Using a Sequential Multiple Assignment Randomized Trial (SMART) to Develop an Adaptive K–2 Literacy Intervention With Personalized Print Texts and App-Based Digital Activities

James S. Kim iD
Catherine A. Asher
Mary Burkhauser
Laura Mesite
Harvard University
Diana Leyva
University of Pittsburgh

This study employs a sequential multiple assignment randomized trial (SMART) design to develop an adaptive intervention with personalized print and digital content for kindergarten to Grade 2 children (n = 273). In Stage 1, we ask whether it is better for children to receive an adaptive intervention based on (a) 10 conceptually coherent texts or (b) 10 leveled texts on a range of topics. In Stage 2, we ask how best to encourage nonresponding children. Findings indicate that children who received either conceptually coherent texts or leveled texts performed similarly on reading comprehension posttests, while augmenting and intensifying follow-up with gamification of the app and text messages to parents improved comprehension outcomes for nonresponders. Descriptively, we find that only 26% (n = 71) of parents accessed the app, highlighting the need for better implementation procedures to increase take up of app-based digital activities.

Keywords: *sequential multiple assignment randomized trial, personalization, adaptive interventions, education apps, literacy, reading, early elementary*

Literacy interventions that use technology to personalize content and activities are an increasingly common evidence-based approach for improving student outcomes (Almirall, Kasari, McCaffrey, & Nahum-Shani, 2018; Connor, 2019). For example, using information on children's prior knowledge and performance to adapt print and digital content is a popular intervention strategy for improving learning (U.S. Department of Education, 2016; Walkington & Bernacki, 2018). Adaptive treatment strategies, or adaptive interventions, are tailored to meet the individual needs of students and include decision rules to guide whether, when, and how best to offer treatment components over time (Almirall, Nahum-Shani, Sherwood, & Murphy, 2014). In particular, adaptive literacy interventions closely mirror typical practice conditions in which complex bundles of treatments are frequently modified to support individual learners. In typical practice conditions, educators are likely to confront two critical questions as they implement an adaptive intervention: Which intervention components should be implemented first? And what kinds of follow-up are needed to support children who do not adequately respond to the initial intervention components?

To date, researchers have employed the standard randomized controlled trial (RCT) design to test the efficacy of adaptive interventions relative to a control group (Hedges, 2018). As a result, it is difficult to identify what works for whom and how best to support individual learners who inevitably differ in how they respond to various intervention components. This article describes the development of a novel adaptive literacy intervention that aims to improve reading comprehension among elementary grade children from low-income households during the summer months (Alexander, Entwisle, & Olson, 2007). We illustrate how a sequential multiple assignment randomized trial (SMART) design can be used to develop an adaptive literacy intervention with personalized print texts and app-based activities for kindergarten to Grade 2 children.
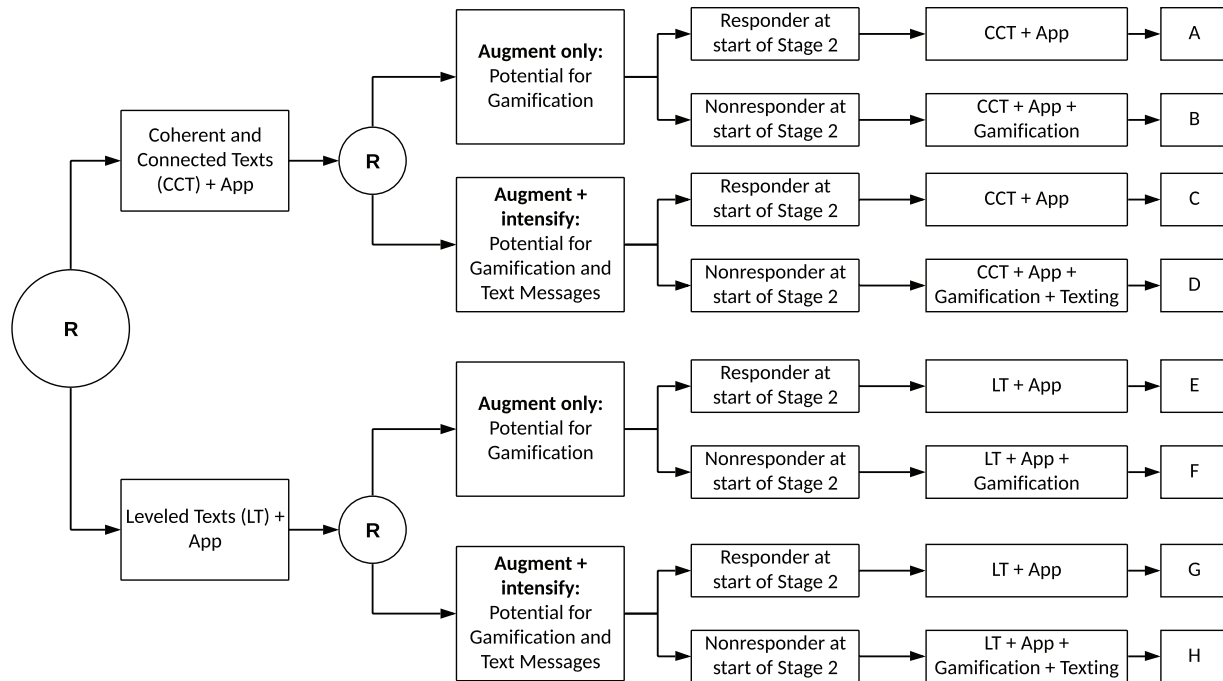
FIGURE 1.   *Design components of the SMART design.*

Matching the Method to the Complexities of an Adaptive Literacy Intervention

The SMART design is ideally suited to help researchers and practitioners develop multicomponent interventions that adapt treatments based on initial or intermediate student outcomes. A central aim of the SMART design is to identify components that best meet the needs of individual learners. Thus, participants are randomly assigned to intervention components at multiple stages of the intervention; these assignments are determined by a sequence of decision rules, often based on proximal outcomes, that determine the modality, dosage, or delivery of a specific treatment at a given point in time (Almirall et al., 2014). Typically, participants who meet some benchmark on this proximal outcome are deemed "responders" to the treatment and continue receiving the same intervention with no adaptations. Participants who do not meet the benchmark established by the decision rule are deemed "nonresponders" and subsequently receive modification to their intervention (Nahum-Shani et al., 2012). It is common for these treatment modifications to be assigned at the time of nonresponder designation; however, they can also be preassigned as a potential modification to be implemented in the case that the participant is identified as a nonresponder.

The SMART design enables researchers and practitioners to explore critical questions regarding the sequence and dosage of intervention activities in real-world contexts like schools and homes. The resulting combination of initial interventions and later modifications or supports are referred to as "embedded interventions," and head-to-head comparisons based on the multistage randomization can yield information about the most promising version to proceed with in future research studies (Nahum-Shani et al., 2012). Figure 1 illustrates how we used a SMART design with preassigned potential modifications to develop the components of our adaptive literacy intervention and to determine the different pathways of activities students can receive. In Stage 1, students are randomly assigned to intervention conditions (conceptually coherent texts [CCT] + App or leveled text [LT] + App). In Stage 2, nonresponding students in the Stage 1 intervention conditions are then randomly assigned to modifications that "augment only" (potential for gamification) or "augment and intensify" supports (potential for gamification and text messages).

Using a SMART Design to Develop Two Approaches to Context Personalization

Personalized learning in education and literacy refers to efforts by researchers and practitioners to leverage technology (e.g., formative assessments, educational apps, text messages) to adapt content and instruction to meet an individual learner's needs (Burke & Psaty, 2007; Connor & Morrison, 2016; Doss, Fahle, Loeb, & York, 2018; Greene, 2018; Hirsh-Pasek et al., 2015; Sabatine, 2018). More recently, a style of personalized learning called *context personalization* emphasizes the role of both cognitive and noncognitive factors in fostering deeper learning (Walkington, 2013; Walkington & Bernacki, 2018). The cognitive factor

refers to personalizing content and activities in ways that tap into children's prior domain knowledge and support word knowledge and text comprehension (Cervetti, Wright, & Hwang, 2016; Goldstone & Son, 2005). The noncognitive factor refers to personalizing content and activities in ways that spark interest in learning from text (Walkington & Bernacki, 2018). In this study, our primary aim was to develop and compare two approaches to context personalization in early literacy.

To operationalize the first approach—context personalization with an emphasis on *cognitive factors*—we grounded the intervention's literacy content in conceptually coherent texts (CCT) with the goal of building children's domain knowledge. Cervetti et al. (2016) define CCT as texts that convey information about the natural or social world (Duke, 2000), cohere around a single topic, and help support the development of children's domain knowledge. Theoretically, CCTs are designed to help children leverage domain knowledge in science and social studies and to connect new learning from text to a preexisting schema, or knowledge structure (Ausubel, 1968; Bransford & Schwartz, 1999; Kintsch, 2009; Perfetti & Stafura, 2014). Although CCTs are not "leveled" or matched to a child's independent reading level, there is growing experimental evidence that children show greater gains in background knowledge and text comprehension during reading activities with CCTs than with leveled texts (LT) (L. T. Brown, Mohr, Wilcox, & Barrett, 2018; Cervetti et al., 2016). CCTs play a critical role in content area literacy instruction in which teachers integrate informational texts into science and social studies lessons. In this study, the CCT approach to context personalization included pre-training in-school lessons in which teachers supported the development of children's domain knowledge (Solis, Miciak, Vaughn, & Fletcher, 2014; Trainin, Hayden, Wilson, & Erickson, 2016) and children and parents were given print texts and an app to promote at-home summer reading activities.

To operationalize the second approach—context personalization with an emphasis on *noncognitive factors*—we grounded the intervention's literacy content in LT on a range of topics. This second approach is based on the theory that reading texts that are close to a student's independent reading level—that is, within their Zone of Proximal Development (Vygotsky, 1978)—will produce the strongest gains in literacy development. LT are scored on a holistic Text Gradient Levels (Fountas & Pinnell, 2010), which includes a continuum of text characteristics that inform the level of support (and challenge) a reader will encounter in the text. There is moderate to strong evidence that providing children with LT can support wide reading of a variety of texts and ultimately improve children's fluency and text comprehension (Foorman et al., 2016; Hassrick, Raudenbush, & Rosen, 2017; Jones, Conradi, & Amendum, 2016). In this study, the

LT condition was designed to foster students' interest in reading leveled fiction and nonfiction texts, thereby increasing exposure to print and a broader range of vocabulary using texts that were neither too hard nor too easy for children (Allington, 2005; Chard, Vaughn, & Tyler, 2002; Mol & Bus, 2011; Stanovich, 2000).

We were interested in understanding whether an adaptive intervention should provide children with CCTs on a single science topic or LTs on a range of topics and how best to leverage print and digital activities to personalize learning at home during the summer months. Given that the Text Gradient System is incorporated into many formative assessments that provide regular information to teachers about each student's progress in literacy (Hassrick et al., 2017; Hoffman, 2017; York, Loeb, & Doss, 2017), comparing the CCT condition with the LT condition has both theoretical and practical implications.

Pairing Print Text and App-based Digital Activities to Enhance Context Personalization

Both print text and app-based digital activities may play a critical role in enhancing the benefits of context personalization. First, the medium matters and reading on paper may benefit learners' recall and memory of expository context (Carr, 2011; Singer & Alexander, 2017a, 2017b). Recent meta-analytic evidence indicates that students perform better on comprehension tasks when they read expository texts on paper than on screens (Clinton, 2019). Furthermore, providing choice of print text can further support intrinsic motivation to read (Deci & Ryan, 2000; Guthrie, McRae, & Klauda, 2007; Patall, Cooper, & Wynn, 2010). Second, accumulating research indicates that "educational" apps should foster meaningful learning activities and social interactions (Hirsh-Pasek et al., 2015; Mayer, 2008; Vaala, Ly, & Levine, 2015). Meaningful learning takes place when children learn with a purpose, learn material that is interesting and personally relevant, and connect new learning to prior knowledge. In other words, meaningful learning enables children to "hook" newly learned content from text into existing information and to grow their domain knowledge (Ausubel, 1968; P. C. Brown, Roediger, & McDaniel, 2014; Fitzgerald, Elmore, Kung, & Stenner, 2017). In addition to meaningful learning, apps should promote social interactions between children and parents (and other family members) to foster language development (Kremar, Grela, & Lin, 2007; Roseberry, Hirsh-Pasek, Parish-Morris, & Golinkoff, 2009; Sawyer, 2006).

In this study, children in both the CCT and LT conditions chose print versions of 10 books and personalized app-based digital activities to foster summer reading activities at home. Based on recommendations from evidence reviews for improving K–2 reading (Foorman et al., 2016; Shanahan et al., 2010), we developed app-based personalized activities

that included a variety of text-dependent questions (i.e., questions about the books) and discussion questions (e.g., questions to stimulate parent–child conversations about the book). Questions were both literal (answers could be found directly in the text) and inferential (required children to use context and prior knowledge to answer questions, for example, what is the main idea?).

To further enhance the potential benefits of pairing print texts with app-based digital content, it is critical to develop strategies for helping families access the resources that foster formal (e.g., word reading activities) and informal (e.g., shared book reading) home literacy activities (Sénéchal, 2006). Thus, the CCT and LT conditions included an after-school family literacy event that focused on helping parents learn how to access the app-based digital activities.

Developing Two Strategies for Addressing Differential Responsiveness

A novel feature of our intervention was the development of a literacy app called MORE@Home that included personalized digital activities. Given prior research indicating that many students would likely not complete the app-based personalized literacy activities paired with each of the 10 print texts (Augustine et al., 2016; Kim et al., 2016), we developed a set of Stage 2 intervention activities that included follow-up protocols to motivate nonresponders. SMART designs enable researchers to compare the effectiveness of different strategies designed to address differential responsiveness to the first stage of an intervention. In this study, because the MORE@Home app is installed on parents' cell phones, parents must be motivated to download and use the app with their children. Self-determination theory posits that motivation derives from both intrinsic and extrinsic sources (Deci & Ryan, 2000). The version of the MORE@Home app initially provided to students relied primarily on their (or their parents) intrinsic motivation. Subsequently, two strategies were used to motivate nonresponders: (1) white hat gamification features, which was directed primarily at students, and (2) text messaging, which was directed primarily at their parents. In Figure 1, we refer to white gamification as supports that "augment only" and text messaging as supports that "augment and intensify" follow-up with nonresponders.

All students identified as nonresponders based on their app usage at either 2 or 4 weeks following implementation of the Stage 1 components received white hat gamification of the MORE@Home app. Gamification refers to the goal of embedding elements of a game within an activity, and "white hat" gamification, is characterized by positive motivators, such as giving the player a purpose or rewarding the player for demonstrating creativity or skill mastery. This approach contrasts with "black hat" gamification, which is characterized by motivators of avoidance, such as instilling uncertainty or fear of loss in the player (Chou, 2014). When the

white hat gamification features were turned on for nonresponders in this study, the MORE@Home framework changed. Instead of seeing a list of activities to complete for each book, reading activities were shown as stepping stones on a path toward creating your own virtual zoo. Each completed set of activities meant a student could choose a new animal for their zoo. These features emphasize meaning, development, empowerment, and ownership—key values of white hat gamification (Chou, 2014)—and they rely on external factors to motivate students to use the app.

Gamification features are designed to motivate the player (in this case, the student) to move through the game (in this case, complete literacy activities in the app). However, even if students are interested in the books and activities, parents must provide them with opportunities to engage, including access to a smartphone with the MORE@Home app installed. Furthermore, some activities encouraged parents to talk with their child about the books. We anticipated that even parents who care about education might fail to download and use the app. For example, parents may have less information than teachers to personalize activities for their children, have logistical difficulties downloading the app, or simply forget to use the app. Thus, in addition to white hat gamification, we employed a second strategy for intensify supports for nonresponders: text messaging. Text messages were sent to parents of nonresponding students twice per week over six weeks, starting a month into summer vacation.

Text messages have become a popular approach to addressing differential responsiveness in part because they are both scalable and cost-effective. Recent studies have demonstrated promising effects of text messages to parents on student outcomes and specifically texts about their children that include guidance for supporting reading habits (Doss et al., 2018; Kraft & Monti-Nussbaum, 2017; York et al., 2017). The text messages used in this study were designed to accomplish three goals. First, to remove barriers to parental access, text messages reminded parents to use the MORE@Home app and included direct links to the iOS/Android apps (Thaler & Sunstein, 2008). Second, to highlight interesting features of the MORE@Home app, text messages mentioned the personalized activities, the gamification features, and the availability of parental discussion questions. Third, to encourage parents to read with their children, text messages shared tips on general reading behaviors, such as leveraging weekends as opportunities to read together (York et al., 2017). Finally, to correct potential misconceptions, text messages presented aggregate usage statistics, such as the number of downloads or the number of books accessed, as way of showing parents that other parents were finding the app to be a useful and fun tool. This final text message type draws on the behavioral theory that correcting parents' beliefs can lead to desirable behaviors (e.g., Robinson, Lee, Dearing, & Rogers, 2018).

TABLE 1
Demographic Characteristics of Participants

Characteristic	Overall Sample Mean	CCT Mean	LT Mean
Male (%)	45.1	46.7	43.2
Race			
African American (%)	67.0	70.5	63.4
Hispanic (%)	16.5	13.7	19.4
White (%)	4.4	5.0	3.7
Asian (%)	3.7	2.9	4.5
Other (%)	7.0	5.8	8.2
English language learner (%)	9.5	9.4	9.7
Have Individualized Education Plan (%)	8.4	10.8	6.0
Attendance 2017–2018 (%)	94.5	95.3	93.7*
Spring MAP RIT score			
Kindergarten	156.8	157.3	156.2
First grade	177.2	177.2	177.2
Second grade	181.5	181.1	181.8
Spring TRC reading level (rescaled)[a]			
Kindergarten	3.2	3.6	2.9
First grade	9.2	9.4	8.9
Second grade	10.4	9.6	10.9
Spring MCLASS proficient or above (%)			
Kindergarten	73.5	68.6	78.4
First grade	79.2	82.6	74.2
Second grade	65.6	61.5	68.6
Number of students	273	139	134

Note. CCT = conceptually coherent texts; LT = leveled texts; MAP = Measure of Academic Progress; TRC = Text Reading Comprehension.
*$p < .05$.

Research Goals and Questions

To improve reading comprehension outcomes for children in kindergarten to Grade 2 during the summer months (Alexander et al., 2007; Kim et al., 2016; Zvoch & Stevens, 2015), we used a SMART design to develop an adaptive literacy intervention with two stages of intervention activities. We asked two research questions[2]:

1. In Stage 1 of an adaptive literacy intervention, is it better for K–2 students' reading comprehension outcomes to receive pretraining in-school lessons, personalized print texts, and app-based digital activities based on (a) 10 conceptually coherent texts (CCT) or (b) 10 leveled texts (LT) on a range of topics?
2. Is it better for students' reading comprehension outcomes (a) to augment the Stage 1 treatment with white-hat gamification of the app alone (Gamification Only) or (b) to augment the Stage 1 treatment with gamification of the app and to intensify supports with text message reminders to parents (Gamfication + Texting) if students did not respond to the initial intervention?

Method

Participants

In the spring of 2018, we recruited 16 kindergarten, first-, and second-grade teachers from a southeastern Title-I elementary school to participate in this study. All students in those classrooms were eligible to participate ($n = 295$) and the 92% of students whose parents provided informed consent subsequently enrolled in the study ($n = 273$). Demographic characteristics of the students as well as baseline test scores on three measures of early reading skill (the Measure of Academic Progress [MAP RIT], Text Reading Comprehension [TRC], and MCLASS Composite) are presented in Table 1. There were no statistically significant differences between the CCT and LT condition on each of the three baseline spring reading tests.

Design and Procedures

The primary aim of the SMART design is to develop an adaptive intervention using the CCT or LT approach to context personalization. Thus, the SMART design in Figure 1 is characterized by sequential random assignment, with

at least one round of this assignment being determined based on a decision rule (i.e., the intermediate tailoring variable indicating whether students are responding to the Stage 1 components). Figure 2 provides a detailed overview of the Stage 1 components, the intermediate tailoring variable used to identify responders and nonresponders, and the Stage 2 components.

Stage 1: CCT Core Components. The CCT core components were designed to instantiate our first approach to context personalization which provides contextual grounding for reading activities at school and home. As shown in the left column of Figure 2, the CCT condition included the three components. In the first component, the CCT pretraining lessons implemented during the last 2 weeks of school, children participated in content literacy instruction and learned about the topic of Arctic animal survival. Children learned to use concept mapping to visualize semantically related words that appeared across conceptually coherent informational texts, and participated in integrated reading, writing, and listening activities to extend their thinking (Connor et al., 2017; Graham et al., 2017; Guthrie et al., 2004; Vitale & Romance, 2012). Concept mapping was used to illustrate the relationship among taught words, including *animal, survival, habitat, behavior, physical characteristic, adapt, advantage, endangered*, and *extinct*. In addition, pretraining in-school lessons also included an end-of-school year family literacy event where parents and adult caregivers learned about the goals of the MORE@Home app, received a unique code to download their app, and began to help their children complete the questions that were included in the app. Overall, 35% of parents (95 of 273) attended the event, including 29% of CCT parents (41 of 139).

In the second component, personalized print texts, children chose 10 print books (i.e., paper books) that cohered around the topic of animals and animal survival; students selected from 34 possible titles (see Appendix A in the online supplemental material for a list of books available in both conditions). Based on the Lexile approach to measuring text difficulty and student reading ability designed by MetaMetrics®, the informational books were quite challenging (most books ranged in Lexile level from 375 to 625) for the majority of students given the age and reading levels of the sample (student Lexile levels typically ranged from 25 to 475).

In the third component, app-based digital activities, children and their parents could access the researcher-developed MORE@Home app, which included personalized reading activities. To personalize activities, we used middle of year data from the Measure of Academic Progress (MAP), Primary Grade Reading, which yields Rasch Unit (RIT) scaled scores for children's performance on four subtests (foundational literacy, comprehension of text, vocabulary,

language and writing). Using the four subtest scores on the MAP, we created activities at three levels of difficulty (low, medium, high), and gave children an appropriately leveled activity that included a range of word-level, sentence, and text-level activities, and a discussion question to stimulate conversations with family members at home about the book (see Appendix B in the online supplemental material for example app activities).

Stage 1: LT Core Components. The right column of Figure 2 displays the LT components. Although the LT condition was composed of the same parallel components as the CCT condition, we developed the LT components using typical instructional practices in wide use across many U.S. urban districts. For example, in the context where we conducted this study, the district enacted a balanced literacy curriculum that provides each classroom teacher with LT based on a holistic measure of Text Gradient Level (Fountas & Pinnell, 2010). Teachers are required to teach a daily literacy block involving word study activities (e.g., phonics, spelling), independent reading, small group guided reading and writing, and an emphasis on Tier 2 academic vocabulary words. Finally, the school district was located in a state where all teachers are required to use a formative assessment screener (MCLASS 3D and TRC) at the beginning, middle, and end-of school year.

The primary technology used in component two of the LT condition—that is, the use of LT and formative assessment data based on a holistic measure text difficulty—is widely used in many U.S. school district contexts and thus, represents a face valid typical practice condition (Hassrick et al., 2017; Hoffman, 2017). For example, the Text Gradient Levels measure the level of difficulty for the narrative texts (i.e., fiction or nonfiction text typically written in narrative form) and informational texts (i.e., nonfiction text typically written in expository form) used in the LT condition. The levels range from A (beginning readers) to Z (advanced readers), and the gradient measure is holistic and based on 10 text characteristics (i.e., genre, text structure, content, themes and ideas, language and literacy features, sentence complexity, vocabulary, word difficulty, graphics, and print features). In essence, the materials and technology that already exist in many U.S. school districts afforded us an opportunity to develop a second approach to context personalization.

As shown in the right column of Figure 2, the LT condition included pretraining school lessons, personalized print texts, and app-based digital activities to stimulate students' interest in reading LT on a range of topics during the summer. In the first component, the LT pretraining in-school lessons, teachers instructed children how to use a word map to build their understanding of target vocabulary (i.e., synonyms, antonyms, examples, and nonexamples of the target word). Word maps are used in the balanced literacy block to
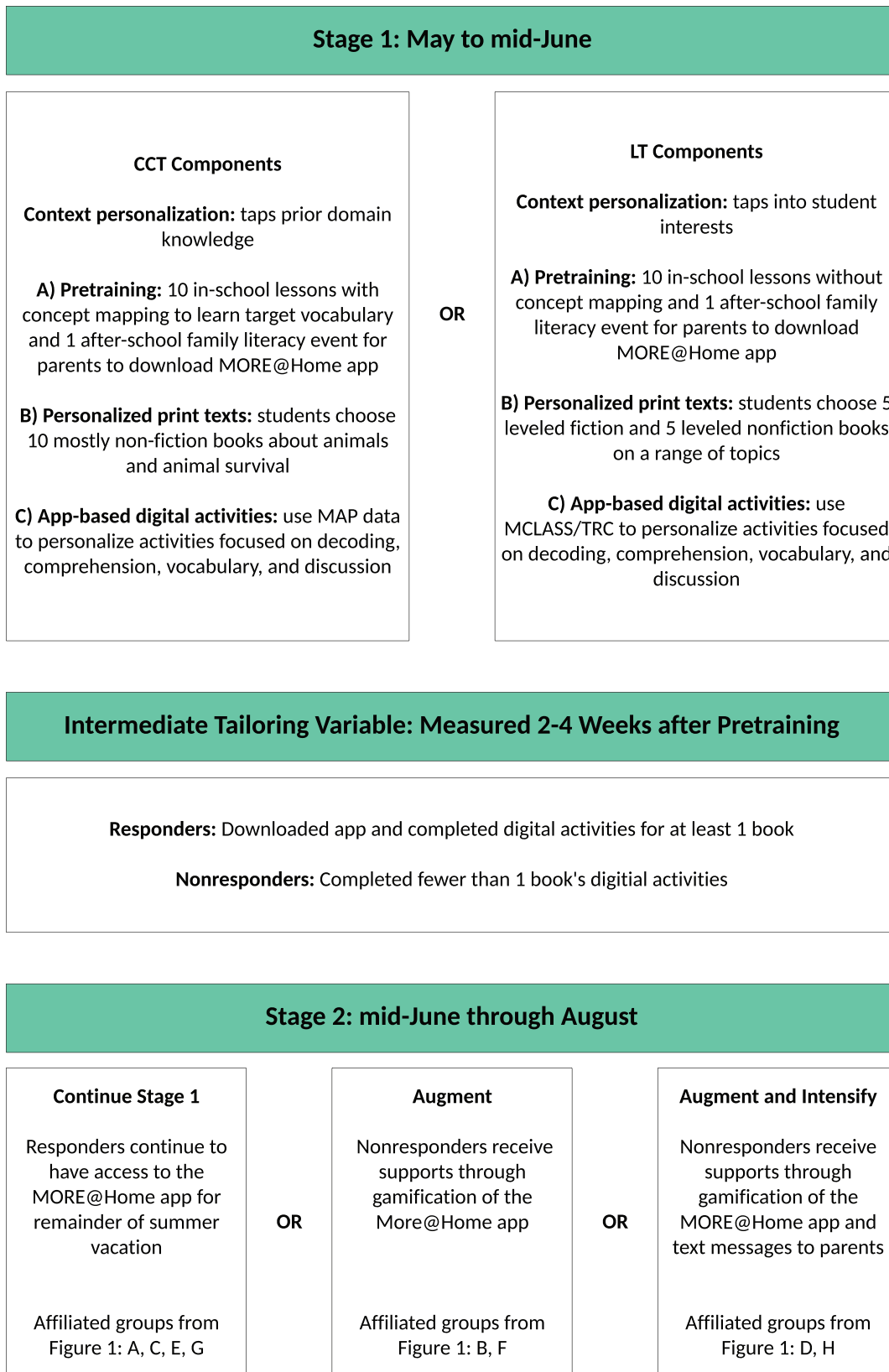
Stage 1: May to mid-June

CCT Components

Context personalization: taps prior domain knowledge

A) Pretraining: 10 in-school lessons with concept mapping to learn target vocabulary and 1 after-school family literacy event for parents to download MORE@Home app

B) Personalized print texts: students choose 10 mostly non-fiction books about animals and animal survival

C) App-based digital activities: use MAP data to personalize activities focused on decoding, comprehension, vocabulary, and discussion

OR

LT Components

Context personalization: taps into student interests

A) Pretraining: 10 in-school lessons without concept mapping and 1 after-school family literacy event for parents to download MORE@Home app

B) Personalized print texts: students choose 5 leveled fiction and 5 leveled nonfiction books on a range of topics

C) App-based digital activities: use MCLASS/TRC to personalize activities focused on decoding, comprehension, vocabulary, and discussion

Intermediate Tailoring Variable: Measured 2-4 Weeks after Pretraining

Responders: Downloaded app and completed digital activities for at least 1 book

Nonresponders: Completed fewer than 1 book's digitial activities

Stage 2: mid-June through August

Continue Stage 1

Responders continue to have access to the MORE@Home app for remainder of summer vacation

Affiliated groups from Figure 1: A, C, E, G

OR

Augment

Nonresponders receive supports through gamification of the More@Home app

Affiliated groups from Figure 1: B, F

OR

Augment and Intensify

Nonresponders receive supports through gamification of the MORE@Home app and text messages to parents

Affiliated groups from Figure 1: D, H

FIGURE 2.   *Overview of the MORE Stage 1 and Stage 2 intervention components.*

teach general purpose academic words that are useful across content areas. In addition to the in-school reading lessons, LT parents were also invited to an end-of-school family literacy event that provided information on the app to parents and adult caregiver. Overall, 40% of LT parents (54 of 134) attended the family literacy event.

In the second component, children chose personalized print texts including five informational texts and five narrative texts. We used the child's TRC level to find LT on a range of topics that had text gradient levels (D to R), which corresponded to the end of kindergarten to the end of grade 2 (see Appendix A in the online supplemental material). The third component involved app-based digital activities that were tied to each LT and focused on activities included in the TRC; the discussion question was not personalized to the book but included general active reading strategies. In sum, the LT components were designed to model typical instructional practices, to help children learn a general vocabulary strategy for independent reading of personalized and leveled print texts, and to include activities focused on text dependent comprehension and vocabulary questions.

Stage 2: Intermediate Tailoring Variable. In June, we used app usage data to determine whether students were responding or not responding to the Stage 1 intervention components. At this point, the decision rule was implemented to identify the "responder" students and the "nonresponder" students. Using a measure of print exposure as a proxy for later reading achievement (Kim et al., 2016; Mol & Bus, 2011), we defined "responders" as students who had completed all the MORE@Home app activities for at least one book. These students continued using the application with no additional supports. Students who had not yet completed all the app activities for at least one book were deemed "nonresponders" and subsequently received additional supports.

Supporting Nonresponders With Gamification Only or Gamification With Text Message Reminders. A second stage of random assignment defined the type of supports we provided these nonresponders. We blocked students within CCT and LT conditions, noted the date they assessed for nonresponder status, and then randomly assigned to receive gamification of the app only or gamification of the app plus text messaging.

Simplified SMART Design. Initially, this study leveraged three different stages of random assignment, with the middle random assignment assigning different dates for the assessment of nonresponder status (either the middle or the end of June). These two dates, 2 weeks and 4 weeks after the family literacy event marking the end of pretraining, were chosen to represent and "early" and "typical" follow-up period. The typical follow-up period was chosen to represent a period of time over which parents would have had time to settle into

their summer routines. The "early" period, 2 weeks after the end of pretraining and one week after the end of the school year, provided an earlier indicator of initial summer vacation behaviors. All randomizations occurred at the beginning of the study and the design and analysis plan were preregistered prior to all program implementation activities (Kim, 2018). However, we decided to simplify our analysis and collapse across the second wave of random assignment for two key reasons. First, while we maintained fidelity of identifying participants as nonresponders at the appropriate time point, there was little difference in the treatment experience for nonresponders in the two conditions. Text messaging only began in July after both groups had been evaluated on the tailoring variable, and while technically, gamification of the app was implemented immediately on identification as a nonresponder, almost all nonresponders had failed to access the app at all. Thus, they would not have learned of the gamification until the first text message was sent. The lack of treatment contrast between the two groups is the primary reason for collapsing across them, but this decision was justified as there was also no significant difference in the tailoring variable between the two groups. This justified the simplification of our analyses to the research design, which is presented in Figure 1.

Stratified Sample of Surveyed Parents. To provide context for our primary research questions and to further explore how families engaged with reading and the MORE@Home app over the summer, we selected a sample of parents for a phone survey in Fall 2018. We randomly identified 2 app users and 2 non–app users (as well as backups) from each of the four embedded interventions, resulting in a total of 16 parents. Both English- and Spanish-speaking parents were called and asked a series of questions about their use of the MORE@Home app over the summer and other ways they engaged with their child about books. We were able to conduct the survey for 12 of the 16 targeted parents. The overall response rate was 75%, and the response rate was higher among app users (88%) than among non–app users (63%).

Measures

Data collection took place immediately after the in-school lessons in Grades K–2, throughout the summer, and following program summer implementation in the fall of Grades 1 to 3. First, we administered three researcher-developed measures immediately after the spring lessons to assess impact on spring science vocabulary, listening comprehension, and word reading skills in Grades 1 and 2. Copies of these can be found in Appendix C in the online supplemental material. Second, we collected usage statistics from the MORE@Home app over the course of the summer, including the books accessed and the activities completed. Finally, we used administrative data to assess impact on two standardized measures of

reading in the fall of Grades 1, 2, and 3 and administered a brief parent survey over the phone.

Measure of Science Vocabulary Knowledge. This assessment is a researcher-developed semantic association task designed to measure vocabulary depth and knowledge of taught science concepts. Students were provided with a set of 10 target words, and for each, were asked to identify which of the four-word option was semantically linked to the target word. Each of the 40 word-options was scored as either incorrect (0) or correct (1). In this sample, the Science Vocabulary Knowledge test demonstrated good reliability, with a Cronbach's alpha score of .94.

Listening Comprehension. This researcher-developed measure is designed to assess whether the student can extract meaning from text that was not taught in any classroom. It includes a short passage about rainforests that was adapted from the Magic Tree House Fact Tracker series (Will Osbourne and Mary Pope Osbourne) and is accompanied by four multiple-choice questions. Students had a copy of the passage in front of them, and the teacher read it aloud. Student scores represent the percent correct among these inferential questions. The assessment had a Cronbach's alpha of .47 in this sample.

Word Study. This measure is a researcher-developed task that assessed students' decoding skills that were addressed in the word study component of the MORE intervention. Student scores represent the percent correct among the six multiple-choice questions that asked students to identify similar phonetic components in the word options as in the target word. The assessment had Cronbach's alpha of .54 in this sample.

Measure of Academic Progress. We used the MAP as the primary pretest and posttest to evaluate impacts on student reading comprehension overall and informational text comprehension in particular. The MAP is a computer-based, adaptive, and vertically scaled test that is administered three times per year and measures a variety of reading skills. It consists of an RIT score (the primary measure) as well as subscores that distinguish between specific types of reading skills. In early grades, these subscores measure both foundation and code-focused skills as well as meaning-focused skills. Once students reach second grade, the measure assesses meaning-focused skills but continues to distinguish between narrative and informational text comprehension ability. The MAP has a reported test–retest reliability ranging from .89 to .96 (R. S. Brown & Coughlin, 2007). For our Grades 1 and 2 baseline sample (also the rising Grades 2 and 3 sample), we report results for the informational text comprehension outcomes. We used RIT scores from the Spring 2018 and Fall 2018 as our primary student outcome.

MCLASS Reading 3D: Dynamic Indicators of Basic Early Literacy Skills (DIBELS) and Text Reading Comprehension. The MCLASS DIBELS (Dynamic Indicators of Basic Early Literacy Skills) battery is a series of teacher-administered assessments that provides systematic information on early grade literacy indicators. From kindergarten to Grade 3, DIBELS subtests include initial sound fluency, phoneme segmentation fluency, letter naming fluency, nonsense word fluency, oral reading fluency, and retell abilities (Good & Kaminski, 2002). We used the DIBELS total composite score since reliability of subtests range from .88 to .98 across tested grades. In addition to the DIBELS composite score, we used the TRC score, which is an individually administered assessment in which teachers use a running record to measure children's reading accuracy, fluency, comprehension, and oral and written comprehension. Validity evidence indicates that the TRC is correlated ($r = .71$) with end-of-Grade 3 reading comprehension tests (North Carolina Department of Public Instruction, 2018).

Analysis

Main Effects. SMART designs use common analytic approaches for traditional RCTs, as random assignment occurs at each phase of the design (Almirall et al., 2014). Our preferred specification is an intent-to-treat (ITT) analysis that leverages the preimplementation random assignments to each potential condition. One reason for preferring an ITT analysis is that not all parents provided a valid phone number. These students would still be able to receive gamification, but were unable to receive any text messages, rendering them as noncompliers in a typical Rubin Causal Model framework (Rubin, 1974). There is a second type of endogenous noncomplier that we include in the analysis by using the ITT approach—these are the responders, those who used the app from the beginning and according to the study design were thus not eligible to receive additional supports. Both these noncompliers were randomly assigned to their respective conditions (including potential additional supports), but due to their own endogenous choices, never received those supports. Using the ITT analytic approach provides a conservative estimate of the effects of additional supports by including individuals who never actually received the intended services.

Figure 1 provides a roadmap for our analyses. For our first research question, our goal was to examine the main effects of the Stage 1 treatments comparing the CCT and LT conditions. Thus, we compared outcomes for all students who received CCT (Groups A + B + C + D) to outcomes for all students who received LT (Groups E + F + G + H). For our second research question, our goal was to examine the main effects of the Stage 2 treatments. Here, we compared outcomes for all students who had the

TABLE 2
Summary App Usage Statistics

Usage statistic	Analytic Sample		Those Who Accessed App	
	M	*SD*	*M*	*SD*
Books accessed				
Kindergarten	1.2	2.5	3.8	3.2
First grade	1.3	2.9	4.6	3.7
Second grade	0.9	2.5	5.3	3.8
All grades combined	1.1	2.6	4.4	3.5
Activities completed				
Kindergarten	13.0	31.0	41.1	43.6
First grade	13.9	32.7	47.8	45.8
Second grade	10.1	31.7	60.8	55.5
All grades combined	12.3	31.6	47.4	46.9
Unique activities completed[a]				
Kindergarten	12.6	29.7	39.8	41.5
First grade	13.5	2.1	46.3	45.5
Second grade	9.1	27.7	54.8	46.9
All grades combined	11.7	29.8	45.1	43.7
Number of days accessed				
Kindergarten	1.1	2.4	3.4	3.3
First grade	0.9	1.9	3.0	2.4
Second grade	0.5	1.5	3.2	2.3
All grades combined	0.8	2.0	3.2	2.8
N for all grades combined	273		71	

Note. In the analytic sample, students who did not access to app at all have values of 0 for each usage statistic.
[a]Unique activities removes any activities that were completed more than one time by a student.

potential to receive gamification and text messaging (Groups C + D + G + H) to students who only had the potential to receive gamification (Groups A + B + E + F). Modeling these treatment effects follows the specifications outlined in the study preregistration: multilevel models nesting students within classrooms.

$$Y_{ij} = \alpha_j + \beta Treat + \Gamma X_{ij} + u_j + \epsilon_{ij}$$

In all models, Level 1 includes student-level covariate adjustments including student pretest reading scores and demographic characteristics such as gender, race/ethnicity, English language learner status, and special education status and an indicator for the random assignment block. We used the spring assessment of the outcome for the MAP and DIBELS outcomes and include the MAP spring assessment as a pretest control in all other regressions. Level 2 includes a teacher-level random intercept to account for the distribution of student achievement around classroom averages. Treatment indicators are included at the appropriate level based on the specific research question (i.e., the classroom-level treatment is included at Level 2 and the individual treatment assignments are included at Level 1).

Results

Use of the MORE@Home App

Prior to looking at the modeled effects, we investigated app take-up and usage over the summer. Table 2 displays wide variability in parents use of the MORE app. Overall, the MORE@Home app was downloaded 134 times during the study period, but because the app was openly accessible on iOS and Android app stores, this meant that nonparticipants may account for some of this volume. Among participants, 71 unique students (26% of the sample) accessed the app. Access rates were similar in kindergarten (32%) and first grade (29%), and slightly lower for second graders (17%). They were also slightly higher among girls (29%) than boys (22%), but this difference was not statistically significant. On average, students who did access the app read 4.4 books and completed 47 activities (see Table 2). A small number of students completed the same activities multiple times on different devices, so we also measured the number of unique activities that students completed, which was slightly lower at 45 activities per student. We also found that students tended to access the app in concentrated chunks of time; the average app user accessed the app over a period of 3.2 days.

TABLE 3

Demographic Characteristics of App Users and Nonusers

Characteristic	Overall Sample	App Users	App Nonusers
Male (%)	45.1	38.0	47.5
Race			
African American (%)	67.0	66.2	67.3
Hispanic (%)	16.5	18.3	15.8
White (%)	4.4	2.8	5.0
Asian (%)	3.7	2.8	4.0
Other (%)	7.0	8.5	6.4
English language learner (%)	9.5	5.6	10.9
Have Individualized Education Plan (%)	8.4	4.2	9.9
Attendance 2017–2018 (%)	94.5	95.2	94.3
Spring MAP RIT score			
Kindergarten	156.8	158.8	155.8
First grade	177.2	186.2	173.5*
Second grade	181.5	183.0	181.2
Spring TRC reading level (rescaled)[a]			
Kindergarten	3.2	3.7	3.0
First grade	9.2	11.5	8.3*
Second grade	10.4	10.6	10.3
Spring MCLASS proficient or above (%)			
Kindergarten	73.5	75.8	72.5
First grade	79.2	86.4	76.4
Second grade	65.6	66.7	65.3
Number of students	273	71	202

Note. MAP = Measure of Academic Progress; TRC = Text Reading Comprehension.
[a]Spring TRC scores converted to numbers ranging from −1 (PR) to 26 (Z). Average kindergarten level is between C and D for kindergarten, between I and J for first grade, and between J and K for second grade.
*$p < .05$.

Analyses comparing demographic characteristics of users and nonusers can shed light on potential sources of selection bias. The results in Table 3 suggest that there were very few statistically significant differences between app users and nonusers on measured baseline demographic and reading skill measures. Grade 1 app users scored significantly higher on spring MAP and TRC, suggesting that Grade 1 children who used the app were better readers than non–app users.

Research Question 1: In Stage 1 of an adaptive literacy intervention, is it better for K–2 students' reading comprehension outcomes to receive pretraining in-school lessons, personalized print texts and app-based digital activities based on (a) 10 conceptually coherent texts (CCT) or (b) 10 leveled texts (LT) on a range of topics?

To address the first research question, we report experimental comparisons in Table 4. The results in the last column display the effect size (ES), the covariate-adjusted standardized mean differences, for three sets of outcome measures. First, we find that the CCT intervention had a slight positive but statistically insignificant effect on app usage statistics relative to LT. Students in this group were 6 percentage points more likely to have accessed the app and completed about 2 more activities. Second, students did score slightly higher on Science Vocabulary Knowledge (ES = 0.10, $p = 0.59$) and Reading Comprehension (ES = 0.14, $p = 0.30$), providing suggestive evidence that the CCT lessons may have enabled students to acquire more science vocabulary and content compared with the LT lessons. Third, we find no statistically significant differences on the MAP reading comprehension test overall or the informational text comprehension subtest or the DIBELS reading outcomes. It is important to note that these comparisons distinguish between those who never used the MORE@Home app and those who ever logged into the app. Thus, the "app user" designation includes those who logged in early (prior to the assessment date) as well as those who first logged in only after receiving additional supports (i.e., the nonresponders). We now turn to the question of how best to intervene to

TABLE 4
Effects of CCT Versus LT

Outcome	*n*	CCT	LT	Diff.	*SE*	*p*	ES
App-based outcomes							
Probability of accessing app	273	0.29	0.23	0.06	0.05	0.27	0.14
Number of books accessed	273	1.22	1.06	0.16	0.42	0.70	0.06
Number of activities completed	273	13.43	11.37	2.06	5.03	0.68	0.07
Number of unique activities completed	273	12.84	10.73	2.11	4.67	0.65	0.07
Number of days accessed	273	0.82	0.88	−0.06	0.33	0.84	−0.03
Standardized reading comprehension outcomes							
Science vocabulary knowledge	152	0.70	0.68	0.02	0.03	0.59	0.10
MAP RIT score	217	174.64	174.70	−0.06	0.31	0.84	0.00
DIBELS composite	226	189.62	190.81	−1.19	−0.04	0.72	−0.01
MAP: Informational text subscores[a]							
Key ideas and details	137	185.50	183.00	2.50	2.19	0.25	0.13
Craft and structure	137	183.94	184.26	−0.32	1.86	0.86	−0.02
TRC level[b]	227	6.86	6.90	−0.04	0.28	0.88	−0.01
Proximal researcher-developed outcomes							
Word study	152	0.45	0.46	−0.02	0.06	0.78	−0.06
Listening comprehension	152	0.47	0.42	0.04	0.04	0.30	0.14

Note. CCT = conceptually coherent texts; LT = leveled text; diff. = raw differences in CCT and LT mean; *SE* = standard error; ES = effect size; MAP = Measure of Academic Progress; TRC = Text Reading Comprehension; DIBELS = Dynamic Indicators of Basic Early Literacy Skills. LT column presents unadjusted group means. All estimates come from two-level hierarchical models with random intercepts to account for nesting within classrooms and the following covariates: pretest, gender, race, English language learner status, Individualized Education Plan status, and grade level (standardized mean difference between CCT and LT).
[a]The proximal student outcomes and MAP informational text subscores were only assessed as outcomes for the first and second grade sample.
[b]TRC scores converted to numbers ranging from −1 (PR) to 26 (Z).
~*p* < .10. *p < .05. **p < .01.

motivate students to participate in the personalized app-based literacy activities.

Research Question 2: Is it better for students' reading comprehension outcomes (a) to augment the Stage 1 treatment with white-hat gamification of the app alone (Gamification Only) or (b) to augment the Stage 1 treatment with gamification of the app and to intensify supports with text message reminders to parents (Gamfication + Texting) if students did not respond to the initial intervention?

Results for Research Question 2 compare the effects of the Gamification + Texting and Gamification Only conditions and are presented in Table 5. Students assigned to the Gamification + Texting support condition were not significantly more likely to use the MORE@Home app by any of the outcomes presented. However, in our sample, these students were 5 percentage points more likely to use the app and completed 3 more activities than students who received just gamification. In the fall, Gamification + Texting students scored marginally significantly higher on the MAP (ES = 0.10, *p* = 0.09) and significantly higher on the MAP subscore related to the craft and structure of informational

text (ES = 0.27, *p* = 0.02) compared with students in the Gamification Only condition. While not significant, Gamification + Texting students also score higher on the MAP subscore related to key ideas and details from informational text (ES = 0.20, *p* = 0.12). These results indicate that the combination of gamification and text messages was more effective than gamification alone in producing student reading comprehension as measured by the MAP.

Using the Parent Surveys to Understand Results

In Table 6, we present description results from a small parent survey. The items were designed to assess whether parents and their children use the personalized print texts and which app features were used. Importantly, the results indicate that parents read an average of 7.7 MORE books to their children and reported their children independently read 6.6 MORE books. Over 50% of parents reported that they "very often" engaged in language activities like reading to their children and talking to their children, and 86% of parents used the "guided questions" in the app, questions designed to foster social interaction and discussion. Across all conditions and combining students who did and did not use the app, parents reported that they read on average 7 of

TABLE 5

Effects of Gamification and Texting (G + T) Versus Gamification Alone (G)

Outcome	n	G + T	G	Diff.	*SE*	p	ES
App-based outcomes							
Probability of accessing app	273	0.28	0.22	0.05	0.05	0.33	0.12
Number of books accessed	273	1.29	0.85	0.45	0.31	0.14	0.19
Number of activities completed	273	13.64	9.42	4.21	3.70	0.26	0.14
Number of unique activities completed	273	13.27	8.66	4.61	3.48	0.19	0.18
Number of days accessed	273	0.92	0.61	0.30	0.24	0.20	0.19
Standardized reading comprehension outcomes							
MAP RIT score	217	176.23	174.37	1.86	1.11	0.09	0.10˜
DIBELS composite	226	194.24	194.61	−0.37	5.46	0.95	0.00
MAP: Informational text subscores[a]							
Key ideas and details	137	184.74	181.44	3.30	2.12	0.12	0.20
Craft and structure	137	185.28	181.16	4.12	1.78	0.02	0.27*
TRC[b]	227	6.92	7.07	−0.15	0.20	0.47	−0.03

Note. MAP = Measure of Academic Progress; TRC = Text Reading Comprehension; DIBELS = Dynamic Indicators of Basic Early Literacy Skills; diff. = raw differences in CCT and LT mean; *SE* = standard error; ES = effect size (standardized mean difference between G + T and G). The Game Only column presents unadjusted group means. All estimates come from two-level hierarchical models with random intercepts to account for nesting within classrooms and the following covariates: pretest, gender, race, English language learner status, Individualized Education Plan status, and grade level.
[a]The MAP informational text subscores were only assessed as outcomes for the first and second grade sample. [b]TRC scores converted to numbers ranging from −1 (PR) to 26 (Z).
˜$p < .10$. *$p < .05$. **$p < .01$.

Discussion

Using a SMART design involving 273 kindergarten to Grade 2 children, we developed an adaptive literacy intervention with multiple components that was designed to improve students' reading comprehension outcomes. The adaptive intervention components included pretraining in-school lessons, personalized print texts, and app-based digital reading activities, and follow-up procedures to support students' literacy development. In sum, the results indicate that (a) children who received CCTs performed similarly to children who received LT on reading comprehension outcomes and (b) nonresponders who received both gamification in their app and parent text messages performed better than those nonresponders who received only gamification. Implementation data revealed that one fourth of participants ($n = 71$) downloaded and accessed the app and, with two exceptions in first grade, there were no differences in the demographic characteristics or reading skills of app users and nonusers. Overall, the results underscore the importance of improving procedures for app take-up and the relevance of pairing print and digital resources to nudge parents to support home literacy activities, particularly in the summer

the 10 books that students were provided. While this survey is limited by sample size, the important story here is that parents and children accessed the print texts and the guided questions in the app were the most frequently accessed parent activity.

months when low-income children are at risk of falling behind in reading (Alexander et al., 2007; Kraft & Monti-Nussbaum, 2017). We discuss findings related to each of the main research questions and then highlight limitations and future research directions.

Our first research question was to examine whether an adaptive intervention with personalized print texts and app-based activities should provide kindergarten to Grade 2 children with CCTs or LTs. In Stage 1 of our SMART design, our aim was to develop and then compare two approaches to context personalization. Although children in the CCT and LT conditions performed equally well on posttest reading outcomes, there are stronger theoretical and practical reasons for testing the CCT approach to context personalization in a confirmatory RCT. From a theoretical perspective, there is growing evidence that opportunities to read CCTs are critical to building domain knowledge and fostering reading comprehension (Cromley & Azevedo, 2007; Hirsch, 2016; Kintsch, 2009). Our results provide suggestive evidence that children in the CCT condition learned more science vocabulary than children in the LT condition. Some scholars have also suggested that "just as it's impossible to build muscle without weight or resistance, it's impossible to build robust reading skills without reading challenging text" (Shanahan, Fisher, & Frey, 2012, p. 58). In other words, the *quality* of print exposure may be as critical as the quantity of print exposure in supporting children's reading comprehension outcomes.

From a practical perspective, increased exposure to informational texts may be especially vital to supporting

TABLE 6
Parent Survey Results (n = 12)

Usage statistic	*M* (*SD*)	Never (%)	Rarely (%)	Sometimes (%)	Often (%)	Very Often (%)
MORE Books read						
Parent and child together	7.7 (4.0)					
Child read alone	6.6 (4.6)					
Frequency of specific reading activities						
Read to child		0.0	16.7	16.7	8.3	58.3
Talk about books		0.0	18.2	9.1	9.1	63.6
Sound out words		10.0	10.0	30.0	20.0	30.0
Read with fluency		10.0	14.3	14.3	14.3	42.9
App feature usage rates (among app-users)						
Book description	42.9					
Guided questions	85.7					
Zoo game	42.9					

informational text comprehension for low-income children. That is, using CCTs rather than LTs in an adaptive literacy intervention may increase low-income children's access to informational texts at school and home (Duke, 2004; Kim et al., 2016). Hoffman (2017) has argued that the widely used classroom practice of using "just right" LTs in elementary literacy instruction may have unintended negative consequences for low-income and struggling readers by limiting access to informational texts that young children want to read. Future experimental research should examine whether the CCT approach to context personalization developed for this study can enhance children's reading comprehension outcomes relative to an untreated control group.

Our second research aim was to compare two strategies for nudging nonresponders to participate in the intervention activities. In a standard RCT, the nonresponders—that is, children whose parents did not download and use the app—would not receive any additional follow-up supports. In many out-of-school literacy interventions, particularly those that target the summer months, there is substantial evidence that individual student heterogeneity in response can attenuate the efficacy of the treatment (Augustine et al., 2016; Cooper, Charlton, Valentine, & Muhlenbruck, 2000). To deal with treatment heterogeneity in how students responded to the Stage 1 components, we included Stage 2 procedure to augment or augment and intensify follow-up procedures for nonresponders. Specifically, we developed text messages to nudge parents to complete the app-based activities with their children, as well as a gamification plan in which children could collect animals for a virtual zoo on completion of the app activities. We found that being assigned to receive parental text messages as well as gamification as supports for nonresponsiveness resulted in better reading outcomes for students than being assigned to only gamification.

However, there were no significant differences in app usage between these groups. These results suggest that not all reading activity related to MORE@Home was captured by the app statistic. In addition, the results indicate that the text-message supports encouraged parents to engage in other, non–app-based reading activities with their children, underscoring the importance of pairing print and digital content as a strategy to support at-home reading activities.

Finally, how do we explain the positive effects of gamification and texting compared with gamification alone on the reading comprehension outcomes of nonresponders? In our SMART design, children in the CCTs chose 10 informational print books and LT children chose 5 informational print books (and 5 narrative print books). Thus, children in both the CCT and LT conditions, including both responders and nonresponders, enjoyed more opportunities to interact with informational texts in print and to complete app-based digital activities. Our results are consistent with intervention studies that have experimentally increased the number of informational texts in children's classrooms and demonstrated positive impact on text comprehension outcomes (Duke, 2004). It may be that the text message reminders nudged nonresponders to interact more with informational than narrative texts in ways that supported informational text comprehension outcomes.

Study Limitations and Future Directions

This study underscores several implementation and analytic challenges that are likely to confront researchers seeking to develop and test app-based literacy tools for parents and families to use in out-of-school contexts. For example, only 26% of students accessed the MORE@Home app, a key element of the intervention. Not only is this concerning from an implementation perspective but it also limits the

effectiveness of the SMART design, because the decision rule and tailoring variable relied on the MORE@Home app. It would be particularly concerning if app usage was associated with baseline reading achievement; however, there was no significant difference between app users and non–app users on their baseline MAP scores. In this study, we implemented a family literacy event prior to summer vacation to demonstrate the app, and while it was highly attended, individual students could not access the app that evening, and many attendees did not attempt to use the app after that evening. Future research could work to tie the family engagement night to app access or to integrate app usage into the pretraining school lesson component of the CCT intervention. Both of these approaches would ensure that students and their families are familiar with the technology before summer vacation begins. Additionally, from a measurement perspective, it is important to note that researcher-developed measures of listening comprehension and word study had low reliability ($\alpha = .47$ and $\alpha = .54$, respectively), which may attenuate the estimated impacts. Future work would benefit from refining these measures to more reliably capture specific dimensions of reading skill. Finally, although we found few observed differences between app users and nonusers, it is critical to explore for whom and under what conditions education apps are most effective in fostering meaningful learning and student engagement (Hirsh-Pasek et al., 2015).

Conclusion

Matching the method to the complexities of developing adaptive interventions is a critical need in education and literacy. To our knowledge, this study represents one of the first efforts to deploy a SMART design to develop an adaptive K–2 literacy intervention with personalized print texts and app-based digital activities. Had we conducted a standard RCT including all the Stage 1 and Stage 2 components in a fixed treatment, we would have limited knowledge of how best to begin an adaptive intervention or how best to help nonresponders participate more fully in the intervention. A SMART design is ideally suited to develop adaptive interventions and to identify potential causal levers of change. Tightly coupling SMART designs that guide the development of adaptive interventions with RCT designs that subsequently evaluate intervention effectiveness should help build usable knowledge to improve young children's reading comprehension outcomes.

Acknowledgments

ORCID iD

James S. Kim https://orcid.org/0000-0002-6415-5496

References

Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2007). Lasting consequences of the summer learning gap. *American Sociological Review*, *72*, 167–180. doi:10.1177/000312240707200202

Allington, R. L. (2005). *What really matters for struggling readers: Designing research-based programs* (What Really Matters Series). Boston, MA: Allyn & Bacon.

Almirall, D., Kasari, C., McCaffrey, D. F., & Nahum-Shani, I. (2018). Developing optimized adaptive interventions in education. *Journal of Research on Educational Effectiveness*, *11*, 27–34. doi:10.1080/19345747.2017.1407136

Almirall, D., Nahum-Shani, I., Sherwood, N. E., & Murphy, S. A. (2014). Introduction to SMART designs for the development of adaptive interventions: With application to weight loss research. *Translational Behavioral Medicine*, *4*, 260–274. doi:10.1007/s13142-014-0265-0

Augustine, C. H., McCombs, J. S., Pane, J. F., Schwartz, H. L., Schweig, J., McEachin, A., & Siler-Evans, K. (2016). *Learning from summer: Effects of voluntary summer learning programs on low-income urban youth*. Santa Monica, CA: RAND Corporation.

Ausubel, D. (1968). *Educational psychology: A cognitive view*. New York, NY: Holt, Rinehart and Winston.

Bransford, J. D., & Schwartz, D. L. (1999). Chapter 3: Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education*, *24*(1), 61–100. doi:10.3102/0091732X024001061

Brown, L. T., Mohr, K. A. J., Wilcox, B. R., & Barrett, T. S. (2018). The effects of dyad reading and text difficulty on third-graders' reading achievement. *Journal of Educational Research*, *111*, 541–553. doi:10.1080/00220671.2017.1310711

Brown, P. C., Roediger, H. L., & McDaniel, M. A. (2014). *Make it stick: The science of successful learning*. Cambridge, MA: Harvard University Press.

Brown, R. S., & Coughlin, E. (2007). *The predictive validity of selected benchmark assessments used in the Mid-Atlantic Region* (Issues & Answers Report, REL 2007–No. 017). Washington, DC: U.S. Department of Education.

Burke, W., & Psaty, B. M. (2007). Personalized medicine in the era of genomics. *JAMA Journal of the American Medical Association*, *298*, 1682–1684. doi:10.1001/jama.298.14.1682

Carr, N. (2011). *The shallows: What the Internet is doing to our brains*. New York, NY: W. W. Norton.

Cervetti, G. N., Wright, T. S., & Hwang, J. (2016). Conceptual coherence, comprehension, and vocabulary acquisition: A knowledge effect? *Reading and Writing: An Interdisciplinary Journal*, *29*, 761–779.

Chard, D. J., Vaughn, S., & Tyler, B.-J. (2002). A synthesis of research on effective interventions for building reading fluency with elementary students with learning disabilities. *Journal of Learning Disabilities*, *35*, 386–406. doi:10.1177/00222194020350050101

Chou, Y. (2014). *Actionable gamification: Beyond points, badges, and leaderboards*. Milpitas, CA: Octalysis Media.

Clinton, V. (2019). Reading from paper compared to screens: A systematic review and meta-analysis. *Journal of Research in Reading*, *42*, 288–325. doi:10.1111/1467-9817.12269

Connor, C. M. (2019). Using technology and assessment to personalize instruction: Preventing reading problems. *Prevention Science*, *20*, 89–99. doi:10.1007/s11121-017-0842-9

Connor, C. M., Dombek, J., Crowe, E. C., Spencer, M., Tighe, E. L., Coffinger, S., . . . Petscher, Y. (2017). Acquiring science and social studies knowledge in kindergarten through fourth grade: Conceptualization, design, implementation, and efficacy testing of Content-Area Literacy Instruction (CALI). *Journal of Educational Psychology*, *109*, 301–320.

Connor, C. M., & Morrison, F. J. (2016). Individualizing student instruction in reading: Implications for policy and practice. *Policy Insights from the Behavioral and Brain Sciences*, *3*, 54–61. doi:10.1177/2372732215624931

Cooper, H., Charlton, K., Valentine, J. C., & Muhlenbruck, L. (2000). Making the most of summer school: A meta-analytic and narrative review. *Monographs of the Society for Research in Child Development*, *65*(1), 1–118.

Cromley, J. G., & Azevedo, R. (2007). Testing and refining the direct and inferential mediation (DIME) model of reading comprehension. *Journal of Educational Psychology*, *99*, 311–325.

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, *11*, 227–268. doi:10.1207/S15327965PLI1104_01

Doss, C. J., Fahle, E. M., Loeb, S., & York, B. N. (2018, March). *More than just a nudge: Supporting kindergarten parents with differentiated and personalized text-messages* (NBER working paper #24450). Cambridge, MA: National Bureau of Economic Research.

Duke, N. K. (2000). 3.6 Minutes per day: The scarcity of informational texts in first grade. *Reading Research Quarterly*, *35*, 202–224. doi:10.1598/RRQ.35.2.1

Duke, N. K. (2004). The case for informational text. *Educational Leadership*, *61*(6), 40–44.

Fitzgerald, W. J., Elmore, J., Kung, M., & Stenner, A. J. (2017). The conceptual complexity of vocabulary in elementary-grades core science program textbooks. *Reading Research Quarterly*, *52*, 417–442. doi:10.1002/rrq.184

Foorman, B., Beyler, N., Borradaile, K., Coyne, M., Denton, C. A., Dimino, J., . . . Wissel, S. (2016). *Foundational skills to support reading for understanding in kindergarten through 3rd grade* (NCEE 2016-4008). Washington, DC: National Center for Education Evaluation and Regional Assistance (NCEE).

Fountas, I. C., & Pinnell, G. S. (2010). *The Fountas & Pinnell Leveled Book List, K-8+* (2010-12 ed.). Portsmouth, NH: Heinemann Educational Books.

Goldstone, R., & Son, J. (2005). The transfer of scientific principles using concrete and idealized simulations. *Journal of the Learning Sciences*, *14*, 69–110. doi:10.1207/s15327809jls1401_4

Good, R. H., & Kaminski, R. A. (Eds.). (2002). *Dynamic indicators of basic early literacy skills*™*: 6th edition: Administration and scoring guide*. Eugene, OR: Institute for the Development of Educational Achievement.

Graham, S., Liu, X., Aitken, A., Ng, C., Bartlett, B., Harris, K. R., & Holzapfel, J. (2017). Effectiveness of literacy programs balancing reading and writing instruction: A meta-analysis. *Reading Research Quarterly*, *53*, 274–304. doi:10.1002/rrq.194

Greene, P. (2018, September 10). Scaling up personalized education. *Forbes*. Retrieved from https://www.forbes.com/sites/petergreene/2018/09/10/scaling-up-personalized-education/#3dbc4b2c735e

Guthrie, J. T., McRae, A., & Klauda, S. L. (2007). Contributions of concept-oriented reading instruction to knowledge about interventions for motivations in reading. *Educational Psychologist*, *42*, 237–250. doi:10.1080/00461520701621087

Guthrie, J. T., Wigfield, A., Barbosa, P., Perencevich, K. C., Taboada, A., Davis, M. H., . . . Tonks, S. (2004). Increasing reading comprehension and engagement through concept-oriented reading instruction. *Journal of Educational Psychology*, *96*, 403–423. doi:10.1037/0022-0663.96.3.403

Hassrick, E. M., Raudenbush, S., & Rosen, L. (2017). *The ambitious elementary school: Its conception, design, and implications for educational equality*. Chicago, IL: University of Chicago Press.

Hedges, L. V. (2018). Challenges in building usable knowledge in education. *Journal of Research on Educational Effectiveness*, *11*, 1–21. doi:10.1080/19345747.2017.1375583

Hirsch, E. D. (2016). *Why knowledge matters: Rescuing our children from failed educational theories*. Cambridge, MA: Harvard Education Press.

Hirsh-Pasek, K., Zosh, J. M., Golinkoff, R. M., Gray, J. H., Robb, M. B., & Kaufman, J. (2015). Putting education in "educational" apps: Lessons from the science of learning. *Psychological Science in the Public Interest*, *16*, 3–34. doi:10.1177/1529100615569721

Hoffman, J. V. (2017). What if "just right" is just wrong? The unintended consequences of leveling readers. *The Reading Teacher*, *71*, 265–273. doi:10.1002/trtr.1611

Jones, J. S., Conradi, K., & Amendum, S. J. (2016). Matching interventions to reading needs: A case for differentiation. *The Reading Teacher*, *70*, 307–316. doi:10.1002/trtr.1513

Kim, J. (2018). *Using a SMART (sequential multiple assignment randomized trial) design to develop a personalized literacy intervention to improve reading engagement and comprehension* (AEA RCT Registry). Retrieved from https://www.socialscienceregistry.org/trials/2983/history/48879

Kim, J. S., Guryan, J., White, T. G., Quinn, D. M., Capotosto, L., & Kingston, H. C. (2016). Delayed effects of a low-cost and large-scale summer reading intervention on elementary school children's reading comprehension. *Journal of Research on Educational Effectiveness*, *9*(Suppl. 1), 1–22. doi:10.1080/19345747.2016.1164780

Kintsch, W. (2009). Learning and constructivism. In S. Tobias & T. M. Duffy (Eds.), *Constructivist instruction: Success or failure?* (pp. 223–241). New York, NY: Routledge.

Kraft, M. A., & Monti-Nussbaum, M. (2017). Can schools enable parents to prevent summer learning loss? A text-messaging field experiment to promote literacy skills. *ANNALS of the American Academy of Political and Social Science*, *674*, 85–112. doi:10.1177/0002716217732009

Kremar, M., Grela, B., & Lin, K. (2007). Can toddlers learn vocabulary from television? An experimental approach. *Media Psychology*, *10*, 41–63. doi:10.108/1521326070130093

Mayer, R. E. (2008). Applying the science of learning: Evidence-based principles for the design of multimedia instruction. *American Psychologist*, *63*, 760–769. doi:10.1037/0003-066X.63.8.760

Mol, S. E., & Bus, A. G. (2011). To read or not to read: A meta-analysis of print exposure from infancy to early adulthood. *Psychological Bulletin*, *137*, 267–296. doi:10.1037/a0021890

Nahum-Shani, I., Qian, M., Almirall, D., Pelham, W. E., Gnagy, B., Fabiano, G. A., . . . Murphy, S. A. (2012). Experimental design and primary data analysis methods for comparing adaptive interventions. *Psychological Methods*, *17*, 457–477. doi:10.1037/a0029372

North Carolina Department of Public Instruction. (2018). *North Carolina formative, diagnostic reading assessment: Reading 3D*. Raleigh, NC: Author.

Patall, E. A., Cooper, H., & Wynn, S. R. (2010). The effectiveness and relative importance of choice in the classroom. *Journal of Educational Psychology*, *102*, 896–915. doi:10.1037/a0019545

Perfetti, C. A., & Stafura, J. (2014). Word knowledge in theory of reading comprehension. *Scientific Studies of Reading*, *18*, 22–37. doi:10.1080/10888438.2013.827687

Robinson, C. D., Lee, M. G., Dearing, E., & Rogers, T. (2018). Reducing student absenteeism in the early grades by targeting parental beliefs. *American Educational Research Journal*, *55*, 1163–1192. doi:10.3102/0002831218772274

Roseberry, S., Hirsh-Pasek, K., Parish-Morris, J., & Golinkoff, R. M. (2009). Live action: Can young children learn verbs from video? *Child Development*, *80*, 1360–1375. doi:10.1111/j.1467-8624.2009.01338.x

Rubin, D. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*, 688–701.

Sabatine, M. S. (2018). Pharmacogenetics and the promise of personalized medicine. *JAMA Cardiology*, *3*, 408–408. doi:10.1001/jamacardio.2018.0586

Sawyer, R. K. (Ed.). (2006). *The Cambridge handbook of the learning sciences*. New York, NY: Cambridge University Press.

Sénéchal, M. (2006). *The effect of family literacy interventions on children's acquisition of reading from kindergarten to grade 3*. Portsmouth, NH: RMC Research Corporation.

Shanahan, T., Callison, K., Carriere, C., Duke, N. K., Pearson, P. D., Schatschneider, C., & Torgesen, J. (2010). *Improving reading comprehension in kindergarten through 3rd grade: A practice guide* (NCEE 2010-4038). Washington, DC: Institute of Education Sciences. Retrieved from whatworks.ed.gov/publications/practiceguides

Shanahan, T., Fisher, D., & Frey, N. (2012). The challenge of challenging text. *Educational Leadership*, *69*(6), 58–62.

Singer, L. M., & Alexander, P. A. (2017a). Reading across mediums: Effects of reading digital and print texts on comprehension and calibration. *Journal of Experimental Education*, *85*, 155–172. doi:10.1080/00220973.2016.1143794

Singer, L. M., & Alexander, P. A. (2017b). Reading on paper and digitally: What the past decades of empirical research reveal. *Review of Educational Research*, *87*, 1007–1041. doi:10.3102/0034654317722961

Solis, M., Miciak, J., Vaughn, S., & Fletcher, J. M. (2014). Why intensive interventions matter: Longitudinal studies of adolescents with reading disabilities and poor reading comprehension. *Learning Disabilities Quarterly*, *37*, 218–228. doi:10.1177/0731948714528806

Stanovich, K. E. (2000). *Progress in understanding reading: Scientific foundations and new frontiers*. New York, NY: Guilford.

Thaler, R., & Sunstein, C. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. New Haven, CT: Yale University Press.

Trainin, G., Hayden, H. E., Wilson, K., & Erickson, J. (2016). Examining the impact of QuickReads' technology and print formats on fluency, comprehension, and vocabulary development for elementary students. *Journal of Research on Educational Effectiveness*, *9*(Suppl. 1), 93–116. doi:10.1080/19345747.2016.1164778

U.S. Department of Education. (2016). *Future ready learning: Reimagining the role of technology in education*. Washington, DC: Office of Educational Technology. Retrieved from http://tech.ed.gov/files/2015/12/NETP16.pdf

Vaala, S., Ly, A., & Levine, M. H. (2015). *Getting a read on the app stores: A market scan and analysis of children's literacy apps*. New York, NY: Joan Ganz Cooney Center at Sesame Workshop. Retrieved from https://files.eric.ed.gov/fulltext/ED574396.pdf

Vitale, M. R., & Romance, N. R. (2012). Using in-depth science instruction to accelerate student achievement in science and reading comprehension in grades 1-2. *International Journal of Science and Mathematics Education*, *10*, 457–472. doi:10.1007/s10763-011-9326-8

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Walkington, C. (2013). Using adaptive learning technologies to personalize instruction to student interests: the impact of relevant contexts on performance and learning outcomes. *Journal of Educational Psychology*, *105*, 932–945.

Walkington, C., & Bernacki, M. L. (2018). Personalization of instruction: Design dimensions and implications for cognition. *Journal of Experimental Education*, *86*, 50–68. doi:10.1080/00220973.2017.1380590

York, B. N., Loeb, S., & Doss, C. (2017, November). *One step at a time: The effects of an early literacy text messaging program for parents of preschoolers* (NBER Working Paper No. 20659). Retrieved from https://www.nber.org/papers/w20659.pdf

Zvoch, K., & Stevens, J. J. (2015). Identification of summer school effects by comparing the in- and out-of-school growth rates of struggling early readers. *Elementary School Journal*, *115*, 433–453.

Authors

JAMES S. KIM is professor of education at the Harvard Graduate School of Education. He is an expert on literacy intervention and experimental design. His current research priority is to understand how building children's domain knowledge and reading engagement can foster long-term improvements in reading comprehension.

CATHERINE A. ASHER is a doctoral candidate at the Harvard Graduate School of Education. She studies large-scale literacy interventions and methodological approaches for understanding and predicting treatment effect heterogeneity.

MARY BURKHAUSER is a senior researcher at the Harvard READS Lab. She designs literacy interventions that integrate a variety of evidence-based practices. She works with teachers to study and improve the feasibility and effectiveness of these interventions.

LAURA MESITE is a doctoral candidate at the Harvard Graduate School of Education. She studies the role of sociocultural factors in reading development as well as the processes used to identify and remediate reading difficulties among children in the United States.

DIANA LEYVA is an associate professor of psychology at the University of Pittsburgh. She examines how parents support children's language, literacy, and math skills in authentic familial practices (e.g., food routines) and uses that information to design and evaluate culturally relevant family interventions.