

Renaissance Learning. (2016b). *STAR Reading technical manual.* Madison, WI: Renaissance Learning, Inc. Retrieved from https://resources.renlearnrp.com/us/manuals/sr/srrptechnicalmanual.pdf

Resendez, M., & Azin, M. (2014*). Raz-Kids and Reading A–Z: A report on the 2013 summer reading field studies*. PRES Associates, Inc.

Schechter, R. Kazakoff, E., Bundschuh, K., Prescott, L., & Macaruso, P. (2017). Exploring the impact of engaged teachers on implementation fidelity and reading skill gains in a blended learning literacy program. *Reading Psychology*, *38*(6), 553–579.

Schweighofer, P., & Ebner, M. (2015). Aspects to be considered when implementing technology-enhanced learning approaches: A literature review. *Future Internet*, *7*(1), 26–49.

Staker, H., & Horn, M. B. (2012). *Classifying K–12 blended learning*. The Innosight Institute. Retrieved from https://www.christenseninstitute.org/wp-content/uploads/2013/04/Classifying-K-12-blended-learning.pdf

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. (2017). *Average reading score of fourth-grade students in 2017 not significantly different compared to 2015.* Retrieved from https://www.nationsreportcard.gov/reading_2017/#/nation/scores?grade=4

U.S. Department of Education, Office of Elementary and Secondary Education, Office of State Support. (2015). *Fundamental change: Innovation in America's schools under Race to the Top*. Retrieved from https://www2.ed.gov/programs/racetothetop/rttfinalrpt1115.pdf

What Works Clearinghouse. (2013). *Procedures and Standards Handbook Version 3.0.* Retrieved from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_v3_0_standards_handbook.pdf

Wilkes, S., Macaruso, P., Kazakoff, E., & Albert, J. (2016). Exploration of a blended learning approach to reading instruction in second grade. In *Proceedings of EdMedia 2016 – World Conference on Educational Media and Technology* (pp. 797–802). Vancouver, Canada: Association for the Advancement of Computing in Education (AACE).

Wolf, F. (1986). *Meta-analysis: Quantitative methods for research synthesis*. Newbury Park, CA: Sage.

Zvoch, K. (2009). Treatment fidelity in multisite evaluation. *American Journal of Evaluation*, *30*(1), 44–61.