A longitudinal randomized trial of a sustained content literacy intervention from first to second grade: Transfer effects on students' reading comprehension.

A Longitudinal Randomized Trial of a Sustained Content Literacy Intervention From First to Second Grade: Transfer Effects on Students' Reading Comprehension

James S. Kim[1]

Mary A. Burkhauser [1]

Jackie Eunjung Relyea[2]

Joshua B. Gilbert[1]

Ethan Scherer[1]

Jill Fitzgerald[3]

Douglas Mosher[1]

Joseph McIntyre[1]

[1]Graduate School of Education, Harvard University

[2]College of Education, North Carolina State University

[3]School of Education, The University of North Carolina at Chapel Hill

Kim, J. S., Burkhauser, M. A., Relyea, J. E., Gilbert, J. B., Scherer, E., Fitzgerald, J., Mosher, D. & McIntyre, J. (2023). Longitudinal Randomized Trial of a Sustained Content Literacy Intervention from First to Second Grade: Transfer Effects on Students' Reading Comprehension. *Journal of Educational Psychology*, *115*(1), 73–98. [DOI]

Abstract

We developed a sustained content literacy intervention that emphasized building domain and topic knowledge from Grade 1 to Grade 2 and evaluated transfer effects on students' reading comprehension outcomes. The Model of Reading Engagement (MORE) intervention emphasizes thematic lessons that provide an intellectual framework for helping students connect new learning to a general schema (i.e., how scientists study past events). A total of 30 elementary schools (*N* = 2,952 students; *N* = 144 teachers) were randomly assigned to a treatment or control group. Over 12 months, the treatment group students participated in (a) spring Grade 1 thematic content literacy lessons in science and social studies followed by wide reading of thematically related informational texts during summer, and (b) fall to spring Grade 2 thematic content literacy lessons in science. After implementation of Grade 1 thematic content literacy lessons and summer support for reading, treatment group students experienced smaller summer losses on a domain-general measure of reading than control group students. Following the sustained implementation of thematic content literacy lessons in science through Grade 2, treatment group students also outperformed their control group peers on a science content reading comprehension outcome (ES = .18). Furthermore, we found transfer effects on science content reading comprehension that varied by passage-item type (near-, mid-, and far-transfer passages determined by the inclusion and number of directly taught words in passages). A sustained content literacy intervention that aligns content and instruction across grades can help students transfer knowledge to novel reading comprehension tasks.

Educational Impact and Implications Statement

This study highlights the benefit of sustaining learning of related topics and aligning science content across grades so that children can read with greater comprehension. Early elementary grade students' acquisition of domain and topic knowledge is critical to reading and understanding complex content-rich informational texts. Sustained and thematic content literacy instruction can help elementary grade students transfer their knowledge to reading texts about related topics in science and social studies. Therefore, sustained content literacy intervention efforts that gradually build thematic connections across grades and across school and home contexts may help young children connect new learning to a general schema and transfer their knowledge to related topics. Compared to students in control group classrooms, students who participated in a sustained content literacy intervention from first to second grade made larger improvements on both general reading comprehension and science content reading comprehension outcomes.

Keywords: content literacy intervention, transfer, domain and topic knowledge, reading comprehension, cluster randomized controlled trial

Correspondence concerning this article should be addressed to James S. Kim, Harvard University, Graduate School of Education, 14 Appian Way, Larsen Hall 505, Cambridge, MA 02138, United States. Email: james_kim@harvard.edu

To succeed in school and to prepare for college, young children must acquire the domain and topic knowledge needed to read complex texts. Domain knowledge refers to how much a learner knows about school subjects like science or social studies, while topic knowledge refers to how much a learner knows about specific topics with a given domain (Alexander, 2003).

Despite the growing literacy demands in the 21st century, fewer than 5% of US schoolchildren in the early elementary grades can evaluate complex nonfiction texts with high knowledge demands (Reardon et al., 2012). Furthermore, there are larger gaps based on family income and student ethnicity on reading comprehension outcomes rather than basic word reading skills (D'Agostino & Rodgers, 2017; Fryer & Levitt, 2006; Reardon et al., 2012). These early difficulties in children's ability to read for understanding have consequences in the later elementary school grades. For example, the 2019 administration of the National Assessment of Educational Progress (National Center for Education Statistics, 2019) revealed that the reading comprehension performance of fourth graders has declined over time, particularly for low-income children and low-performing readers. Why are so many elementary-grade children struggling to read complex informational texts with high knowledge demands? Limited opportunities for children to participate in sustained content literacy instruction that provides meaningful learning and continuity across grades may play a pivotal role (Pearson et al., 2020, Stipek et al., 2017).

Researchers and theorists have long underscored the need for developing content literacy interventions that sustain and align content and practices across grades and across school and home contexts (Alexander et al., 2007; Bailey et al., 2017; Bronfenbrenner & Morris, 2006; Nagy, 2005; Newmann et al., 2001; Pressley et al., 2007). In this study, we argue that sustained content literacy instruction is necessary to facilitate students' ability to transfer their newly acquired knowledge to related topics. We also argue that transfer of knowledge to better understand passages about new topics involves schema instantiation. A critical feature of students' domain knowledge development is mental representation of knowledge in the form of schemas (Alexander, 2003), and schema representation plays a significant role in a reader's comprehension (Anderson, 1984; Anderson & Pearson, 1984). It is likely that considerable time is required for students to develop strong, elaborated generalized schemas that can be accessed and deployed when new, but related, topics are encountered in science and social studies texts. However, sustained content literacy interventions are rare in the early elementary grades and intervention effectiveness for transfer to new topic comprehension is even rarer.

The purpose of this study was to test the hypothesis that a sustained 12-month content literacy intervention for students as they moved from first grade into second grade could improve reading comprehension outcomes. Students were gradually introduced to increasingly complex and related science topics that were structured and sequenced from first to second grade (cf. Hirsch, 1987; Kim et al., 2021b; McKeown et al., 2017). Students were also provided recurrent exposure to thematically related topics through wide reading of informational texts during summer. Special features of the instruction were aimed at supporting students' instantiation of generalized mental schemas that could facilitate comprehension of new topics.

Theoretical and Conceptual Foundation for a Sustained Content Literacy Intervention

In recent years, numerous educational psychologists and literacy researchers have designed and evaluated the effectiveness of content literacy interventions in the early elementary grades (Cabell & Hwang, 2020; Connor et al., 2017; Duke et al., 2021; Williams et al., 2016). Although there is no consensus definition of the term content literacy intervention (Hwang et al., 2021), early elementary grade content literacy programs share the common goal of improving the quantity and quality of students' domain and topic knowledge as a lever for improving reading comprehension (Cromley & Azvedo, 2007; Kendeou & van den Broek, 2016; Kintsch, 1988, 2009). For this study, we define sustained content literacy intervention as an approach in which thematic lessons are designed to provide sustained and recurrent exposure to schema-building activities so that knowledge is more accessible to children as they build coherent text representations (Cervetti & Wright, 2020; Kim et al., 2021b; Smith et al., 2021).

Expert Readers Represent Domain and Topic Knowledge in the Form of Schemas

A key purpose of schooling is to help novice learners build schemas that will serve as intellectual structures for the continued acquisition, organization and application of knowledge needed to learn from school texts (Alexander, 2003). Schemas can be defined as domain and topic-specific knowledge structures that enable students to retrieve, organize, and apply knowledge. In essence, a schema is an abstract knowledge structure that builds domain expertise and organizes a variety of topics within a given domain (Anderson & Pearson, 1984; Bartlett, 1932; Kimball & Holyoak, 2000; Rist, 1989).

Schemas are stored in long-term memory, and as novices acquire greater domain knowledge, they represent and organize that knowledge in larger, abstract, interconnected, and robust schemas (Kimball & Holyoak, 2000). To be maximally useful to novice learners, however, schemas can be neither too concrete nor too decontextualized, but "linked to abstract, generalizable features of situations" (Kintsch, 2009, p. 231). These generalizable schemas serve as scaffolds, helping novice learners make sense of novel situations across a variety of related topics. For example, when young children begin to learn about animal survival, they might start with concrete cases of living animals such as polar bears—what they look like, how they behave, where they live, how they survive, and how they adapt. Then, in subsequent grades, they can leverage this knowledge to study more complex cases of how scientists study the fossils of non-living animals like dinosaurs. Over time, children begin to form a topic-specific schema for animal survival and are introduced to key concepts such as physical characteristics and behaviors, habitats, survival, adaptation, extinction, and fossil evidence. The induction of a general schema from concrete topics can then be employed to situate and understand newly encountered topics about animal survival, including animals that have gone extinct such as dinosaurs.

One of the central purposes of the present study was for children to develop a general schema. General schemas are particularly strengthened when students gradually establish topic-specific schemas, through reading and hearing about specific topics within a particular domain (Kimball & Holyoak, 2000). Moreover, hypothetically, when a general schema is induced through repeated exposure in one domain (e.g., science), it is theoretically possible that the general schema might be used to comprehend texts in another domain (e.g., social studies), particularly if the texts involve the same general schema and include networks of vocabulary that co-occur across domains (Ausubel, 1960; Cervetti & Wright, 2020; Hirsch, 2016; Smith et al., 2021).

Thus, the topic-specific schemas in the present study were structured and sequenced to help students connect knowledge across grades and to build a general schema that could be applied to a variety of novel reading passages in science. In our conceptual framework described in Table 1, the topic-specific schemas share structural components that were designed to help students connect newly acquired knowledge to the general schema (cf. Table 1, Gick & Holyoak, 1983). Table 1 shows the study's intended general schema for how scientists study past events. First, antecedent conditions exist; second, there is a solution plan; and third, an outcome. In

Grade 1, students encountered the topic-specific schema for animal survival (a science domain), and then the general schema was encountered again in Grade 2 via the topic-specific schema about how paleontologists study pre-historic events and animals, like dinosaur mass extinctions. Thus, the schema learned in Grade 2 was built on previously learned topics, as students broadened and deepened their understanding of the generalizable schema (i.e., how scientists use multiple sources of evidence to study past events that cannot be directly observed).

Domain and Topic Knowledge Schemas Support Reading Comprehension

Theorists have long noted that a reader's pre-existing schema is a critical determinant of how well they comprehend a text (Anderson & Pearson, 1984; Bartlett, 1932; Hirsch, 1987; Kintsch & van Dijk, 1978; van Dijk 1977). In the construction-integration model of reading comprehension (Kintsch, 1993; 2009), a significant feature of the text comprehension processes entails building a coherent situation model that incorporates the literal understanding of text with prior knowledge. Both novices and experts enact similar processes to construct the literal textbase—that is, they activate word meanings, hold information in working memory, and form propositions. During the integration phase, however, experts use retrieval structures—that is, instantiations of domain knowledge schemas—to form a superior situation model. There is strong evidence that experienced readers retrieve general and topic-specific schemas held in long-term working memory to store and update the textbase and the situation model (Ericsson & Kintsch, 1995; Kimball & Holyoak, 2000). Schemas make prior knowledge stored in long-term memory available to readers as they integrate that knowledge with new information in texts to make gap-filling inferences and build coherent mental representations of texts. The availability of well-structured schemas enables the expert to retrieve knowledge that is relevant for comprehending text without taxing working memory.

Cognitive psychologists have emphasized, however, that experts tend to represent knowledge in a well-structured schema to a greater extent than novices and generalized schemas are likely to facilitate transfer to new topics (Kendeou & O'Brien, 2016; Kimball & Holyoak, 2000). For example, experts are able to use and transfer a general schema while reading about a variety of topics that differ only in their superficial characteristics (Ericsson, 2018; Ericsson & Pool, 2016; Kintsch, 1988). To acquire domain knowledge from text, children must not only understand what they are reading, but also connect newly learned content to an existing schema that expands their knowledge base and creates new understandings about a topic. For example, a student who recognizes that concepts like dinosaurs, paleontologists, evidence, and theory rest on a general schema for how scientists study pre-historic events are more likely to comprehend passages about topics that share this general schema. Therefore, schema instantiation can facilitate transfer across topics within a domain and enhance comprehension of novel material that fits the same situation. In essence, transfer is facilitated through a process of mapping whereby the reader can find a clear correspondence between each component of the general schema and the topic-specific schemas, as shown in Table 1 (Gick & Holyoak, 1983; Schustack & Anderson, 1979). As novices develop greater domain and topic knowledge, they build more structured and elaborate schemas that facilitate transfer in reading comprehension.

Domain-Specific Academic Vocabulary Networks Activate and Build Schemas

If schema development is critical to fostering transfer in reading comprehension, a key theoretical and practical question becomes: How can we help students build and activate schemas that are general and useful for school learning? To a large extent, embedding domain-specific academic words and word parts that appear across topic-specific schemas may provide the retrieval cues that activate the general schema in the reader's mind and, at the same time, enhance memory of a word's meaning while reading complex texts (Chilton & Ehri, 2015; Karpicke & Blunt, 2011). Theoretically, repeated exposures to the form and meaning of words, and other words with semantic overlap, can enhance a reader's ability to retrieve and apply words' meanings to topic-related texts (Perfetti, 2007). As novices begin to develop well-structured schemas that are supported by a rich and dense network of semantically related vocabulary (Alexander, 1997), they begin to accumulate a deeper knowledge of domain-specific words that are "highly interdependent semantically" (Kimball & Holyoak, 2000, p. 114) and further catalyze schema development, restructuring, and growth.

Experts mentally organize domain-specific vocabulary as rich connections among words whose meanings are built through varied exposures in contexts that convey information about the word's meaning (Kim et al., 2021a; McKeown et al., 2017; Perfetti, 2007). For instance, over time, as children connect the words to the general schema for what scientists (e.g., biologists, paleontologists, archeologists) do, they deepen their knowledge of words in contexts that reveal the meaning of the words and the associated network. The schema fosters a cohesive instructional context that makes it easier for learners to store and retrieve words from memory, particularly as students read and learn from related texts on science and social studies topics.

In all likelihood, there is a synergistic relationship between domain and topic knowledge development, growth in academic vocabulary networks, and, ultimately, improvements in students' reading comprehension (Adams et al., 1995; Kendeou & O'Brien, 2016; Nagy, 2005; Pearson et al., 2010; Perfetti & Stafura, 2014). As students read and study related topics in school, they begin to develop larger and more generalizable schemas, as well as networks of vocabulary knowledge, that support text comprehension.

Little is Known About Sustained Content Literacy Instruction Involving Emphasis on Schema Development for Transfer on Reading Comprehension

Although content literacy interventions for early adolescents have demonstrated success in improving reading comprehension and developing content knowledge (Guthrie & Klauda, 2014; Romance & Vitale, 2001; Vaughn et al., 2013; Williams et al., 2016), less is known about early elementary-grade content literacy intervention and their effects on young children's reading comprehension outcomes. Only a few content literacy interventions in the early elementary grades have been conducted, yielding mixed evidence of success in promoting transfer on reading comprehension (Connor et al., 2017; Kim et al. 2021a, 2021b). Notably, few early grades content literacy interventions have focused on development of generalized schema as a facilitator of transfer for reading novel material in science (Cabell & Hwang, 2020; Vitale & Romance, 2007). For example, in their synthesis of findings from the *Reading for Understanding* initiative, Pearson and colleagues (2020) underscored that transfer is difficult to achieve and measure in reading comprehension.

In other words, many literacy intervention studies use distal measures of reading comprehension, but these measures are not adequately sensitive to assess whether students were able to apply their knowledge to new comprehension tasks. It is critical, therefore, to assess how far intervention effects can travel by using comprehension measures that differ from the instructional content and context. Fundamentally, the concept of transfer (cf. Figure 1 and Barnett & Ceci, 2002) can be broken down into two factors, including the content (i.e., what is transferred) and the context (i.e., when and where learning is transferred from and to).

Stated succinctly, the explanatory theory undergirding the present intervention was: Helping students acquire domain and topic knowledge could simultaneously develop students' mental knowledge schemas. Hypothetically, development of deep, embellished schemas would require time—sustained effort across grades. Knowledge schema instantiation and development should also support students' transfer for comprehension of new, but schema-related topics. Importantly, more sensitive comprehension measures are needed to determine whether and to what extent a sustained content literacy intervention can help students transfer their knowledge to novel reading tasks.

Research Supporting a Sustained Content Literacy Intervention

In developing the Model of Reading Engagement (MORE) intervention, we identified research-based practices to build and deepen students' domain and topic knowledge schemas and, ultimately, to enable students to transfer and apply knowledge to novel reading and writing tasks (Alexander 2000; McKeown et al., 2017; Vaughn et al., 2013). There is growing research that sustained content literacy instruction creates an ideal context for structuring and sequencing exposure to topic-specific schemas that enable students to build and deepen knowledge in science and social studies (e.g., Connor et al., 2017; Kim et al., 2021a, 2021b; Williams et al., 2016). When students are exposed to aligned curriculum and instruction, they are more likely to build on skills and knowledge learned in previous grades and less likely to repeat content covered in previous grades (Engel et al., 2013; Newmann et al., 2001).

Accordingly, we designed and evaluated the effectiveness of sustaining MORE over 12 months, including (a) Grade 1 thematic content literacy lessons in science and social studies followed by wide reading of reading thematically related informational texts in the summer and (b) Grade 2 thematic content literacy lessons in science that built upon and extended schema-related topics in Grade 1. In the MORE thematic content literacy lessons, teachers interleaved practices involving direct teaching of domain-specific academic vocabulary networks, interactive read-alouds to provide recurrent exposures to schema-related vocabulary, and wide reading of thematically related informational texts with autonomy and competence supports. In addition, teachers provided further opportunities for students to transfer their knowledge to collaborative research and structured word inquiry activities. Teachers infused each practice with activities that fostered student engagement, which is a multi-dimensional construct that encompasses students' affective/motivational, cognitive, and behavioral characteristics (Fredricks, Blemenfeld, & Paris, 2004; Guthrie & Wigfield, 2000; Sinatra, Heddy, & Lombardi, 2015).

Interleaving Practices to Build Domain and Topic Knowledge Schemas

Thematic lessons to promote schema development and student engagement. Thematic content literacy lessons (Stoddart et al., 2002) provide students with an intellectual structure for connecting newly acquired knowledge to a general schema. Thematic lessons are built on a pyramid structure where a general schema helps to unify and organize topics in a domain (Guthrie et al., 2007; Wiske, 1998). In the MORE instructional sequence, the Grade 1 thematic unit ("How Do Animals Survive in Their Habitat?") laid the foundation for further and deeper learning of the Grade 2 thematic unit ("How Do Paleontologists Study Prehistoric Events and Animals, Like Dinosaur Mass Extinctions?").

Thus, students were able to "hook" newly acquired knowledge to a general schema ("How do scientists study past events?") by studying fewer related topics in greater depth over time.

During the implementation of the thematic lessons, teachers enacted two research-based practices to build students' reading engagement to acquire knowledge. First, teachers used thematically related informational texts that were designed to spark students' situational interest and sustain their persistence to learn from texts (Guthrie et al., 2006; Kulikowich & Hepfer, 2018; Schiefele, 1999). Second, teachers emphasized mastery goals that were designed to foster students' affective, cognitive, and behavioral engagement. In particular, teachers employed interesting texts and emphasized mastery goal structures, enabling their students to perceive progress in their understanding of topics, to gradually acquire conceptual and vocabulary knowledge, and to build self-competence beliefs and intrinsic value in literacy activities (Guthrie et al., 2012; Hwang et al., 2021; Romance & Vitale, 2001).

Teaching domain-specific vocabulary networks through concept mapping. Within the context of thematic lessons, concept mapping is an instructional practice that helps children "hook" newly acquired domain-specific academic vocabulary into an existing semantic network. A larger network for a focal word facilitates a student's vocabulary acquisition and ability to comprehend passages that include the network of related vocabulary (Kim et al., 2021a; Stahl & Fairbanks, 1986). Teachers enacted concept mapping to build and strengthen connections among domain-specific academic vocabulary in semantic memory (Gelman, 2009; Imai, Gentner, & Uchida, 1994; Levin, 1988; Stahl & Fairbanks, 1986). A concept map is a visual tool for organizing concepts in which pictures are paired with labels, and a node-link-node syntax is used to build a schematic representation of related vocabulary (Nesbit & Adesope, 2006). Through concept mapping of domain-specific academic words, teachers were able to illustrate how well-structured schemas can be visually displayed as a network of semantically related vocabulary (Neuman et al., 2011).

Interactive read-alouds to provide recurrent exposure to vocabulary networks. A primary aim of the thematic lessons was to promote oral language activities and exposure to schema-related vocabulary through interactive read-alouds and discussion. There is substantial evidence that deeper engagement with thematically related informational texts enables student to build their oral language skills and to facilitate the integration of newly acquired vocabulary to a pre-existing schema. For example, interactive read-alouds may help children comprehend sentences with target words (Beck et al., 1982), acquire knowledge of untaught words through incidental learning (Carlisle et al., 2000), and, in some instances, transfer knowledge to novel reading tasks (Elleman et al., 2009; Stahl & Fairbanks, 1986). Furthermore, recurrent exposure to target vocabulary and concepts can foster children's topic interest and task values as well as their persistent to learn from texts with high knowledge demands (Kulikowich & Hepfer, 2018; Marinak et al., 2015).

Wide reading of thematically related texts with autonomy and competence supports. Wide reading affords young children additional opportunities to read about topic-specific schemas, to build depth of conceptual and vocabulary knowledge, and to strengthen intertextual connections (Cervetti et al., 2016, Lupo et al., 2020). Theoretically, it is possible to induce the general schema through repeated exposures to topic schemas that organize the thematically related text sets used in wide reading activities. Furthermore, because the schematic structures associated with narrative text are usually more familiar to young children than informational texts (Best et al., 2008; Smith et. al., 2021; Williams et al., 2016), it is critical for teachers to provide structured opportunities for students to build coherent mental representations using thematically related science and social studies texts. Thus, wide reading may provide additional opportunities for readers to integrate prior knowledge with text-based ideas during independent reading (Kintsch, 1988).

To enhance students' intrinsic motivation (Ryan & Deci, 2000), teachers can also use autonomy and competence supportive practices. Teachers can scaffold wide reading by giving children choice of both grade-level matched texts (i.e., texts used in school year thematic lessons) and reader-level matched texts (i.e., texts personalized to each students' independent reading level). We adapted a previously validated teacher-scaffolded approach to wide reading (Kim & White, 2008; Kim et al., 2016) to increase behavioral engagement (i.e., time on task, amount of reading) with schema-related texts during the summer months to improve domain general reading.

Interleaving Practices to Transfer Domain and Topic Knowledge Schemas

Collaborative research with open literacy tasks. Within the context of a thematic content literacy lesson, teachers can help students transfer knowledge to open literacy tasks without a clear correct answer. In particular, collaborative research activities were enacted to help students develop deep expertise in a topic of choice during the MORE lessons. For example, in Grade 1 science lessons, students could choose one Arctic animal (e.g., snowy owls, Arctic foxes, narwhals) to study in greater depth; and in Grade 2 science lessons, students could choose one dinosaur (e.g., Tyrannosaurus rex, Oviraptor, Microratpro, Megalosaurs) to study in greater depth, such that students became experts who organized themselves into teams to read, write, and discuss self-selected topics with their peers. Thus, collaborative research activities provide students with autonomy to make choices, to perform authentic tasks, and to sustain attention on completing appropriately challenging tasks (Guthrie et al., 2007; Parsons et al., 2015). Students were provided an additional context for reading (e.g., reading newspaper and magazine articles) and writing about words and concepts that were critical to developing topic-specific schemas.

Structured word inquiry to build metalinguistic skills. To deepen students' content and vocabulary knowledge, teachers focused on building students' metalinguistic awareness—the ability to reflect on the nature of language and manipulate the structural aspects of language (McKeown et al., 2017; Nagy, 2005). Children's metalinguistic awareness plays a critical role in the development of vocabulary knowledge and reading comprehension (Nagy, 2007). Emerging evidence indicates that structured word inquiry activities that involve delving deeply into the lexical structure of morphologically complex words and identifying relations among semantically related words can improve morphological awareness and vocabulary knowledge (Colenbrander et al., 2021; Georgiou et al., 2021). Students learn about morphological structure and etymology of the words that they encountered via concept mapping, interactive read-alouds, and wide reading activities. For example, the Grade 2 MORE science lessons include several morphologically complex words (e.g., *paleontologist*) in the concept mapping and interactive read-alouds. Teachers explicitly defined the word, explained how the word was built, discussed related words, and highlighted important sounds (Bowers et al., 2010; Bowers & Bowers, 2018).

Summary: Making Thematic Approaches to Content Literacy Effective in the Early Elementary Grades

This study builds on previous research evaluating the design, implementation, and effectiveness of the core components of the MORE intervention. Previous research highlights the unique contributions of each component of MORE (e.g., Kim et al., 2021a; 2021b). In early efficacy studies involving Grade 1 science thematic lessons, MORE improved reading, vocabulary, and writing outcomes relative to control students in Tier I literacy instruction. In addition, previous research suggests that knowledge development through wide reading of thematically related texts in the summer can further extend and build on school lessons and improve student reading comprehension outcomes (Kim et al., 2016; 2017).

The Present Study

The main goal of the present study was to design and test a full 12-month intervention involving (a) spring 2019 Grade 1 thematic content literacy lessons in science and social studies followed by wide reading of reading thematically related informational texts in the summer, and (b) fall 2019 to spring 2020 Grade 2 thematic content literacy lessons in science. Our original implementation plan had to be modified due to the COVID-19 pandemic, which led to school closings in our study site in spring 2020. As a result, most Grade 2 teachers were unable to implement the subsequent social studies unit after the completion of the science unit. Thus, we were unable to administer the originally planned Grade 2 social studies content reading comprehension assessments and a domain-general reading comprehension assessment in spring 2020. Details on the original study are in our preregistration plan (Kim, 2021) and the online supplementary materials. Despite this limitation, we were able to test the impact of the MORE intervention on domain-general reading comprehension outcomes for students in the longitudinal Grade 1 to Grade 2 cohort. Accordingly, we addressed the following research questions:

1. What is the effect of the Grade 1 MORE thematic content literacy lessons in spring followed by wide reading of thematically related informational texts in summer on students' growth in domain-general reading comprehension and vocabulary from the winter of Grade 1 to the fall of Grade 2?
2. Following implementation of the Grade 1 MORE thematic content literacy lessons and wide reading of thematically related informational texts, what is the effect of sustaining MORE thematic content literacy lessons on Grade 2 students' science content reading comprehension and vocabulary knowledge depth?
3. Does the magnitude of the transfer effect on Grade 2 science content reading comprehension outcome differ by item-passage type?

Method

Research Design

We conducted a longitudinal cluster (school) randomized controlled trial (RCT) to test the effectiveness of sustaining the MORE content literacy intervention from Grade 1 to Grade 2 on students' reading comprehension outcomes. A total of 30 elementary schools in one urban school district located in the southeastern United States participated in the 12-month program implementation. Prior to program implementation, we pre-registered the study design, conducted a power analysis to determine the minimum detectable effect size, and implemented a blocked, cluster (school-level) randomized controlled trial (Kim et al., 2021b). In particular, schools were blocked by three levels of school size (i.e., small, medium, large schools based on total student enrollment), two levels of prior reading proficiency (i.e., above and below the median on end-of-grade 3 reading performance), and prior experience with MORE and then randomized to treatment or control conditions for the 12-month study period.

Figure 1 provides a visual description of the timeline for implementing the intervention activities and outcomes measures over a 12-month period. At the beginning of year 1 in spring 2019, half the schools were randomly assigned to treatment in Grade 1 treatment lessons and half were assigned to Grade 2 treatment lessons. Grade 1 treatment schools included controls at Grade 2 and Grade 2 treatment schools included controls at Grade 1. Thus, Grade 1 treatment schools provided valid counterfactuals for the Grade 2 schools, and vice versa. The planned assessments in Grade 3 MORE were not implemented due to school closures in spring 2020 (see Supplementary Materials S-A). Thus, we were able to examine the intervention treatment effects for students who started the study in the spring of Grade 1 and continued through the end of the study in the spring of Grade 2.

To address our first research question, we implemented the spring Grade 1 content lessons and wide reading activities and measured growth using data from three measurement occasions of the Measure of Academic Progress (MAP) reading comprehension and vocabulary assessment. To address our second and third research questions, we sustained content literacy instruction through the fall and spring of Grade 2 and then administered the domain-specific (science) measure of reading comprehension and vocabulary knowledge depth.

Consort Flow and Study Participants

The consort flow diagram in figure 2 provides detailed information on attrition rates. At the first follow-up in grade 2, there were a total of 2,275 students included in the piecewise linear growth models analyzing growth in domain-general reading comprehension. Thus, 23% of the students were lost to attrition from the initially randomized sample (N = 2,952) to the analyzed sample at first follow-up (N = 2,275). At the second follow-up, 27% of students were lost to attrition and included in the second follow-up (N = 2,156) analyses of domain-specific reading comprehension at the end of grade 2. At the final follow-up, there was no statistically significant difference in attrition rates based on experimental condition, $\chi^2(1) = 1.05$, $p > .05$. Overall, the analytic sample of students who completed the final science content comprehension outcome included 72% of the treatment group (n = 1,176) and 74% of the control group (n = 980).

Table 2 displays descriptive characteristics of participating teachers (N = 144) and students (N = 2,275) at the beginning of the study in Grade 1. Participating teachers were mostly female (97%) and white (62%), about half had earned or were working towards a master's degree, and 47% had participated in a professional development program focused on the science of reading (reading research to classroom practice). In terms of students' demographic characteristics, Black and Hispanic students comprised the majority of the student sample. Finally, 41% of students were from low socioeconomic status backgrounds and 24% of students had limited English proficiency.

To assess baseline equivalence on student pretest measures for the analytic sample of students included in the first follow-up in fall Grade 2 and the second follow-up in spring grade 2, we fit multi-level models in which the pretest MAP total reading and MAP total mathematics scores were predicted by the treatment indicator, randomization block, and random effects for schools. For the first follow-up analytic sample, there was a statistically significant difference of -0.22 SDs (SE = 0.09, z = -2.50) on the pretest MAP total reading test score and -0.20 SDs (SE = 0.09, z = -2.32) on the pretest MAP total mathematics score. For the second follow-up analytic sample, the standardized mean difference was -0.23 SDs (SE = 0.09, z = -2.45) on the pretest MAP total reading and -0.21 SDs (SE = 0.09, z = -2.44) on the pretest MAP total mathematics. For both analytic samples, the baseline difference between treatment and control conditions on pretest scores was below the threshold of 0.25 SDs established by What Works Clearinghouse (WWC) Standards (2021) and all statistical models included pretest scores to adjust for these baseline differences.

Intervention Description and Study Context

Grade 1 and Grade 2 English Language Arts Tier I Curriculum and Instruction

In Grade 1, the Tier I English language arts (ELA) curriculum was based on a balanced literacy program involving word study, guided reading and writing activities in small groups, teacher-directed lessons, and independent reading. In Grade 2, the district adopted a new Tier I ELA curriculum, Expeditionary Learning (EL), a content-based literacy designed to simultaneously build content knowledge and language and literacy skills (Woodfin & Plaut, 2017). For example, the Grade 2 EL curriculum included a unit on “Learning About Paleontologists” which revolved around the guiding question, “What Do Paleontologists Do?” As a result, there were more similarities between the treatment and control activities in Grade 2.

The MORE Grade 1 lessons and procedures for promoting wide reading of thematically related texts and the MORE Grade 2 lessons were implemented during the supplemental science and social studies content block. In Grade 1 and Grade 2, there was no district-wide science and social studies curriculum and schools had considerable flexibility in the scope and sequence of content instruction.

Grade 1 MORE Thematic Content Literacy Lessons and Professional Development

To implement the MORE Grade 1 lessons, teachers participated in a two-hour in-person professional development meeting and received ongoing support during the implemented MORE content literacy lessons, including daily video clips that further explained critical features of the online lesson guide, materials, and procedures. School-based literacy facilitators also provided support by answering questions about the implementation. The research team hosted monthly meetings for literacy facilitators to communicate information about the MORE lesson components and to address teachers' questions. The intervention began with a 10-lesson sequence thematic unit in science ("How Do Animals Survive in Their Habitat?) followed by a 10-lesson thematic unit in social studies (How Do Explorers Overcome Obstacles?). Teachers were given substantial flexibility in their pacing calendar to teach the 20 lessons over a 5- to 10-week window from February to March 2019.

During the last two weeks of the school year, Grade 1 teachers implemented a procedure for promoting wide reading of thematically related informational texts in the summer. Students completed an online survey where they were given an opportunity to choose 10 thematically related science and social studies print books to read during the summer (Supplementary Materials S-B includes the list of the 33 thematically related Grade 1 print books and 6 digital books). First, teachers provided students with a choice of five grade-level matched texts, which included texts that teachers used during grade 1 thematic lessons. Students were prompted to choose books from the *Who would win?* series and two books from the series on *what if you had animal feet* (animal eyes, animal ears, or T. rex teeth). The grade-level matched texts were designed to provide repetition and extended examples of the topic-specific schema in grade 1 school and to foster inter-textual connections related to the theme of animal survival. Second, teachers organized students into four groups (low, low-mid, mid-high, high) based on their assessment of students' reading level and then students chose 5 reader-level matched books that were related to science (i.e., animal survival) and social studies themes (i.e., explorers). The reader-level matched texts were designed to be just difficult enough to support active processing as students drew upon their knowledge to read independently at home (Kim & Quinn, 2013; Smith et al., 2021). Treatment group students received their print books during the last week of school during first grade. As an incentive for participation in the study, both treatment and control group parents were given access to an app where their children could choose six digital books to read on contemporary and historical figures who overcame obstacles.

Grade 2 MORE Thematic Content Literacy Lessons and Schedule

In the Grade 2 science unit, there were 45 lessons that focused on a single theme (*dinosaurs*) and organizing question (*How do paleontologists study dinosaurs?*). The unit was designed to be taught over a period of 10 weeks, starting in late October and ending in late January, and teachers were given a pacing calendar to support their planning. The unit included nine lesson activity types (see Table 3), which comprised a lesson sequence that was repeated a total of five times across the 10-week unit. The first part of the four lesson activity types in each lesson sequence focused on building and deepening domain and topic knowledge through interactive read-alouds and concept mapping. Specifically, each sequence began with a lesson type called “zoom,” which was designed to generate situational interest in the upcoming read-aloud. This activity type was always followed by read-aloud lessons to build background knowledge; a concept mapping lesson to both build academic vocabulary and think together about the relationships between and among the concepts under study; and a notetaking lesson to organize relevant textual details.

The second part of the five lesson activity types in each lesson sequence focused on deepening domain and topic knowledge. Teachers used an inquiry-based approach to build morphological and phonological awareness of words that appear in the concept maps. In particular, the structured word inquiry instruction (Bowers et al., 2010) emphasized morphology, phonology, and etymology by helpings students recognize that target words like paleontologist refer to a scientist who studies ancient things, that the morpheme *paleo-* (old) and *-gist* (one who studies) signal the word’s meaning, and that related words like *paleontologist* share morphemic and phonemic units. They also engaged in collaborative research using common text features and structures (e.g., compare and contrast) to obtain additional information on given concepts about dinosaurs during reading and writing activities and participated in in-depth academic group discussions while incorporating relevant details and evidence from their research. Finally, each lesson sequence ended with a reflection activity during which students considered what they had learned across the sequence, what they had enjoyed, and what they still wanted to learn.

Grade 2 Professional Development

Before program implementation started, teachers participated in an interactive, three-hour professional development session. The session focused on the core MORE practices, as well as core lesson components and instructional routines. During the implementation period, the research team provided treatment teachers with ongoing support through daily video clips (5-10 minutes) that were anchored to that day's lesson components. Treatment teachers also received continuing support for lesson implementation and data collection processes from their school's literacy facilitator. The research team hosted a series of monthly meetings for literacy facilitators to communicate information on lesson implementation and to address teachers' questions about implementation.

Procedures for Selecting Domain-Specific Vocabulary in Thematic Lessons

We used an iterative procedure to identify domain-specific vocabulary. First, we anchored the words to the state's science standards and the Next Generation Science Standards (National Research Council, 2012). We conducted a content analysis of each standards document as well as our lesson texts to identify the related vocabulary. Second, we cross-validated these words against content standards that predated the state standards and NGSS to ensure that the words were relatively stable features of U.S. school curricula over time (Hirsch, 2016; National Research Council, 2012). Third, we created an automated concept network for the Grade 1 and Grade 2 thematic lessons containing the target academic words and additional semantically associated words also appearing in the lesson texts. Each target word was represented as a node with weighted connections between nodes indicating the degree of similarity. The Grade 1 science words cohered around the theme of animal survival and were designed to help students apply these words to animals in a variety of habitats. Grade 2 science words (i.e., *theory, fossil, evidence*) focused on the question of how scientists use evidence to develop theories with specific applications to the topic of mass extinctions and dinosaurs.

Using the concept maps for all target concepts, we identified two types of academic vocabulary words for each grade: seven semantically associated words directly taught in the lessons of each unit, particularly in the concept mapping and structured word inquiry activities;

and five related words that appeared in the lesson texts, but were not explicitly taught during the lessons. The five untaught words were in the range of lower frequency, higher age of acquisition, and/or lower concreteness at each grade level. We selected these untaught words for the purpose of assessing transfer on the vocabulary knowledge depth measure. We replicated these procedures to choose domain-specific vocabulary used in social studies lessons.

Fidelity of Implementation for Grade 1 and Grade 2 MORE Intervention

Table 4 displays the results of Grade 1 and 2 fidelity of implementation. The following section reports the fidelity of implementation results by grade level because slightly different measures were used for each grade. In Grade 1, we measured fidelity using audio-recorded lessons and teacher surveys. In Grade 2, only teacher surveys were administered.

Grade 1 Fidelity of Implementation

Adherence to Core Intervention Components. Grade 1 teachers' adherence to the MORE lesson components was assessed using audio recordings. We first created an adherence checklist that outlined 11 indicators that were essential to the intervention core components (see Supplementary Materials S-A for the adherence checklist). Then, a research assistant listened to 25 audio-recorded lessons randomly selected from the treatment teachers and tallied the presence (1 point) and absence (0 point) of each of the 11 indicators. A percentage score was obtained by multiplying the average score by 100 to estimate the degree of adherence for each teacher. After an initial coding, the second research assistant independently double-coded a randomly selected 44% subset of the lesson audio-recordings to obtain inter-rater agreement. Overall agreement was 91% (Cohen's $\kappa = 0.63$). More detailed information on Grade 1 fidelity of implementation results has been reported in our previous study (i.e., Kim et al., 2021b). Average adherence was high (100%) across all teachers.

Program Differentiation. We evaluated program differentiation between treatment and control groups to identify unique features of two conditions by measuring complexity of interactive read-aloud texts and openness of the literacy tasks (Parsons et al., 2015) A higher level of openness meant that literacy tasks were more likely to promote student engagement and student-centered learning than tasks with lower openness scores (Duke et al., 2006).

To measure complexity of read-aloud texts, we first conducted a teacher survey to collect information on the titles of the books that Grade 1 teachers used for interactive read-aloud activities over the implementation period and then obtained the Lexile levels of those books. The average Lexile levels of the books used in the treatment group (science books M = 696.25L; social studies books M = 941.67L) was much higher than those from the control group (science books M = 522.63L; social studies books M = 439.09L). The list of science and social studies books used in Grade 1 treatment and control groups is provided in the Supplementary Materials S-B.

To evaluate the openness of the literacy tasks, we audio-recorded ELA lessons of 12 randomly selected treatment and 11 randomly selected control classrooms, identified the types of literacy tasks, and rated the following five characteristics of each literacy task on a 3-point Likert scale (1 = closed, 2 = moderately open, and 3 = open task): authenticity (simulating real-life activity), collaboration (collaborative or independent activity), challenge (intellectually stimulating activity), student-directed work (involvement of student input), and sustained effort (sustainability over time). Overall inter-rater agreement ranged from 91 to 97%. As shown in Table 4, literacy tasks in the treatment classrooms were more likely characterized as authentic, collaborative, challenging, student-directed, and sustained relative to those observed from the control classrooms.

Students' Exposure to ELA, Science, and Social Studies Instructional Time. We administered a survey for all teachers to estimate the amount of time teachers devoted to ELA, science, and social studies instruction during this study. Overall, treatment group teachers (M = 177.34, SD = 130.46) reported spending approximately 68 more minutes per week on science instruction than control teachers (M = 109.05, SD = 99.37). Likewise, treatment group teachers (M = 181.82, SD = 133.05) devoted nearly 72 more minutes per week to social studies instruction relative to control teachers (M = 108.90, SD = 100.60). However, the amount of time on ELA/reading instruction was similar for treatment and control groups.

Wide Reading of Thematically Related Texts. We administered a school-based survey in the early fall of Grade 2 to assess the amount of wide reading that occurred in the summer months following Grade 1. Both treatment and control students' survey asked: "On the MORE @ Home app, there were six digital books you could read. About how many of these digital books did you read over the summer?" Since only treatment group students received print books, their surveys also asked: "Last spring, you picked 10 free books to read at home. About how many of your free books did you read over the summer?" Both digital and print books were related to the thematic units taught during the Grade 1 science and social studies lessons. Treatment group students reported reading more than half the print books (M = 5.39, SD = 4.16) and one-third of the digital books (M = 2.13, SD = 2.63). Control group students reportedly read less than one-third of their digital books (M = 1.57, SD = 2.40).

In summary, the fidelity results from the Grade 1 lessons and summer support suggest that teachers delivered the core components with high fidelity that was comparable to previous studies of content literacy intervention (Vaughn et al., 2013; Williams et al., 2016); furthermore, treatment group children reported reading five more books than control children, consistent with previous research on summer reading (Kim et al., 2017; Phillips & Chin, 2004).

Grade 2 Fidelity of Implementation

Adherence to Core Intervention Components. Table 4 reports Grade 2 treatment teachers' adherence to the MORE lesson sequence and the use of the MORE read-aloud books. Treatment teachers were asked to report the total number of each lesson activity type that they implemented during the implementation period. Adherence scores of MORE lesson implementation for each treatment teacher was computed by taking the number of each lesson type that teachers reported teaching (e.g., the number of concept mapping lessons taught) divided by the total number of lessons to teach during the implementation period (e.g., five concept mapping lessons). Overall, the average adherence rates ranged between 87% and 94%, which is comparable to adherence levels from the previous studies of MORE (cf. Kim et. al., 2021a; 2021b). In addition, teachers were asked to report whether they used the MORE books recommended for the interactive read-aloud lessons. On average, nearly 88% of teachers reported using these books during the implementation period.

Students' Exposure to ELA and Science Instructional Time. Both treatment- and control-group Grade 2 teachers reported the amount of instructional time that they spent on ELA and science content each week over the implementation period (late October through late January). As shown in Table 4, treatment and control teachers devoted an average of 678.04 and 673.74 minutes per week, respectively, on ELA/reading instruction, and 104.53 and 108.04 minutes per week, respectively, on science instruction, though neither difference was statistically significant ($p > .05$). This result suggests that the treatment teachers, though asked to take on 45 additional lessons, did not do so at the expense of instructional time dedicated to the district's ELA curriculum.

Framework for Measuring Transfer Effects on Grade 2 Science Content Comprehension. The final aim of this study was to understand whether a sustained content literacy intervention could help students leverage and transfer newly acquired knowledge while reading about new, schema-related topics in science. Importantly, the issue of whether schema activation supports transfer in reading comprehension involves questions about what is transferred (the content) and from where to where (the context). Therefore, we sought to develop more sensitive measures for the final, end of Grade 2 domain-specific (science) reading comprehension along a continuum of transfer. Consistent with Barnett and Ceci (2002), we developed a framework for assessing transfer displayed in Table 5, which includes two dimensions of transfer. First, the content of the near-, mid-, and far-transfer passages provided extensive, limited, and no exposure, respectively, to the domain-specific vocabulary networks that children learned in the lessons. In other words, the content of the passages in the near- and mid-passages were designed to activate the schema children learned whereas the far transfer passage provided no cues for schema activation. Thus, the content dimension captures the extent to which students were cued to instantiate the schema underlying the passages (i.e., via extensive or limited exposure to vocabulary networks) or had to spontaneously recall the schema (i.e., no exposure to vocabulary networks).

Second, the context dimension captures the extent to which the passages are similar to the thematic lessons that were taught during content literacy instruction. The context for near transfer was similar to the instructional context whereas the mid and far transfer passages were different. In essence, if children's exposure to topic-specific schemas served to instantiate a general schema for problem-solving in children's minds, then the likelihood for impact on mid transfer, or possibly far transfer, for comprehension should be enhanced.

Student Measures

Domain-General Reading Comprehension and Vocabulary

The MAP reading assessment (Northwest Evaluation Association, 2011) was used to assess students' domain-general reading comprehension and vocabulary. The MAP is a vertically scaled and computer-adaptive reading assessment that measures student growth in reading comprehension from kindergarten to second grade using the Rasch unit (RIT) scale (Northwest Evaluation Association, 2011). The MAP reading comprehension subtest assesses students' ability to understand what they can read independently and their ability to make inferences, cite evidence from text, and understand main ideas in both narrative and informational texts. The MAP reading vocabulary subtest assesses students' ability to determine the meaning of words in context, analyze word parts, and understand figurative language. The test-retest reliabilities of an overall RIT total reading and subtests of reading comprehension and vocabulary ranged from .89 to .96 (Brown & Coughlin, 2007). The MAP reading assessments were administered at three-time points: winter of Grade 1 (January), spring of Grade 1 (April), and fall of Grade 2 (September).

Science Content Reading Comprehension of Near-, Mid-, and Far-Transfer Passages

We created a transfer measure of science content reading comprehension to assess Grade 2 students' ability to comprehend texts about schema-related topics. Specifically, the assessment included three passages designed to assess near, mid, and far transfer from the MORE curriculum by including progressively fewer explicitly taught curricular vocabulary words. The comprehension assessment rested on the construction-integration model of reading comprehension (Kintsch, 1993; 2009), which requires literal understanding of the propositions in text and integration with textual evidence with prior topic and vocabulary knowledge. Each of the three passages was followed by seven comprehension items (21 total) and one item (the fourth item) on the far-transfer passage was removed due to poor item functioning in item response theory (IRT) analysis. The items were designed to assess a student's ability to identify the main idea, determine the meaning of words or phrases in relevant text, describe scientific concepts, and integrate knowledge and ideas. The transfer assessment is a 20-item, three-option, untimed multiple-choice test that measures reading comprehension in science. Students read the passages silently and answered the multiple-choice questions in paper-pencil format. The assessment consisted of three passages and were equalized on Lexile measures (near = 640L, mid = 640L, far = 620L), mean sentence length (near = 9.0, mid = 8.2, far = 9.5), mean log word frequency (near = 3.45, mid = 3.28, far = 3.63), and word count (near = 180, mid = 164, far = 163).

We used two-parameter logistic (2PL) IRT models to estimate item difficulty and discrimination parameters. Using classical test statistics, the final 20-item assessment had Cronbach's alpha of .78. Descriptive statistics indicated that students scored 10.73 items correctly (SD = 4.88). IRT item difficulties ranged from -1.71 to 1.68, discriminations ranged from 0.47 to 2.46 (M = 1.05), and exploratory factor analysis provided strong evidence for unidimensionality on the full 20-item assessment, as the first factor explained most of the total variance. The Supplementary Materials S-D includes detailed information on the passage and items, the IRT model results, and the factor analysis.

To prepare the scores for the impact analysis, we created both total sum scores and 2-PL and 3-PL IRT scaled scores. The correlations between the total sum score and the 2-PL and 3-PL scores were .98 and .97 respectively. Given strong correlations among the scores, we used the standardized score for all analyses.

Networks of Vocabulary Knowledge Depth in Grade 2 Science

We used the semantic association task (Read, 1998; 2004) to assess students' vocabulary knowledge depth of taught words in science thematic units and their ability to identify semantical relations of the taught words with other words that were not explicitly targeted in the units (Collins & Loftus, 1975; Schmitt, 2014; Stahl & Fairbanks, 1986). The seven taught words were pre-selected words that treatment teachers directly taught via the concept mapping activity involving the structured word inquiry and argumentative writing activities. The five untaught words were not explicitly taught during the activities but were semantically associated with the taught words and were incidentally encountered through reading, listening, and discussion activities. In the semantic association task, there were four-word options for each of the 12 items that presented a target (taught or untaught) word and students were prompted to "circle two words that go with" the target word. Each item was scored 0 to 4 (see Supplementary Materials S-E for the test items and scoring system). Internal consistency (Cronbach's α) was .91.

Data Analysis

Analysis for Research Question 1: Effects of Grade 1 MORE Intervention on Domain-General Reading Comprehension and Vocabulary Growth

To address our first research question, we specified a multilevel piecewise linear growth curve modeling (Rabe-Hesketh & Skrondal, 2021) to examine whether implementation of the MORE Grade 1 thematic content literacy lessons followed by wide reading of thematically related texts in summer improved growth in students' domain-general reading comprehension and vocabulary (measured by MAP assessment). We used a piecewise linear growth model as we expected the slope for the treatment effect in the interval between the baseline (January 2019) and Time 1 (April 2019) in Grade 1 would differ from that in the interval from the baseline through Time 2 (September 2019) in Grade 2.

We specified three models for the following outcome measures: domain-general reading comprehension (Model A), vocabulary (Model B), and overall MAP reading score of domain-general reading comprehension and vocabulary (Model C). Each model was fitted as follows:

Level 1 (time): $Y_{tjk} = \pi_{0jk} + \pi_{1jk}\text{TIME1}_t + \pi_{2jk}\text{TIME2}_t + \varepsilon_{tjk}$

Level 2 (student): $\pi_{0jk} = \gamma_{00k} + \gamma_{01k}\text{PRETEST\_MATH}_{jk} + \sum_{q=2}^{13} \gamma_{0qk}\,\text{COV}_{jk} + \zeta_{0jk} + \zeta_{01k}\text{TIME}$

$\pi_{1jk} = \gamma_{10k}$

$\pi_{2jk} = \gamma_{20k}$

Level 3 (school): $\gamma_{00k} = \beta_{000} + \beta_{001}\text{MORE}_k + \sum_{w=2}^{7} \beta_{00w}\text{BLOCK}_k + \nu_{00k}$

$\gamma_{01k} = \beta_{010}$

$\gamma_{0qk} = \beta_{0q0}$

$\gamma_{10k} = \beta_{100} + \beta_{101}\text{MORE}_k$

$\gamma_{20k} = \beta_{200} + \beta_{201}\text{MORE}_k$

The composite model is written as follows:

$$Y_{tjk} = \beta_{000} + \beta_{100}\text{TIME1}_t + \beta_{200}\text{TIME2}_t + \beta_{001}\text{MORE}_k + \beta_{101}(\text{MORE}_k \times \text{TIME1}_t) + \beta_{201}(\text{MORE}_k \times \text{TIME2}_t) + \beta_{010}\text{PRETEST\_MATH}_{jk} + \sum_{q=2}^{13} \beta_{0q0}\,\text{COV}_{jk} + \sum_{w=2}^{7} \beta_{00w}\text{BLOCK}_k + \varepsilon_{tjk} + \zeta_{0jk} + \zeta_{01k}\text{TIME} + \nu_{00k}$$

$$\varepsilon_{tjk} \sim N\,(0, \sigma_\varepsilon^2),\ \begin{bmatrix}\zeta_{0jk}\\ \zeta_{01k}\end{bmatrix} \sim N\left(\begin{bmatrix}0\\0\end{bmatrix}\begin{bmatrix}\tau_{00} & \tau_{01}\\ \tau_{10} & \tau_{11}\end{bmatrix}\right),\ \nu_{00k} \sim N\,(0, \tau_\beta),$$

where Y_{tjk} represents the respective score at time t for student j in school k. The parameter β_{000} represents the mean initial status for control students. The parameters β_{100} and β_{200} represent the control group students' mean score differences between the baseline (Grade 1 January 2019) and Time 1 (Grade 1 April 2019) and between the baseline and Time 2 (Grade 2 September 2019), respectively. The parameter β_{001} refers to the treatment-control difference (0 = control; 1 = treatment) at the baseline; and β_{101} and β_{201} represent the treatment-control differences in the rate of growth between Time 0 and Time 1 and between Time 0 and Time 2, respectively. The parameters β_{010}, β_{0q0}, and β_{00w} represent the effects of the MAP mathematics pretest score (PRETEST_MATH), student-level demographic covariates (COV; q = 2, …, 13), and dummy-coded randomization blocks (BLOCK; w = 2, …, 7), respectively. The model also includes level-1, -2, and -3 residuals representing the random error ε_{tjk}; the student random intercept, ζ_{0jk}; the school random intercept, ν_{00k}; and a level-2 random effect for rate of change, ζ_{01k}, assumed to be normally distributed with mean 0 and uncorrelated with the predictors.

Analysis for Research Question 2: Effects of Sustaining MORE Intervention on Grade 2 Science Content Reading Comprehension and Vocabulary Knowledge Depth

To address the second research question about the main treatment effect on Grade 2 science content reading comprehension and vocabulary knowledge depth outcomes, we specified a series of two-level hierarchical linear models (HLM; Raudenbush & Bryk, 2002) for the analysis of cluster RCT to account for the nested structure of the data (i.e., students nested within schools and schools within school-level randomization blocks). A full combined model is specified as follows:

$$Y_{ij} = \gamma_{00} + \gamma_{01}\text{MORE}_j + \sum_{p=1}^{2} \gamma_{pj}\,\text{PRETEST}_i + \sum_{q=3}^{14} \gamma_{q0}\,\text{COV}_i + \sum_{w=2}^{7} \gamma_{0w}\text{BLOCK}_j + \varepsilon_{ij} + \zeta_{0j},$$

$$\varepsilon_{ij} \sim N\,(0, \sigma^2) \text{ and } \zeta_{0j} \sim N\,(0, \tau_{00}),$$

where Y_{ij} represents the posttest score of science content reading comprehension (overall scores and scores for near-, mid-, and far-transfer passages) and science vocabulary knowledge depth (overall, taught, and untaught vocabulary) assessments of student *i* in school *j*. The parameters γ_{00} and γ_{01} represent the intercept and the MORE treatment effect, respectively. The parameters γ_{pj}, γ_{q0}, and γ_{0w} represent the effects of the MAP reading and mathematics pretest scores (p = 1 and 2), demographic covariates (q = 3, …, 14), and randomization blocks (w = 2, …, 7), respectively. The model includes level-1 and level-2 residuals, ε_{ij} and ζ_{0j}, respectively, assumed to be normally distributed with mean zero and uncorrelated with the predictors.

Analysis for Research Question 3: Transfer Effects on Grade 2 Science Content Reading Comprehension by Item-Passage Type

To examine whether treatment effects on the science content reading comprehension outcome differed by passage item type, we fit an explanatory item response model (de Boeck, 2008; Wilson & de Boeck, 2004) to examine item-specific transfer effects. We specified a cross-classified multilevel logistic regression model to accommodate the nature of the cross-classified data structure, in which students were given all subtest passage items (i.e., near, mid, and far transfer). Level 1 was specified to estimate the predicted probability of a correct response (0 = incorrect, 1 = correct) as a function of cross-classification between passage item types and students and the effects of within-school covariates. Level 3 included the treatment effect and randomization blocks. The full model is expressed as follows:

Level 1: $$\eta_{(ij)k} = \text{logit}\left(\varphi_{(ij)k}\right) = \gamma_{ik} + \pi_{jk}$$

Level 2 (passage item): $$\gamma_{ik} = \psi_{1k}\text{MID}_i + \psi_{2k}\text{FAR}_i + \zeta_{ik}$$

Level 2 (student): $$\pi_{jk} = \gamma_{0k} + \sum_{p=1}^{2} \beta_{pk}\text{PRETEST}_j + \sum_{q=3}^{14} \beta_{qk}\text{COV}_j + \epsilon_{jk}$$

Level 3 (school): $$\psi_{1k} = \beta_{10} + \beta_{11}\text{MORE}_k$$

$$\psi_{2k} = \beta_{20} + \beta_{21}\text{MORE}_k$$

$$\gamma_{0k} = \beta_{00} + \beta_{01}\text{MORE}_k + \sum_{w=2}^{7} \beta_{0w}\text{BLOCK}_k + v_{0k}$$

Full combined model:

$$\eta_{(ij)k} = \text{logit}\left(\varphi_{(ij)k}\right) = \beta_{00} + \beta_{10}\text{MID}_i + \beta_{20}\text{FAR}_i + \text{MORE}_k \times (\beta_{11}\text{MID}_i + \beta_{21}\text{FAR}_i) +$$

$$\beta_{01}\text{MORE}_k + \sum_{p=1}^{2} \beta_{pk}\text{PRETEST}_j + \sum_{q=3}^{14} \beta_{qk}\text{COV}_j + \sum_{w=2}^{7} \beta_{0w}\text{BLOC}K_k + \zeta_{ik} + \epsilon_{jk} + v_{0k},$$

$$\text{Y}_{(ij)k} \sim \text{Bernoulli}\ (\varphi_{(ij)k})$$

$$\zeta_{ik} \sim \text{N}\ (0, \tau_\gamma),\ \epsilon_{jk} \sim \text{N}\ (0, \tau_\pi),\ \nu_{0k} \sim \text{N}\ (0, \tau_{00}),$$

where the predicted values for $\eta_{(ij)k}$ is the log-odds (logit) of a correct response of subtest passage item *i* for student *j* and school *k.* The parameter β_{00} represents the average expected log-odds of a correct response for near-transfer passage items among students in the control group in the reference randomization block. The parameters β_{10} and β_{20} represent the average difference in the log-odds of a correct response between near- and mid-transfer passage items and between near- and far-transfer passage items, respectively. The parameter β_{01} represents the main effect of MORE treatment on near-transfer passage items. The parameters β_{11} and β_{21} represent the interaction effects between MORE treatment and mid-transfer passage items and between MORE and far-transfer passage items, respectively. The parameters β_{pk}, β_{qk}, and β_{0w} represent the effects of the baseline MAP reading and mathematics scores (p = 1 and 2), demographic covariates (q = 3, …, 14), and randomization blocks (w = 2, …, 7), respectively. The parameters ζ_{ik}, ϵ_{jk}, and ν_{0k} denote random effects for passage items type (level 2), students (level 2), and schools (level 3), respectively.

Sensitivity Analyses

Missing values existed for baseline MAP reading (3.5%) and mathematics (3.7%) scores. With the assumption that data were missing at random (MAR), we used multiple imputation by simulating 10 data sets with plausible values in place of missing observations as a sensitivity check for the main effect models (research question 2). The procedure was performed using a Markov Chain Monte Carlo (MCMC) algorithm with a multivariate normal distribution using the *mi impute mvn* command in Stata.

Results

Descriptive and Correlational Analyses

Table 6 displays descriptive statistics of key pretest and posttest measures by treatment condition and a pairwise correlation matrix. Descriptive statistics for the MAP total reading RIT scaled scores were below the national norm in the winter and spring of Grade 1 for both conditions (Thum et al., 2015). In general, mean MAP reading scaled scores also increased from the winter to spring of Grade 1 and then remained flat from spring Grade 1 to fall Grade 2. Furthermore, the pairwise correlational analysis results indicated that the MAP reading scores at Time 0, 1, and 2 were positively and strongly correlated with a total score of science content reading comprehension (r range = .60 to .65) and moderately correlated with a total score of vocabulary knowledge depth (r range = .55 to .59). Similarly, there was a moderately strong correlation between the total score on the science content reading comprehension and vocabulary knowledge depth (r = .57).

Research Question 1: Effects of Grade 1 MORE Intervention on Domain-General Reading Comprehension and Vocabulary Growth

The first research question examined the piecewise linear growth trajectories of domain-general reading comprehension, vocabulary, and overall MAP reading outcomes of the treatment and control group students. Table 7 displays the results of the fitted three-level piecewise linear growth curve models for the three outcomes. For the domain-general reading comprehension outcome (Model A), the treatment group students started out at a significantly lower level of domain-general reading comprehension than the treatment group students at the baseline (β_{001} = -1.78, $p < .01$), accounting for all covariates. However, treatment group students demonstrated a marginally faster rate of growth between the baseline and Time 1 (Grade 1 April) (β_{101} = 0.90; $p < .10$) and a statistically significantly faster rate of growth between the baseline and Time 2 (Grade 2 September) (β_{201} = 1.89; $p < .001$) relative to the control group students. Results from the piecewise growth models for vocabulary (Model B) and MAP reading total score (Model C) revealed no statistically significant interactions between treatment and time.

Figure 3 visually displays the difference in fitted trajectories of MAP reading comprehension between the treatment and control group students based on the piecewise linear growth modeling. The difference between the two groups with a higher average score of the control group at the baseline (Time 0) was gradually narrowed during the MORE implementation period in the spring semester and over the summer due to the treatment group students' accelerating growth over time. As a result, there was no statistically significant difference between the two groups at Time 2.

Research Question 2: Effects of Sustaining MORE Intervention on Grade 2 Science Content Reading Comprehension and Vocabulary Knowledge Depth

The second research question was to investigate the treatment effect of sustaining MORE thematic content literacy lessons on Grade 2 students' (a) science content reading comprehension and (b) science vocabulary knowledge depth outcomes. Table 8 shows the results of the treatment effects on both sets of outcomes. We found statistically significant treatment effects on the overall science content reading comprehension test (ES = .18, $p < .01$), after accounting for the control variables. When the treatment effects by subtests were examined, the treatment effect were largest for the near-transfer subtest (ES = .22, $p < .001$), moderately sized on the mid-transfer subtest (ES = .17, $p < .01$), and smallest on the far-transfer subtest (ES = .04, $p > .05$). We found statistically non-significant (*ps* > .05) treatment effects on overall science vocabulary knowledge depth test (ES = .05) and the subtests with taught words (ES = -.01) and untaught words (ES = .11). There was also no evidence of treatment effect that varied by students' initial domain-general reading comprehension levels (see the moderation analysis results in the online supplemental material).

Research Question 3: **Transfer Effects on Grade 2 Science Content Reading Comprehension by Item-Passage Type**

The third research question was to examine whether the magnitude of transfer effects of sustaining Grade 2 MORE thematic content literacy lessons on Grade 2 science content reading comprehension differed by item-passage type. Table 9 shows the results of the explanatory item response model specified as a cross-classified multilevel logistic regression model to estimate item-passage-specific transfer effects. We found that the treatment condition was associated with a higher log-odds of a correct response ($\beta_{01} = 0.31$, $SE = 0.07$, $p < .001$) for a near-transfer item-passage type, indicating that the expected odds of a correct response among the treatment group students were $e^{0.31} = 1.36$ times the odds of a correct response among the control group students, holding other covariates and random effects constant. The interaction effect between treatment and mid-transfer item passage was statistically significant ($\beta_{11} = -0.12$, $SE = 0.05$, $p < .05$), suggesting that the treatment effect was slightly smaller on the mid-transfer item-passage type than the near-transfer item-passage type. Likewise, the interaction effect between treatment and far-transfer passage items was also statistically significant (β_{21}= -0.26, *SE* = 0.05, *p* < .001), indicating that the treatment effect was much smaller on the far-transfer passage item type than the near-transfer one. Post-hoc analysis also revealed that the treatment effect was statistically significantly higher on the mid-transfer than far-transfer passage item types (Difference = 0.14 logits, *SE* = 0.05, *p* < .05).

Figure 4 visually displays the model-implied treatment effects in percentage points compared to control students with a 50% correct response probability to treatment group students on each passage item type. The fitted bar graphs indicate that, holding constant all covariates in the model, the treatment group had a significantly larger advantage on the probability of correctly answering both near-transfer passage items (7.7 percentage points) and mid-transfer (4.9 percentage points) but no significant advantage on the far-transfer passage items.

Sensitivity Analyses

We report the results of sensitivity analyses for research question 2 (treatment effects on Grade 2 science content reading comprehension and vocabulary knowledge depth) based on multiple imputation in the online supplementary material. Overall, the results of sensitivity analyses replicated findings for Research Question 2 and support the robustness of the treatment effects.

Discussion

Using a longitudinal cluster randomized trial design involving 30 elementary schools, we examined the effectiveness of a sustained content literacy intervention that aligned thematic content literacy lessons across grades and promoted wide reading of thematically related texts. The MORE content literacy intervention was anchored in research underscoring the role of prior domain and topic knowledge schemas in supporting students’ ability to read informational texts, both in school lessons and at-home summer reading activities (Anderson & Freebody, 1981; Duke, 2000; Kim et al., 2019; Kintsch, 1998; McCombs et al., 2019; O'Reilly et al., 2019). Following the implementation of the Grade 1 thematic content literacy lessons and summer support for reading, treatment group children enjoyed a faster rate of growth on a measure of domain-general reading comprehension from the winter of Grade 1 to the fall of Grade 2 relative to control group children. Sustaining the thematic science lessons in Grade 2 improved students' performance on science content reading comprehension. In particular, transfer effects were observed on both the near- and mid-transfer passages that contained semantically related, directly taught science vocabulary. The findings support the theoretical proposition that sustaining and aligning content and practices across grades can improve students' domain knowledge and, ultimately, their reading comprehension outcomes (Anderson & Freebody, 1981; Kintsch, 2009; Stahl & Fairbanks, 1986). We discuss findings related to each of the three research questions and conclude with study limitations and implications.

Sustaining and Aligning Content and Practices Can Improve Domain-General Reading Comprehension

Our first aim was to examine whether the implementation of Grade 1 thematic science content literacy instruction followed by wide reading of thematically related texts during summer could promote transfer on a domain-general measure of reading comprehension. Overall, results from the piecewise linear growth models indicated that the implementation of Grade 1 content literacy lessons followed by wide reading of thematically related informational texts enabled treatment group children to enjoy larger gains in domain-general reading compared to the control group. How do we explain these results?

Results from the Grade 1 fidelity of implementation analyses indicate that teachers were able to adhere closely to the lesson scripts, implement more open literacy tasks, and spend more time on content-area instruction. In addition, the results from surveys administered in the fall of Grade 2 indicated that treatment group students reported reading an average of six more print and digital texts than did control group students. Although self-report data are vulnerable to over-reporting bias, the average number of print books read by rising second-graders in our study is consistent with the average of five print books reported in previous research on summer reading in the early elementary grades (Kim, 2007; Phillips & Chin, 2004), while also suggesting that digital reading may be another medium for increasing print exposure (Furenes et al., 2021).

Some might argue that the treatment group children's additional reading volume in the summer alone could explain the positive effect on outcomes. However, the results from the piecewise growth model clearly indicate that the additional summer reading was necessary but not sufficient to mitigate the loss in domain-general reading in the summer months. Outcomes were also attributable to school year, thematic content literacy instruction, which preceded the summer reading opportunities. Overall, then, our findings suggest that building knowledge of topic-specific schemas in both Grade 1 science and social studies lessons, and then scaffolding wide reading through choice of grade-level and reader-level matched informational books during summer, were both critical levers for impacting a domain-general reading comprehension.

Students domain and topic knowledge development relies extensively on exposure to informational texts (Best et al., 2008; Smith et al., 2021). Accordingly, early grade content literacy instruction that focuses on building topic-specific schemas through deep and focused instruction during school followed by wide reading of thematically related science and social studies texts during summer may work synergistically. That is, sustaining and aligning the informational texts that children read at school and home may work together to increase the amount of reading and build the background knowledge needed to read informational texts (Cunningham & Stanovich, 1991; Hirsch, 2016; Mol & Bus, 2011). Results from our piecewise growth models indicate that giving children opportunities to read thematically related informational texts at home that are tightly coupled to topics learned in school can mitigate summer reading loss.

Building Schemas to Improve Domain-Specific Reading Comprehension for All Learners

Our second research aim was to examine whether building a general schema that was infused in the instructional content and teacher practices across two grades could improve *all* second graders' domain-specific reading comprehension. We designed the MORE thematic science lessons to provide students with continual exposures to networks of domain-specific vocabulary words that were undergirded by the general schema. Findings further underscore the value of embedding domain-specific vocabulary networks in thematic lessons to instantiate schemas that support language and literacy activities for all students during Tier I instruction (Bolger et al., 2008; Graves, 2015; Kim et al., 2021a, 2021b; Neuman et al., 2011). There is growing evidence that vocabulary instruction alone is necessary but not sufficient to help students build the schemas and depth of vocabulary knowledge for reading challenging informational texts (Apthorp et al., 2012; Elleman et al., 2009; Pressley et al., 2007; Wright & Cervetti, 2016). In other words, thematic approaches to content literacy instruction may help children develop high quality lexical representations of semantically related words in networks, thus facilitating efficient retrieval and application of those word networks during text processing (Perfetti & Stafura, 2014).

In particular, children were gradually exposed to related domain-specific academic vocabulary words in Grade 1 and 2 content literacy instruction that helped to reduce cognitive load for learning those words (Sweller, 2011). Based on data from Grade 2 teacher surveys, it is also clear that treatment group teachers adhered closely to the MORE lesson sequence and used the read-aloud books that provided recurrent exposure to semantically related vocabulary. Because there was no difference in the amount of time teachers spent on content literacy instruction, we can also rule out additional time as a factor that explains the impact on domain-specific reading comprehension. Importantly, the Grade 2 lessons provided students with opportunities to learn how to apply networks of domain-specific academic words through listening, reading, writing, and discussion activities that focused on topic-specific schemas that were tightly connected to the general schema. The development of high-quality lexical representations may require sustained interventions that incrementally build a learner's knowledge of word meanings and gradually situates new word meanings in an existing network (Perfetti, 2007). In this study, students were given recurrent exposures to domain-specific vocabulary networks learned in Grade 1 thematic content literacy lessons about animal survival that laid the foundation for the Grade 2 thematic content literacy lessons about how paleontologists study dinosaurs. Therefore, sustained content literacy interventions that gradually build thematic connections across grades can help students connect new learning to a general schema and transfer knowledge to novel, schema-related passages.

Importantly, our approach to content literacy improved outcomes for all learners. There was no evidence of aptitude by treatment interaction effects (Cronbach, 1975bg) whereby initially high-performing students reap greater gains on the Grade 2 transfer measures of reading than initially low-performing students (see Supplementary Materials S-F). In the current debates over the science of reading (Goodwin & Jimenez, 2021), there is an emerging consensus that improving the quantity and quality of Tier I content area instruction can help to build the domain knowledge schemas that all children need to read informational text with greater understanding (Cabell & Hwang, 2020; Duke et al., 2021). Given that Black students and Latinx students and students from low and middle socioeconomic status families comprised a majority of our sample, our results further underscore how large-scale implementation of the MORE intervention can support national efforts to equitably improve outcomes for all learners. Unfortunately, however, there is growing evidence that the overemphasis on English language arts and math instruction has restricted students' opportunities to acquire science and social studies content and exacerbated academic achievement disparities based on students' ethnicity and family income (Arold & Shakeel, 2021; Blank, 2012). To address these unintended negative consequences, our findings highlight the importance of providing evidence-based Tier I content literacy instruction to all learners.

Reconceptualizing Transfer

Our third research aim was to examine how much transfer occurred on the final Grade 2 science content reading comprehension outcome. Our findings highlight the importance of promoting and measuring transfer along a continuum. In developing our transfer passages and items, we systematically varied two dimensions along a continuum. First, the content dimension varied the number of domain-specific vocabulary words such that the near, mid, and far passages provided children with full, extensive, and no exposures to taught words. Second, the context dimension varied the degree to which the passage was similar or different from the instructional context, which focused on the topic of how paleontologists study dinosaurs. Compared to control group students, MORE students performed better on the near and mid-transfer items drawn from the passages that included networks of domain-specific words that were directly taught by teachers in the MORE Grade 2 lessons.

Notably, however, there were no significant effects on the far-transfer items drawn from the passage that provided no exposure to the vocabulary network and that was different from the instructional context. More broadly, our findings support the theoretical proposition that thematic content literacy lessons can build a novice learner's schema, but spontaneous retrieval and application of that schema in understanding far-transfer passages may require more extensive practice and experience (Barnett & Ceci, 2002; Ericsson & Kintsch, 1995; Kimball & Holyoak, 2000; Mayer & Wittrock, 1996; Salomon & Perkins, 1989). Our findings are also consistent with research on other large-scale literacy programs that have not demonstrated impact on far-transfer measures of reading comprehension (Pearson et al., 2020; Quint et al., 2015). Finally, the use of cross-classified multilevel logistic regression models (Compton & Pearson, 2016; Francis et al., 2018) enabled us to pinpoint "pressure points" and mechanisms that promote transfer effects in reading comprehension, including the degree to which the transfer passages included directly taught words and described scenarios that were similar to the instructional context.

Study Limitations and Future Research

Several limitations should be addressed in future research. Given the impact of COVID-19 on our original research plan, we were unable to continue the longitudinal growth model analysis with the domain-general reading comprehension measure or to assess social studies reading comprehension in the spring of Grade 2. As a result, it is unclear whether such an approach would promote broader gains on domain-general reading comprehension outcomes through the end of Grade 2. An in-depth focus on a single schema across two grades may also have opportunity costs by curtailing opportunities to read about a wide variety of topics in multiple domains (e.g., science, social studies, literature). Keeping these limitations in mind, we encourage future replications to determine whether the full implementation of Grade 1 and Grade 2 science and social studies lessons promotes gains in reading comprehension (both domain-general reading and domain-specific reading in social studies) across a 12-month time period.

More pointedly, what would it take to help students transfer schemas learned in science to another domain like social studies? To date, early elementary grade content literacy instruction has focused on only science (Cervetti et al., 2012; Gray et al., 2021) or social studies (Duke et al., 2021; Williams et al., 2016). Furthermore, content literacy programs that target multiple domains have shown limited evidence of far transfer in reading in the early elementary grades (Connor et al., 2017). Because many of the Grade 2 words and concepts in the MORE science lessons appear across domains (e.g., evidence, theory), teachers could potentially interleave (Bjork, 1994; Snow et al., 2009) those concepts in topic-specific schemas that are directly taught in both science and social studies lessons. Future research should explore whether such efforts could improve reading outcomes simultaneously in two domains without having adverse effects on children's foundational word knowledge.

It is also beyond the scope of this study to pinpoint the precise mediating mechanisms driving the transfer effects on both domain-general and domain-specific reading comprehension outcomes. Models of reading comprehension (Cromley & Azevedo, 2007; Talwar et al., 2018) clearly suggest that vocabulary and background knowledge (e.g., science and social studies topics) are separable and malleable pathways through which content literacy ultimately improves reading comprehension. Thus, a crucial next step for future research is to employ structural equation modeling to probe whether gains in vocabulary knowledge depth and domain and topic knowledge are critical mediating mechanisms that drive improvement in student reading comprehension outcomes. It is clear that vocabulary knowledge, background knowledge, and print exposure are all active ingredients that are likely to foster students' ability to read for understanding, above and beyond word reading skills (Kim et al., 2016; Connor et al., 2017; Mol & Bus, 2011; Neuman et al., 2011; Shanahan et al., 2010). Thus, follow-up research should examine the mediating processes that underlie gains in reading comprehension.

Implications for Policymakers and Practitioners

The findings related to our three main research questions have direct implications for policymakers and practitioners. First, our findings highlight the core components that reduced summer loss in reading comprehension. The thematic content literacy lessons and materials, the alignment of content across school year and summer, and the stimulation of wide reading at home are all distinctive core components of MORE. Alignment of content materials across school year and summer further strengthens the possibilities for enhanced learning of specific topics within and across domains. For example, in making choices about which books to make available to students over the summer – and how to scaffold students' engagement with those books (Kim & White, 2008; Stein, 2017) – educators might start by identifying the topic-specific schemas covered during the school year and give students opportunities to read both grade- and reader-level matched texts during the summer. Although there are some efforts to strengthen home-school connections around science book reading activities (e.g., Mantzicopoulos et al., 2013), school-year content literacy instruction is largely de-coupled from home-based summer reading interventions. Our approach to sustaining MORE Grade 1 lessons through the summer with opportunities to read thematically related texts may be more cost-effective and easier to scale than either school-year content literacy instruction (Pearson et al., 2020) or classroom-based summer programs (McCombs et al., 2020). Our findings also replicate and extend previous studies indicating that both teacher-directed instruction and wide reading of thematically-related texts enhance growth in students' domain-general reading (Hwang et al., 2021; Kim et al., 2016).

Second, our findings highlight how teachers can interleave practices to build schemas and meta-linguistic skills to help students read novel informational texts. For example, teachers might begin their lessons with the general schema they want children to acquire as they study and read about related topics. During the Grade 2 MORE science lessons, for example, a teacher could introduce the general schema for *how scientists study past events* and the common goals, the resources, and the constraints that scientists confront as they explore this question. The schema might then be written in abbreviated form on a chart. Then, while reading aloud from the informational texts like *T. Rex: Hunter or Scavenger* (Holtz, 2003), a teacher could make explicit connections between parts of the general schema and the specific topic of how paleontologists study dinosaur mass extinctions. While reading from the chapter on "The Science of Paleontology," a teacher could explain that "paleontologists" are scientists who study "fossil remains of ancient creates" and "all scientists use observations about the world around us to try and answer questions" (Holtz, 2003, p. 16). Next, a teacher could also pause and reflect on semantically related words in the text T. Rex and build students' meta-linguistic skills—that is, the ability to manipulate the structural features of language (McKeown et al., 2017; Nagy, 2005; Tunmer et al., 1988). In the MORE lessons, teachers lead students in an investigation of the word "paleontologists," with subsequent discussion of the meaning of derivational morphemes *paleo-, -ist*, and the inflectional morpheme *-s*. By interleaving practices to build a general schema and simultaneously teach key words and morphemes, teachers can help students read about topics that are unified by a generalizable schema and include similar domain-specific vocabulary (e.g., mid-transfer passage on how archaeologists study Pompei). In essence, teachers can build general schemas and knowledge of semantically related words that give their students a leg up when reading new informational texts.

Finally, to assess transfer in reading comprehension, policymakers should not only test students' mastery of discrete literacy skills, but also their mastery of specific topics learned in school. According to the recent synthesis of findings from the Reading for Understanding Initiative, Pearson and colleagues (2020) underscored the need to address the question of transfer in reading comprehension by developing "better approaches to scaling and describing the degree of alignment between assessments and interventions" (p. 276). In developing our Grade 2 science assessments, we wanted to move as far away as possible from the typical practice of using proximal measures that are too close to intervention activities and distal measures that are too far from intervention actives to be sensitive to improvements in instruction. Importantly, claims about intervention effectiveness depend crucially on our answer to the question: how far do intervention effects travel? To address this question, policymakers could develop and use near-, mid-, and far-transfer passages that vary the content (the number of domain-specific words taught in lessons) and the context (the distance from the instructional context) to measure transfer on domain-specific reading comprehension outcomes. In short, it is critical to measure students' mastery of the topics that are taught during classroom instruction and their ability to transfer knowledge to new passages.

Conclusion

This longitudinal randomized trial of MORE represents a novel approach to promoting and measuring transfer in reading by (a) tightly coupling school year content literacy instruction with wide reading of informational texts during the summer months, (b) simultaneously building readers' schemas and meta-linguistic skills during thematic science lessons, and (c) assessing student mastery of topics learned in school on near, mid, far transfer passages. To date, large-scale literacy interventions have demonstrated mixed success in improving elementary grade students' reading comprehension outcomes (Pearson et al., 2020; Quint et al., 2015). A sustained content literacy intervention may represent an innovative and replicable approach for improving all learners' ability to read challenging informational texts with greater understanding.

References

Adams, B. C., Bell, L. S., & Perfetti, C. A. (1995). A trading relationship between reading skill and domain knowledge in children's text comprehension. *Discourse Processes, 20*, 307–323.

Alexander, P. A. (1997). Mapping the multidimensional nature of domain learning: The interplay of cognitive, motivational, and strategic forces. In Maehr, M. L. & Pintrich, P. R. (Eds.), *Advances in motivation and achievement* (Vol. 10, pp. 213–250). Greenwich, CT: JAI Press.

Alexander, P. A. (2000). Research news and comment: Toward a model of academic development: schooling and the acquisition of knowledge. *Educational Researcher*, *29*(2), 28–44. https://doi.org/10.3102/0013189X029002028

Alexander, P. A. (2003). The development of expertise: The journey from acclimation to proficiency. *Educational Researcher, 32*(8), 10–14. https://doi.org/10.3102/0013189X032008010

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

Anderson, R. (1984). Some reflections on the acquisition of knowledge. *Educational Researcher, 13*, 5-10.

Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. T. Guthrie (Ed.), *Comprehension and teaching: Research reviews* (pp. 77–117). Newark, DE: International Reading Association.

Anderson, R.C., & Pearson, P.D. (1984). A schema-theoretic view of basic processes in reading. In R. Barr, M. L. Kamil, & Mosenthal (Ed.), *Handbook of Reading Research* (pp.255-291).

Apthorp, H., Randel, B., Cherasaro, T., Clark, T., McKeown, M., & Beck, I. (2012). Effects of a supplemental vocabulary program on word knowledge and passage comprehension. *Journal of Research on Educational Effectiveness, 5*(2), 160–188.

Arold, B. W., & Shakeel, M. (2021). The Unintended Effects of Common Core State Standards on Non-Targeted Subjects. Program on Education Policy and Governance Working Papers Series. PEPG 21-03. *Program on Education Policy and Governance*.

Ausubel, D. P. (1960). The use of advance organizers in the learning and retention of meaningful verbal material. *Journal of Educational Psychology*, *51*(5), 267–272. https://doi.org/10.1037/h0046669

Bailey, D., Duncan, G. J., Odgers, C. L., & Yu, W. (2017). Persistence and fadeout in the impacts of child and adolescent interventions. *Journal of research on educational effectiveness*, *10*(1), 7–39. https://doi.org/10.1080/19345747.2016.1232459

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin, 128*, 612-637.

Bartlett, F.C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge University Press.

Beck, I. L., Perfetti, C. A., & McKeown, M. G. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology, 74*(4), 506–521. https://doi.org/10.1037/0022-0663.74.4.506

Best, R. M., Floyd, R. G., & Mcnamara, D. S. (2008). Differential competencies contributing to children’s comprehension of narrative and expository texts. *Reading Psychology*, *29*(2), 137–164. https://doi.org/10.1080/02702710801963951

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe and A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp.185-205). Cambridge, MA: MIT Press.

Blank, R. K. (2013). Science instructional time is declining in elementary schools: What are the implications for student achievement and closing the gap?. *Science Education*, *97*(6), 830-847.

Bolger, D. J., Balass, M., Landen, E., & Perfetti, C. A. (2008). Contextual variation and definitions in learning the meaning of words. *Discourse Processes, 45*(2), 122–159. https://doi.org/10.1080/01638530701792826

Bowers, J. S., & Bowers, P. N. (2018). Progress in reading instruction requires a better understanding of the English spelling system. *Current Directions in Psychological Science*, *27*(6), 407–412. https://doi.org/10.1177/0963721418773749

Bowers, P. N., Kirby, J. R., & Deacon, S. H. (2010). The effects of morphological instruction on literacy skills: A systematic review of the literature. *Review of Educational Research*, *80*(2), 144–179. https://doi.org/10.3102/0034654309359353

Bronfenbrenner, U., & Morris, P. A. (2006). The bioecological model of human development. In W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology: Theoretical models of human development* (5th ed., Vol. 1, pp. 795–825).

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, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Mid-Atlantic; Retrieved from http://ies.ed.gov/ncee/edlabs

Cabell, S.Q., & Hwang, H. (2020). Building content knowledge to boost comprehension in the primary grades. *Reading Research Quarterly, 55*(1), 99-107. https://doi.org/10.1002/rrq.338

Carlisle, J. F., Fleming, J. E., & Gudbrandsen, B. (2000). Incidental word learning in science classes. *Contemporary Educational Psychology*, 25, 184-211.

Cervetti, G. N., Wright, T. S., & Hwang, H. (2016). Conceptual coherence, comprehension, and vocabulary acquisition: A knowledge effect? *Reading and Writing*, *29*(4), 761–779. https://doi.org/10.1007/s11145-016-9628-x

Cervetti, G. N., Barber, J., Dorph, R., Pearson, P. D., & Goldschmidt, P. (2012). The impact of an integrated approach to science and literacy in elementary school classrooms. *Journal of Research in Science Teaching, 49*, 631-658.

Cervetti, G. N., & Wright, T. S. (2020). The role of knowledge in understanding and learning from text.

Chilton, M. W., & Ehri, L. C. (2015). Vocabulary learning: Sentence contexts linked by events in scenarios facilitate third graders’ memory for verb meanings. *Reading Research Quarterly*, 50(4), 439-458. http://doi.org/10.1002/rrq.106

Colenbrander, D., Parsons, L., Bowers, J.S., & Davis, C.J. (2021). Assessing the effectiveness of structured word inquiry for students in grades 3 and 5 with reading and spelling difficulties: A randomized controlled trial. *Reading Research Quarterly*. Advance online publication. https://doi.org/10.1002/rrq.399

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82,* 407-428. http://dx.doi.org.10.1037/0033-295X.82.6.407

Compton, D. L., & Pearson, P. D. (2016). Identifying robust variations associated with reading comprehension skill: The search for pressure points. *Journal of Research on Educational Effectiveness*, *9*(2), 223–231. https://doi.org/10.1080/19345747.2016.1149007

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*(3), 301-320.

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.

Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist, 30*(2), 116-127.

Cunningham, A. E., & Stanovich, K. E. (1991). Tracking the unique effects of print exposure in children: Association with vocabulary, general knowledge, and spelling. *Journal of Educational Psychology, 83,* 264-274

D'Agostino, J. V., Rodgers, E., (2017). Literacy achievement trends at entry to first grade. *Educational Researcher, 46*(2), 78-89.

de Boeck, P. (2008). Random item IRT models. *Psychometrika, 73,* 533–559. https://doi.org/10.1007/s11336-008-9092-x

Duke, N. K. (2000). 3.6 minutes per day: The scarcity of informational texts in first grade. *Reading Research Quarterly*, *35*(2), 202–224. https://doi.org/10.1598/RRQ.35.2.1

Duke, N. K., Purcell-Gates, V., Hall, L. A., & Tower, C. (2006). Authentic literacy activities for developing comprehension and writing. *The Reading Teacher, 60*(4), 344–355. https://doi.org/10.1598/RT.60.4.4

Duke, N. K., Halvorsen, A.-L., Strachan, S. L., Kim, J., & Konstantopoulos, S. (2021). Putting PjBL to the test: The impact of project-based learning on second graders' social studies and literacy learning and motivation in low-ses school settings. *American Educational Research Journal*, *58*(1), 160–200. https://doi.org/10.3102/0002831220929638

Elleman, A. E., Lindo, E. J., Morphy, P., & Compton, D. L. (2009). The impact of vocabulary instruction on passage-level comprehension of school-age children: A meta-analysis. *Journal of Research on Educational Effectiveness, 2*, 1– 44. http://dx.doi.org/10.1080/19345740802539200

Engel, M., Claessens, A., & Finch, M. (2013). Teaching students what they already know? The (mis)alignment between mathematics instructional content and student knowledge in kindergarten. *Educational Evaluation and Policy Analysis, 35*(2), 157–178.

Ericsson, K. A. (2018). Superior working memory in experts. In K. A. Ericsson, R. R. Hoffman, A. Kozbelt, & A. M. Williams (Eds.), *Cambridge handbooks in psychology. The Cambridge handbook of expertise and expert performance* (p. 696–713). Cambridge University Press. https://doi.org/10.1017/9781316480748.036

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*(2), 211-245.

Ericsson, K. A. & Pool, R. (2016). *Peak: Secrets from the new science of expertise*. Boston: Houghton Mifflin Harcourt.

Francis, D. J., Kulesz, P. A., & Benoit, J. S. (2018). Extending the simple view of reading to account for variation within readers and across texts: The complete view of reading (CVR*i*). *Remedial and Special Education*, *39*(5), 274–288. https://doi.org/10.1177/0741932518772904

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*, 59-109.

Fryer, R. G., & Levitt, S. D. (2006). The black-white test score gap through third grade. *American Law and Economics Review, 8*(2), 249–281.

Furenes, M. I., Kucirkova, N., & Bus, A. G. (2021). A comparison of children's reading on paper versus screen: A meta-analysis. *Review of Educational Research*, *91*(4), 483–517. https://doi.org/10.3102/0034654321998074

Gelman, S. A. (2009). Learning from others: Children's construction of concepts. *Annual Review of Psychology, 60,* 115–140.

Georgiou, G.K., Savage, R., Dunn, K., Bowers, P., & Parrila, R. (2021). Examining the effects of Structured Word Inquiry on the reading and spelling skills of persistently poor grade 3 readers. *Journal of Research in Reading, 44*(1), 131– 153. https://doi.org/10.1111/1467-9817.12325

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15*(1), 1–38. https://doi.org/10.1016/0010-0285(83)90002-6

Goodwin, A. P., & Jimenez, R. T. (2021). The science of reading: Supports, critiques, and questions. *Reading Research Quarterly, 56*(S1), S7-S22.

Graves, M.F. (2015). Building a vocabulary program that really could make a significant contribution to students becoming college and career ready. In P.D. Pearson & E.H. Hiebert (Eds.), *Research-based practices for Common Core literacy* (pp. 123–142). New York, NY: Teachers College Press.

Gray, A. M., Sirinides, P. M., Fink, R. E., & Bowden, A. B. (2021). Integrating literacy and science instruction in kindergarten: Results from the efficacy study of zoology one. *Journal of Research on Educational Effectiveness*, *15*(1), 1–27. https://doi.org/10.1080/19345747.2021.1938313

Guthrie, J. T., & Klauda, S. L. (2014). Effects of classroom practices on reading comprehension, engagement, and motivations for adolescents. *Reading Research Quarterly, 49*, 387-416.

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*(4), 237–250.

Guthrie, J., & Wigfield, A. (2000). Engagement and motivation in reading. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Volume III, pp. 403–422

Guthrie, J. T., Wigfield, A., Humenick, N. M., Perencevich, K. C., Taboada, A., & Barbosa, P. (2006). Influences of stimulating tasks on reading motivation and comprehension. *The Journal of Educational Research*, *99*(4), 232–246. https://doi.org/10.3200/JOER.99.4.232-246

Guthrie, J. T., Wigfield, A., & You, W. (2012). Instructional contexts for engagement and achievement in reading. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engage- ment* (pp. 601–634). Boston, MA: Springer. http://dx.doi.org/10.1007/978-1-4614-2018-7_29

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

Hirsch, P. M. (1987), The other side of assessment. *New Directions for Community Colleges,* 15-24. https://doi.org/10.1002/cc.36819875904

Holtz, T. R. (2003). *T. Rex, Hunter or Scavenger?* New York, NY: Random House Children's Books.

Hwang, H., Cabell, S. Q., & Joyner, R. E. (2021). Effects of integrated literacy and content-area instruction on vocabulary and comprehension in the elementary years: A meta-analysis. *Scientific Studies of Reading*. Advance online publication. https://doi.org/10.1080/10888438.2021.1954005

Imai, M., Gentner, D., & Uchida, N. (1994). Children's theories of word meaning: The role of shape similarity in early acquisition. *Cognitive Development, 9*(1), 45-75.

Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, *331*(6018), 772–775. https://doi.org/10.1126/science.1199327

Kendeou, P. & O'Brien, E. J. (2016). Prior knowledge, acquisition and revision. In P. Afflerbach (Ed.), *Handbook of individual differences in reading: reader, text, and context* (pp. 151-163). New York, NY: Routledge Taylor & Francis.

Kim, J. S. (2007). The effects of a voluntary summer reading intervention on reading activities and reading achievement. *Journal of Educational Psychology, 99*(3), 505–515.doi:10.1037/0022-0663.99.3.505

Kim, J. S. (2021). "A Randomized Controlled Trial to Replicate and Personalize with Technology the Effects of a Model of Reading Engagement (MORE) on First- and Second-Graders' Science and Social Studies Domain Knowledge, Reading Engagement, Reading Comprehension, and Basic Literacy Skills ." AEA RCT Registry. February 06. https://doi.org/10.1257/rct.3489-1.2000000000000002

Kim, J. S., Asher, C. A., Burkhauser, M. A., Mesite, L., & Leyva, D. (2019). 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. *AERA Open, 5*(3), 1-18. DOI: 10.1177/2332858419872701

Kim, J. S., Burkhauser, M. A., Mesite, L. M., Asher, C. A., Relyea, J. E., Fitzgerald, J., & Elmore, J. (2021a). Improving reading comprehension, science domain knowledge, and reading engagement through a first-grade content literacy intervention. *Journal of Educational Psychology, 113*(1), 3–26. https://doi.org/10.1037/edu0000465

Kim, J. S., Burkhauser, M. A., Quinn, D. M., Guryan, J., Kingston, H. C., & Aleman, A. (2017). Effectiveness of structured teacher adaptations to an evidence-based summer literacy program. *Reading Research Quarterly, 52*(4), 443-467. doi:10.1002/rrq.178

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

Kim, J. S., & Quinn, D.M. (2013). The effects of summer reading on low-income children's literacy achievement from kindergarten to grade 8. *Review of Educational Research*, *83*(3), 386–431. doi:10.3102/0034654313483906

Kim, J. S., Relyea, J. E., Burkhauser, M. A., Scherer, E., & Rich, P. (2021b). Improving elementary grade students' science and social studies vocabulary knowledge depth, reading comprehension, and argumentative writing: A conceptual replication. *Educational Psychology Review, 33*, 1935-1964. doi:10.1007/s10648-021-09609-6

Kim, J. S., & White, T. G. (2008). Scaffolding voluntary summer reading for children in grades 3 to 5: An experimental study. *Scientific Studies of Reading, 12*(1), 1–23. doi:10.1080/10888430701746849

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95,* 163-182. http://dx.doi.org/10.1037/0033-295X.95.2.163

Kintsch, W. (1993). Information accretion and reduction in text processing: Inferences. *Discourse Processes, 16*, 193–202.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition.* New York, NY: Cambridge University Press.

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.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*(5), 363-394.

Kulikowich, J. M., & Hepfer, M. L. (2018). The model of domain learning (MDL): Stages of component interplays in academic development. In H. Fives & D. L. Dinsmore (Eds.), *The model of domain learning: Under- standing the development of expertise* (pp. 3–18). New York, NY: Routledge.

Levin, J. R. (1988). Elaboration-based learning strategies: Powerful theory = powerful application. *Contemporary Educational Psychology*, *13*(3), 191–205. https://doi:10.1016/0361-476X(88)90020-3

Lupo, S.M., Berry, A., Thacker, E., Sawyer, A., & Merritt, J. (2019). Rethinking text sets to support knowledge building and interdisciplinary learning. *The Reading Teacher, 73*(4), 513-524. https://doi.org/10.1002/trtr.1869

Mantzicopoulos, P., Patrick, H., & Samarapungavan, A. (2013). Science literacy in school and home contexts. *Cognition and Instruction, 31,* 62–119. http://dx.doi.org/10.1080/07370008.2012.742087

Marinak, B. A., Malloy, J. B., Gambrell, L. B., & Mazzoni, S. A. (2015). Me and my reading profile. *The Reading Teacher*, *69*(1), 51–62. https://doi.org/10.1002/trtr.1362

Mayer, R. E., & Wittrock, M. C. (1996). Problem solving transfer. In D. C. Berliner & R. Calfee (Eds.), *Handbook of Educational Psychology* (pp. 47–62). New York, NY: Simon & Schuster Macmillan.

McCombs, J., Augustine, C., Unlu, F., Ziol-Guest, K., Naftel, S., Gomez, C. J., Marsh, T., Akinniranye, G., & Todd, I. (2019). *Investing in Successful Summer Programs: A Review of Evidence Under the Every Student Succeeds Act*. RAND Corporation. https://doi.org/10.7249/RR2836

McCombs, J. S., Augustine, C. H., Pane, J. F., & Schweig, J. (2020). *Every summer counts: A longitudinal analysis of outcomes from the national summer learning project*. RAND Corporation. https://doi.org/10.7249/RR3201

McKeown, M. G., Deane, P. D., Scott, J. D., Krovetz, R., & Lawless, R. R. (2017). *Vocabulary assessment to support instruction: Building rich word-learning experiences*. New York, NY: Guilford Press.

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

Nagy, W.E. (2005). Why instruction needs to be long-term and comprehensive. In E.H. Hiebert & M.L. Kamil (Eds.), *Teaching and learning vocabulary: Bringing research to practice* (pp. 27–44). New York, New York: Routledge.

Nagy, W.E. (2007). Metalinguistic awareness and the vocabulary-comprehension connection. In R. Wagner, A. Muse, & K. Tannenbaum (Eds.), *Vocabulary acquisition: Implications for reading comprehension* (pp. 53-77). New York, New York: Guilford Press.

National Center for Education Statistics. (2019). National Assessment of Educational Progress (NAEP) Report Card: 2019 NAEP Reading Assessment.

Nesbit, J. S., Adesope, O. O. (2006). Learning with concept and knowledge maps: A meta-analysis. *Review of Educational Research, 76*, 413-448.

Neuman, S. B., Newman, E. H., & Dwyer, J. (2011). Educational effects of a vocabulary intervention on preschoolers' word knowledge and conceptual development: A cluster-randomized trial. *Reading Research Quarterly, 46*, 249–272.

Newmann, F. M., Smith, B. A., Allensworth, E., & Bryk, A. S. (2001). Instructional program coherence: What it is and why it should guide school improvement policy. *Educational Evaluation and Policy Analysis, 23*(4), 297–321.

Northwest Evaluation Association. (2011). *Technical manual for measures of academic progress and measures of academic progress for primary grades*. Portland, OR: Author.

O'Reilly, T., Wang, Z., & Sabatini, J. (2019). How much knowledge is too little? When a lack of knowledge becomes a barrier to comprehension. *Psychological Science, 30*(9), 1344–1351. https://doi.org/10.1177/0956797619862276

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

Pearson, D. P., Moje, E., & Greenleaf, C. (2010). Literacy and science: Each in the service of the other. *Science, 323*, 459-463.

Pearson, P. D., Palincsar, A. S., Biancarosa, G., & Berman, A. I. (Eds.). (2020). *Reaping the rewards of the Reading for Understanding Initiativ*e. Washington, DC: National Academy of Education.

Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, *11*, 357–383.

Perfetti, C. A., & Stafura, J. (2014). Word knowledge in theory of reading comprehension. *Scientific Studies of Reading, 18*, 22-37.

Phillips, M., & Chin, T. 2004. "How Families, Children, and Teachers Contribute to Summer Learning and Loss." In *Summer Learning: Research, Policies, and Programs*, edited by G. Borman and M. Boulay. Mahwah, NJ: Lawrence Erlbaum.

Pressley, M., Disney, L., & Anderson, K. (2007). Landmark vocabulary instructional research and the vocabulary instructional research that makes sense now. In R. K. Wagner, A. E. Muse, & K. R. Tannenbaum (Eds.), *Vocabulary acquisition: Implications for reading comprehension* (pp. 205–232). Guilford Press.

Quint, J., Zhu, P., Balu, R., Rappaport, S., & DeLaurentis, M. (2015). *Scaling up the success for all model of school reform: Final report from the investing in innovation (i3) evaluation*. MDRC.

Rabe-Hesketh, S., & Skrondal, A. (2021). *Multilevel and longitudinal modeling using Stata* (Fourth Edition). College Station, TX: Stata Press. Volume I: Continuous Responses.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models*. Thousand Oaks, CA: Sage.

Read, J. (1998). Validating a test to measure depth of vocabulary knowledge. In A. Kunnan (Ed.), *Validation in language assessment* (pp. 41-60). New York, New York: Erlbaum.

Read, J. (2004). Plumbing the depths: How should the construct of vocabulary knowledge be defined. In B. Laufer & P. Bogaards (Eds.), *Vocabulary in a second language: Selection, acquisition and testing* (pp. 209-227). Amsterdam, the Netherlands: John Benjamins. http://dx.doi.org/10.1075/lllt.10.15rea

Reardon, S. F., Valentino, R. A., & Shores, K. A. (2012). Patterns of literacy among U.S. Students. *Future of Children, 22*(2), 17-37.

Rist, R.S. (1989). Schema creation in programming. *Cognitive Science 13*, 389–414. http://doi.org/10.1207/s15516709cog1303_3

Romance, N. R., & Vitale, M. R., (2001) Implementing an in-depth expanded science model in elementary schools: Multi-year findings, research issues, and policy implications, *International Journal of Science Education, 23*(4), 373-404.

Ryan, R. M., & Deci, E. L. (2000). Facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68-78.

Salomon, G., & Perkins, D. N. (1989). Rocky roads to transfer: Rethinking mechanism of a neglected phenomenon. *Educational Psychologist*, *24*(2), 113-142.

Schiefele, U. (1999). Interest and learning from text. *Scientific Studies of Reading*, *3*(3), 257–279. https://doi.org/10.1207/s1532799xssr0303_4

Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning, 64*(4), 913–951.

Schustack, M. W., & Anderson, J. R. (1979). Effects of analogy to prior knowledge on memory for new information. *Journal of Verbal Learning & Verbal Behavior*, *18*(5), 565–583. https://doi.org/10.1016/S0022-5371(79)90314-1

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, U.S. Department of Education.

Sinatra, G. M., Heddy, B. C., & Lombardi, D. (2015). The Challenges of Defining and Measuring Student Engagement in Science. *Educational Psychologist, 50*(1), 1-13.

Smith, R., Snow, P., Serry, T., & Hammond, L. (2021). The role of background knowledge in reading comprehension: A critical review. *Reading Psychology*, *42*(3), 214–240. https://doi.org/10.1080/02702711.2021.1888348

Snow, C. E., Burns, M. S., & Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

Snow, C. E., Lawrence, J. F., & White, C. (2009). Generating knowledge of academic language among urban middle school students. *Journal of Research on Educational Effectiveness*, *2*(4), 325–344. https://doi.org/10.1080/19345740903167042

Stahl, S. A., & Fairbanks, M. M. (1986). The effects of vocabulary instruction: A model-based meta-analysis. *Review of Educational Research, 56*, 72–110.

Stein, M. (2017). Supporting the summer reading of urban youth: An evaluation of the Baltimore SummerREADS program. *Education and Urban Society, 49*(1), 29-52. https://doi.org/10.1177/0013124516630595

Stipek, D., Franke, M., Clements, D., Farran, D., & Coburn, C. (2017). PK-3: What does it mean for instruction? *Social Policy Report, 30*, 1–23. http://dx.doi.org/10.1002/j.2379-3988.2017.tb00087.x

Stoddart, T., Pinal, A., Latzke, M., & Canaday, D. (2002). Integrating inquiry science and language development for English language learners. *Journal of Research in Science Teaching, 39*(8), 664–687.

Sweller, J. (2011). Cognitive load theory. In J. P. Mestre & B. H. Ross (Eds.), *The psychology of learning and motivation: Cognition in education* (pp. 37–76). Elsevier Academic Press. https://doi.org/10.1016/B978-0-12-387691-1.00002-8

Talwar, A., Tighe, E. L., & Greenberg, D. (2018) Augmenting the Simple View of Reading for Struggling Adult Readers: A Unique Role for Background Knowledge. *Scientific Studies of Reading*, 22(5), 351-366. https://doi.org/10.1080/10888438.2018.1450410

Thum Y. M., & Hauser, C. H. (2015). NWEA 2015 MAP Norms for Student and School Achievement Status and Growth. NWEA Research Report. Portland, OR: NWEA

Tunmer, W. E., Herriman, M. L., & Nesdale, A. R. (1988). Metalinguistic abilities and beginning read. *Reading Research Quarterly, 23*(2), 134-158.

van Dijk, T. A. (1977). Knowledge frames, macrostructures, and discourse comprehension. In M. Just & P. Carpenter (Eds.), *Cognitive processes in comprehension* (pp. 3-32). Hillsdale, NJ: Lawrence Erlbaum Associates.

Vaughn, S., Swanson, E.A., Roberts, G., Wanzek, J., Stillman-Spisak, S. J., Solis, M., & Simmons, D. (2013). Improving reading comprehension and social studies knowledge in middle school. *Reading Research Quarterly, 48*(1), 77–93.

Vitale, M. R., & Romance, N. R. (2007). A knowledge-based framework for unifying content area reading comprehension and reading comprehension strategies. In D. McNamara (Ed.) *Reading comprehension strategies: Theory, interventions, and technologies* (pp. 73-104). Erlbaum.

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(2), 457–472.

What Works Clearinghouse (2021). WWC Reporting guide for study authors. Washington, DC: Institute of Education Sciences; Retrieved February 1, 2022 from https://ies.ed.gov/ncee/wwc/Docs/ReferenceResources/WWC_Author_Guide_Jul2021.pdf

Williams, J. P., Kao, J. C., Pao, L. S., Ordynans, J. G., Atkins, J. G., Cheng, R., & DeBonis, D. (2016). Close analysis of texts with structure (CATS): An intervention to teach reading comprehension to at-risk second graders. *Journal of Educational Psychology, 108*(8), 1061–1077. https://doi.org/10.1037/edu0000117

Wilson, M., & de Boeck, P. (2004). Descriptive and explanatory item response models. In P. de Boeck & M. Wilson (Eds.), *Explanatory item response models: A generalized linear and nonlinear approach* (pp. 43–74). New York, NY: Springer.

Wiske, M. S. (1998). What is teaching for understanding? In M. S. Wiske (Ed.), *Teaching for understanding* (pp. 61–86). San Francisco: Jossey-Bass.

Woodfin, L., & Plaut, S. N. (2017). *The essential guide to teaching the EL education K-5 language arts curriculum.* New York, NY: EL Education.

Wright, T. S., & Cervetti, G. N. (2016). A systematic review of the research on vocabulary instruction that impacts comprehension. *Reading Research Quarterly, 52 (2)*, 203-226.

Table 1

Conceptual Framework for General and Topic-Specific Schemas in Grade 1 and Grade 2 and Corresponding Key Questions and Structural Components

Types of schema	Key questions and structural components
General schema	How do scientists study past events? A. Antecedent conditions 1. Goal: Use evidence to study past events, people, places, things 2. Resources: Collect multiple pieces of evidence 3. Constraint: Cannot directly observe people, places, things B. Solution plan: Use multiple sources of evidence to theorize C. Outcome: Multiple sources of evidence support claims
Topic-specific schema in Grade 2 science	How do paleontologists study pre-historic events and animals, like dinosaur mass extinctions? A. Antecedent conditions 1. Goal: Use evidence to study pre-historic animals like dinosaurs 2. Resources: Collect multiple pieces of evidence about dinosaurs 3. Constraint: Cannot directly observe dinosaurs and their physical characteristics, behaviors, habitats B. Solution plan: Use multiple sources of evidence to theorize about dinosaurs' physical characteristics and behaviors C. Outcome: Multiple sources of evidence to support claims about dinosaurs
Topic-specific schema in Grade 1 science	How do animals survive in their habitat? A. Antecedent conditions 1. Goal: Animals want to survive in their habitat 2. Resources: Animals need food, water, air, protection from external events (e.g., weather) 3. Constraint: Lacking resources, not having the advantage over predators B. Solution plan: The physical characteristics and behaviors of animals adapt over time C. Outcome: Animals stay alive (don't become extinct) through adaptations

Table 2

*Descriptive Characteristics of Grade 1 Teachers (*N *= 144) and Students (*N *= 2,275) at the Beginning of the Study*

Characteristics	Overall	Treatment	Control
Teachers			
N	144	78	66
Female	97%	97%	97%
White	62%	67%	56%
Has or working toward Masters	48%	51%	44%
Attended PD for reading research to classroom practice	47%	42%	53%
Students			
N	2,275	1,260	1,015
Male	51%	51%	51%
White	18%	15%	21%
Black	36%	36%	35%
Asian	8%	7%	9%
Latinx	35%	38%	31%
Other race	3%	3%	4%
Limited English proficiency	24%	26%	21%
Individual education plan	8%	7%	9%
English as a primary home language	61%	59%	63%
Low SES	41%	48%	32%
Mid SES	38%	33%	44%
High SES	21%	18%	24%
Baseline MAP reading, M (*SD*)	169.27 (15.87)	167.35 (15.70)	171.64 (15.76)
Baseline MAP mathematics, M (*SD*)	170.59 (17.09)	168.59 (16.76)	173.06 (17.19)

Note. PD = professional development. SES = Socioeconomic status. MAP = Measure of Academic Progress.

Table 3

Weekly Lesson Activity Type and Sequence, Description, and Duration for the MORE Grade 2 Science Thematic Unit

Goal	Sequence	Activity Type	Description of Practices	Duration
Part 1: Interleaving Practices to Build Domain and Topic Knowledge Schemas	1	Zoom	Students were introduced to MORE mission and established learning goals and interest with a guiding question (e.g., *Why are we learning about fossils?*)	40 min.
	2	Interactive Read-Aloud I	Teacher and students read aloud thematically related informational texts (one book per week). Students engaged in a structured equitable academic discussion.	40 min.
	3	Concept Mapping	Students built a concept network using vocabulary and resources from text and supplemental materials.	40 min.
	4	Interactive Read-Aloud II	Teacher and students continued to read aloud the selected books, followed by a structured equitable academic discussion.	40 min.
	5	Collaborative research with argumentative writing and reading I	Students first conducted research by reading grade-level texts in their collaborative research groups followed by in-depth academic group discussions on relevant details and evidence from their research. Students collaboratively identified and wrote down relevant textual details from class notes on the walls to make hypotheses and argumentation about dinosaurs.	40 min.
Part 2: Interleaving Practices to Transfer Domain and Topic Knowledge Schemas	6	Word sleuthing	Teachers used the structured word inquiry approach to build metalinguistic awareness of domain specific words in concept maps.	40 min.
	7	Collaborative research with argumentative writing and reading II	Students engaged in verbal and/or written argumentation and collaborative research using text features (e.g., compare and contrast) to obtain additional information on given concepts during reading and writing activities related to dinosaurs.	40 min.
	8	Sharing our Research	Students groups presented what they investigated and discovered to the rest of the class.	40 min.
	9	Reflection	Teacher and students reflected on the week's activities and discussed which activity they liked the most and why.	40 min.

Note. MORE = Model of Reading Engagement.

Table 4
Descriptive Statistics for Fidelity of Implementation in Grade 1 and Grade 2

Fidelity of implementation components	Treatment % or M (*SD*)	Control M (*SD*)
Grade 1		
Adherence to MORE lessons[a] (randomly selected 12 treatment classrooms)	100.00%	n/a
Program differentiation		
Read-aloud text Lexile level (78 treatment & 66 control classrooms)		
All (science & social studies)	801.43L (153.62L)	498.92L (118.85L)
Science	696.25L (82.80L)	522.63L (138.72L)
Social studies	941.67L (102.26L)	439.09L (100.64L)
Openness of literacy tasks (randomly selected 12 treatment & 11 control classrooms)		
Authenticity	2.61 (.31)	1.66 (.17)
Collaboration	2.60 (.29)	1.57 (.31)
Challenge level	2.67 (.42)	1.11 (.17)
Student-directed work	2.69 (.26)	1.47 (.35)
Sustained efforts	3.00 (.00)	2.26 (.62)
Instructional time exposure[b] (78 treatment & 66 control classrooms)		
English language arts/reading instruction	507.66 (239.45)	498.51 (282.77)
Science instruction	177.34 (130.46)	109.05 (99.37)
Social studies instruction	181.82 (133.05)	108.90 (100.60)
Wide reading of thematically related texts (1,071 treatment & 810 control students)		
N of grade- and reader-level matched print books that students read in summer (max. 10 books)	5.39 (4.16)	n/a
N of digital books that students read in summer (max. 6 books)	2.13 (2.63)	1.57 (2.40)
Grade 2 (69 treatment & 64 control classrooms)		
Adherence to MORE lessons[a]		
Part 1: Interleaving practices to build and deepen domain and topic-knowledge schemas		
Developing situational interest and learning goals through zoom lessons (max. 5 lessons)	92.65%	n/a

Interactive read aloud of conceptually related informational texts (max. 10 lessons)	87.10%	n/a
Direct vocabulary instruction through concept maps (max. 5 lessons)	92.17%	n/a
Part 2: Interleaving practices to transfer and apply domain and topic-knowledge schemas		
Collaborative research with argumentative writing and reading (max. 15 lessons)	93.91%	n/a
Word Sleuthing (max. 5 lessons)	88.41%	n/a
Final reflection (max. 5 lessons)	91.30%	n/a
Adherence to the use of MORE books[a] (% of teachers who used MORE books)		
T. Rex: Hunter or Scavenger?	98.55%	n/a
Fossils: A True Book	97.10%	n/a
Dinosaurs Before Dark: Magic Tree House, or *Who Would Win? Trex v. Velociraptor*	75.36%	n/a
Boy, Were We Wrong About Dinosaurs!	98.55%	n/a
Instructional time exposure[c]		
English language arts/reading instruction	678.04 (173.79)	673.74 (182.62)
Science instruction	104.52 (108.98)	108.04 (126.30)

Note. Two additional fidelity of implementation components for Grade 1 (i.e., participant responsiveness and program differentiation) were evaluated and found not statistically significantly different between treatment and control groups ($ps > .05$). The detailed descriptions of the measures, analyses, and results are reported in Kim et al., (2021b).

[a]Treatment teachers' adherence to MORE lessons core components was assessed using audio-recordings for Grade 1 and a teacher self-report survey for Grade 2. Grade 2 classroom observation and audio-recording data collection were not conducted due to COVID-19 school closing.

[b, c]Minutes per week spent on instruction was based on teachers' self-report to the following question from the Early Childhood Longitudinal Survey, Kindergarten Class of 2010-11 (ECLS-K: 2011).

[c]Grade 2 treatment group teachers did not implement the social studies unit due to COVID-19 school closings.

Table 5

A Framework for Measuring Transfer Effects of Grade 2 Science Passages

Dimensions of Transfer	Instructional Context	Science Content Reading Comprehension Passage		
		Near Transfer	Mid Transfer	Far Transfer
General schema: How do scientists study past events?	*How do paleontologists study pre-historic events and animals, like dinosaur mass extinctions?*	*How do paleontologists study ammonites?*	*How do archaeologists study Pompei?*	*How do genealogists study our ancestors?*
Content: Academic Vocabulary Networks (N of words in the networks)	Full Exposure (12 words)	Extensive Exposure (6 words)	Limited Exposure (4 words)	No Exposure (0 words)
Context: Similarity to the Instructional context	n/a	Similar	Different	Different

Table 6

Descriptive Statistics for Analytic Sample (Mean and Standard Deviation) of Study Variables and Pairwise Correlation Matrix

Variable	Treatment		Control		Correlation								
	M (*SD*)	*n*	M (*SD*)	*n*	1	2	3	4	5	6	7	8	9
Reading: Total MAP Scores across time periods													
1. T0	167.73 (15.61)	1,176	171.98 (15.38)	980	-								
2. T1	173.47 (15.70)	1,168	177.91 (16.01)	960	.89	-							
3. T2	173.06 (16.27)	1,169	176.50 (16.51)	956	.80	.80	-						
Science content reading (T3)													
4. Overall	10.97 (4.41)	1,176	10.97 (4.04)	980	.60	.61	.65	-					
5. Near-transfer subtest	3.66 (1.75)	1,176	3.49 (1.58)	980	.42	.42	.45	.73	-				
6. Mid-transfer subtest	3.66 (1.93)	1,176	3.65 (1.81)	980	.50	.52	.54	.83	.40	-			
7. Far-transfer subtest	3.65 (1.83)	1,176	3.83 (1.80)	980	.50	.51	.55	.81	.36	.54	-		
Vocabulary knowledge depth (T3)													
8. Overall	33.27 (9.76)	1,176	34.23 (9.07)	980	.55	.57	.59	.57	.49	.43	.44	-	
9. Taught words	19.77 (5.82)	1,176	20.64 (5.47)	980	.52	.54	.55	.53	.45	.40	.42	.94	-
10. Untaught words	13.50 (4.68)	1,176	13.59 (4.42)	980	.49	.51	.53	.53	.45	.40	.40	.90	.70

Note. Domain-general MAP reading percentile ranks for the treatment group at Time 0, Time 1, and Time 2 MAP. RIT scaled scores correspond to the 38th, 38th, 45th percentile ranks for the treatment group, and 49th, 49th, and 54rd percentile ranks for the control group. MAP = Measure of Academic Progress, Reading Total RIT Scaled Score. T0 = Time 0 (January of Grade 1), T1 = Time 1 (April of Grade 1), T2 = Time 2 (Follow-up 1, September of Grade 2), T3 = Time 3 (Follow-up 2, March of Grade 2).

Table 7

*Results of Piecewise Linear Growth Curve Modeling for MAP Reading Comprehension, Vocabulary, and Reading Total (Reading Comprehension and Reading Vocabulary) Outcomes (*N *= 2,275)*

Source	Model A: MAP Reading Comprehension	Model B: MAP Vocabulary	Model C: MAP Reading Total
Fixed effects			
Intercept, β_{000}	170.45 (1.47)***	168.82 (1.37)***	169.33 (1.32)***
Time 1, β_{100}	5.72 (0.38)***	6.01 (0.42)***	5.78 (0.28)***
Time 2, β_{200}	2.83 (0.39)***	4.74 (0.49)***	4.71 (0.32)***
Treatment, β_{001}	-1.78 (0.67)**	-0.97 (0.59)	-0.90 (0.60)
Treatment × Time 1, β_{101}	0.90 (0.51)†	-0.10 (0.56)	-0.03 (0.38)
Treatment × Time 2, β_{201}	1.89 (0.52)***	0.89 (0.66)	0.73 (0.43)†
Variance components			
Level 1, ε_{tjk}	69.74	77.51	35.09
Level 2			
Intercept, ζ_{0jk}	42.60	44.77	32.00
Rate of change, ζ_{01k}	2.82	21.85	8.98
Correlation, τ_{01}	-0.04	-0.43	0.05
Level 3, ν_{00k}	1.89	1.01	2.15
Goodness of fit			
Deviance (-2LL)	50,262.55	51,958.63	47,770.81
Akaike information criterion	50,312.55	52,008.63	47,820.81
Bayesian information criterion	50,482.91	52,179.23	47,991.41

Note. Student demographic characteristics (i.e., gender, race/ethnicity, grade, limited English proficiency/individual education plan status, socioeconomic status), Measures of Academic Progress (MAP) mathematics pretest scores, and school randomization blocks were included in the analyses as covariates.

†$p < 0.10$, **$p < 0.01$, ***$p < 0.001$

Table 8

*Results of Hierarchical Models Predicting Treatment Effects on Science Content Reading Comprehension and Science Vocabulary Knowledge (*N *= 2,156)*

	Coefficient (*SE*)					
	Science content reading comprehension				Science vocab	
Source	Overall	Near transfer	Mid transfer	Far transfer	Overall	Ta
Fixed effects						
Intercept, γ_{00}	.09 (.16)	.003 (.17)	-.04 (.17)	.25 (.17)	-.15 (.17)	-.2
Treatment, γ_{01}	.18 (.06)**	.22 (.06)***	.17 (.06)**	.04 (.06)	.05 (.06)	-.0
Variance components						
Level 1, ε_{ij}	.57	.76	.69	.72	.64	.68
Level 2, ζ_{0j}	.02	.02	.02	.02	.02	.02

Note. Student demographic characteristics (i.e., gender, race/ethnicity, grade, limited English proficiency/individual education plan status, socioeconomic status), Measures of Academic Progress (MAP) reading and mathematics pretest score, and school randomization blocks were included in the analyses as covariates.

†$p < 0.10$, **$p < 0.01$, ***$p < 0.001$.

Table 9

Fixed and Random Effects from Cross-Classified Multi-Level Logistic Regression Models Examining Item-Specific Transfer Effects on Science-Content Reading Comprehension (N = 2,156)

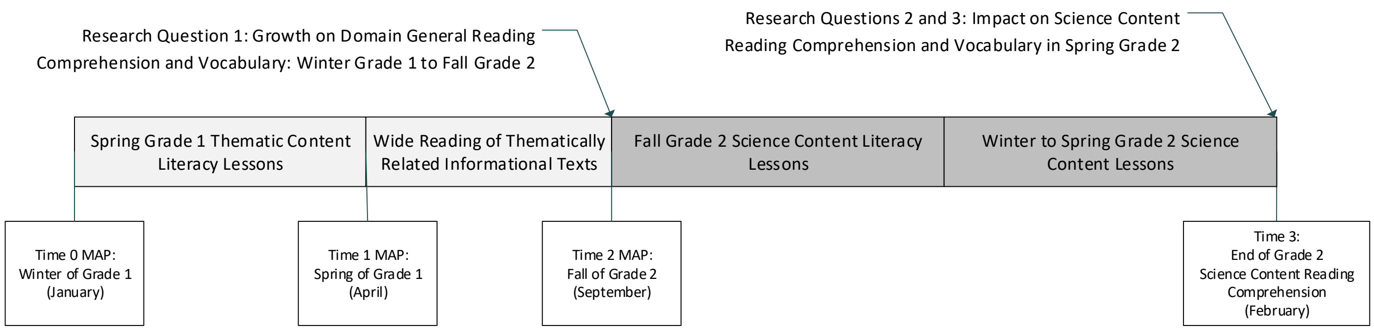
	Item-specific transfer effect		
Source	Logit (*SE*)	OR	95% CI
Fixed effects			
Intercept, β_{00}	0.10 (0.28)	1.11	-0.46 to 0.66
Treatment, β_{01}	0.31 (0.07)***	1.37	0.17 to 0.46
Mid transfer, β_{10}	0.09 (0.32)	1.10	-0.54 to 0.73
Far transfer, β_{20}	0.64 (0.34)†	1.90	-0.02 to 1.30
Treatment × Mid transfer, β_{11}	-0.12 (0.05)*	0.89	-0.22 to -0.01
Treatment × Far transfer, β_{21}	-0.26 (0.05)***	0.77	-0.36 to -0.15
Variance components			
Passage item, ζ_{ik}	0.36		
Student, ϵ_{jk}	0.39		
School, ν_{0k}	0.03		

Note. Demographic characteristics (i.e., gender, race/ethnicity, grade, limited English proficiency/individual education plan status, socioeconomic status), Measures of Academic Progress (MAP) reading and mathematics pretest scores, and school randomization blocks were included in the analyses as covariates but suppressed in Table 9. OR= odds ratio. CI= confidence interval.

†$p < 0.10$, *$p < 0.05$, ***$p < 0.001$.

Figure 1

Visual Displaying Study Design for Implementation of Intervention Activities and Measurement of Reading Outcomes



Note. MAP = Measures of Academic Progress. Light gray indicates Grade 1 school year and summer. Dark gray indicates Grade 2 school year.

Figure 2

MORE Intervention Consort Diagram

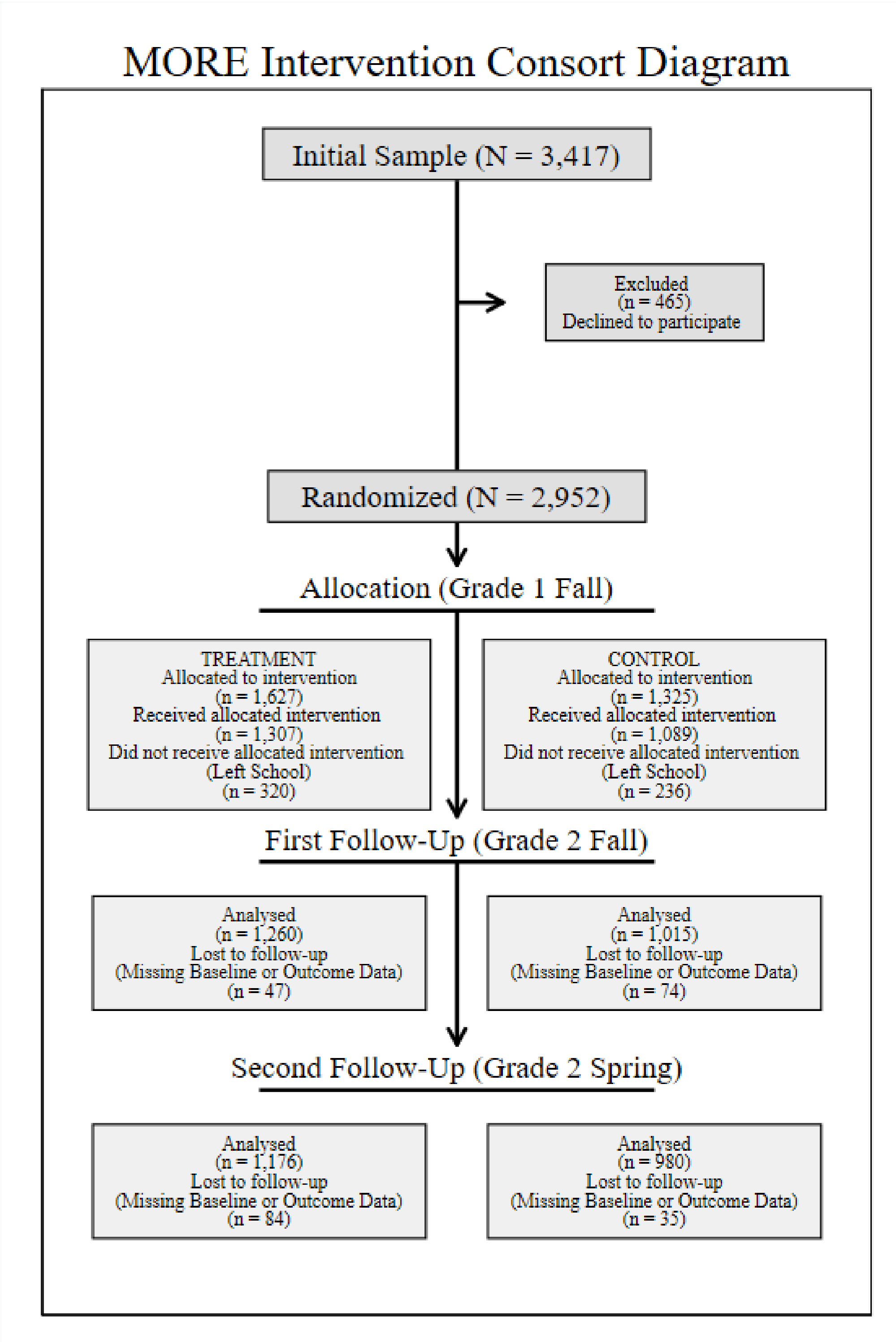
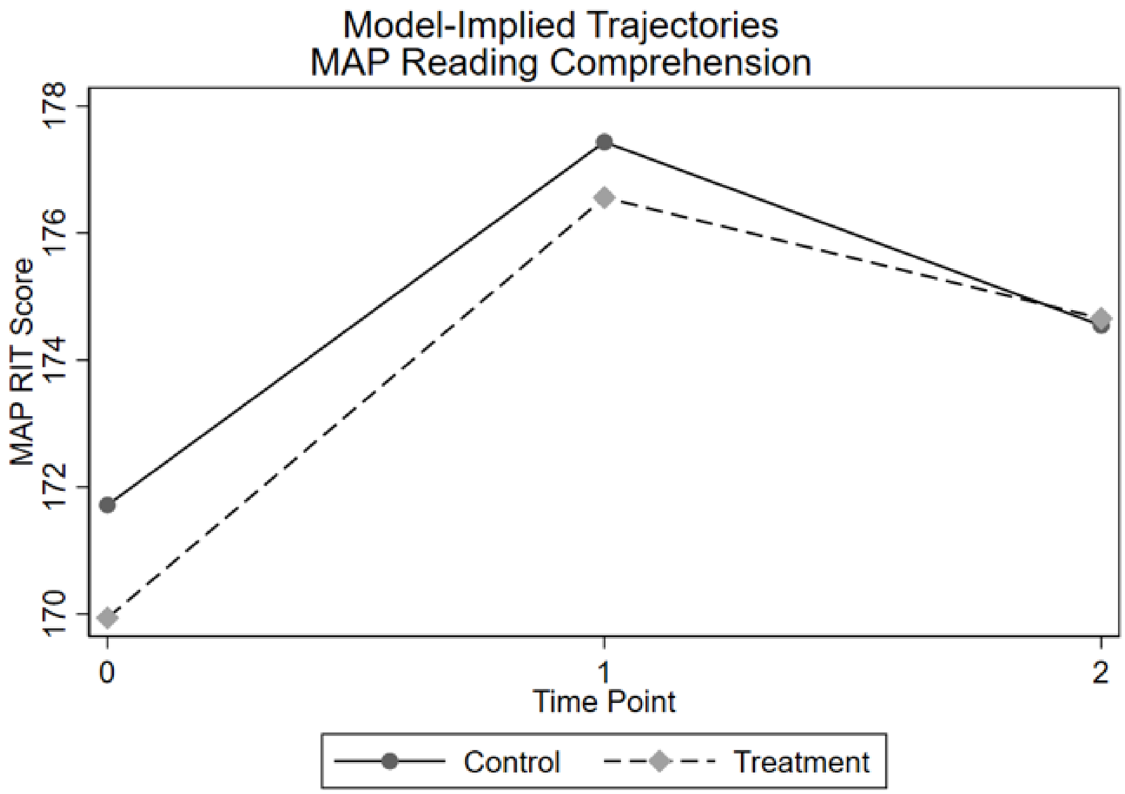


Figure 3

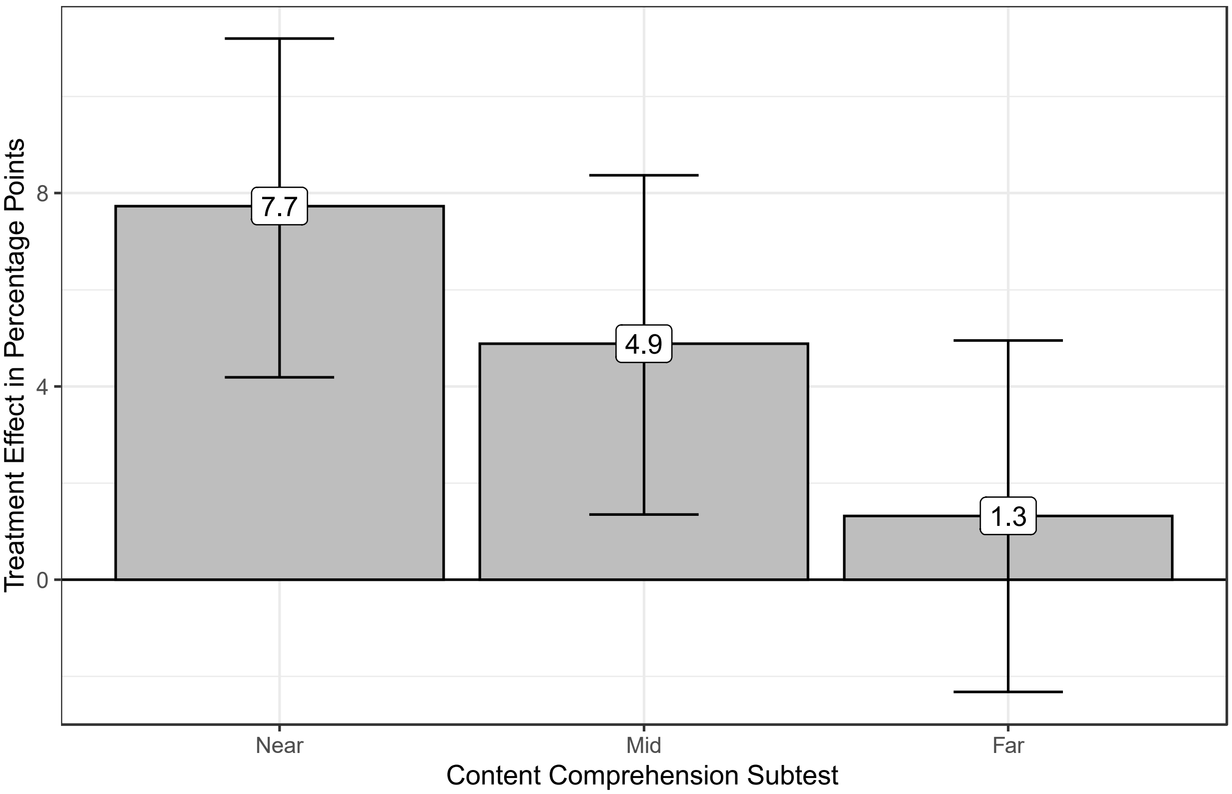
Prototypical Trajectories for Growth on Students' MAP Reading Comprehension by Experimental Condition



Note. MAP = Measures of Academic Progress. All outcomes were measured by MAP Rasch unit (RIT) scale rea
assessment using within-person predictors for time. Horizontal axis represents assessment time points: Time 0 =
2019; Time 1 = Grade 1 April 2019, and Time 2 = Grade 2 September 2019. A solid line represents a growth tra
control group students and a dashed line represents that for the treatment group students (N = 2,275).

Figure 4

Treatment Effects Derived from Explanatory Item Response Model



Note. Results based on cross-classified multi-level logistic regression models (see Table 9). Graph of model-implied treatment effects on the correct response probability relative to control group students with a 50% correct response probability on the three subtests of the domain-specific (science) reading comprehension outcome (N = 2,156).

A Longitudinal Randomized Trial of a Sustained Content Literacy Intervention from First to Second Grade: Transfer Effects on Students' Reading Comprehension

Online Supplementary Materials

Supplementary Materials S-A

Table S-A
Modifications to Original Design for the MORE Intervention.

Sustained Content Literacy Intervention: Core Components				
	Spring 2019	Summer 2019		Fall 2019 to Spring 2020
School A (Grade 1-2)*	Grade 1 MORE Thematic Content Literacy Lessons and Assessments (Science and Social Studies)	Grade 1 Wide Reading of Thematically Related Informational Texts	New School Year; Students promoted and surveyed about summer reading	Grade 2 MORE Thematic Content Literacy Lessons and Assessments (Science and ~~Social Studies~~)
School B (Grade 2-3)**	Grade 2 MORE Thematic Content Literacy Lessons and Assessments (Science and Social Studies)	Grade 2 Wide Readingg of Thematically Related Informational Texts		Grade 3 MORE Thematic Content Literacy Lessons and ~~Assessments (Science and Social Studies~~)

Figure 1. Actual plan for MORE sustained content literacy intervention from first to second grade.
*Control group for Grade 1 students in School B in Grade 1
**Control group for Grade 2 students in School A in Grade 2

Note. Strikethrough indicates research plan that was not implemented due to school closures in spring 2020. *Control group for Grade 1 students in School B in Grade 1. **Control group for Grade 2 students in School A in Grade 2. MORE = More of Reading Engagement.

Table S-A visually displays the original research plan as well as the modified plan to adjust to the school closure situation. The original plan was to follow the Grade 1 treatment group children into Grade 2 and Grade 2 treatment children into Grade 3 to focus on the sustained effects of the intervention over time. The final spring 2020 MORE lessons were to occur in both science and social studies units. The spring 2020 intervention core components was designed to be identical to spring 2019 implementation except that new topics were selected for the lessons in science and social studies. The statistical models were originally designed as individual growth models to account for repeated measures across the 12-month period, including four waves of a standardized domain-general reading comprehension assessment (i.e., Measures of Academic Progress [MAP]) given before and after lesson Grade 1 lesson implementation in school year 2018-2019 and before and after Grade 2 lesson implementation in school year 2019-2020. In the piecewise linear growth models, our first aim was to examine the impact of Grade 1 lessons followed by wide reading of thematically related informational texts over the summer on reading growth across three time points on the domain-general reading comprehension measure (MAP). Our second aim was to examine the impact of the full sequence of Grade 1 lessons, summer support for reading, and Grade 2 lessons on reading growth across four time points.

The assessment administration plan was affected by school closures in late spring 2020, but the first two waves of the assessment were successfully completed during the spring 2019 data collection and those results have been reported in a published study (Kim, Relyea, Burkhauser, Scherer, & Rich, 2021).

However, as the second round of implementation in spring 2020 was about to occur for Grade 2 and Grade 3 children who had moved from Grade 1 and Grade 2, respectively, the COVID-19 pandemic triggered school closings in our study site. Accordingly, modifications had to be made to our original plan on lesson implementation, post-test assessment, and data analytic approach, as shown in Table S-A. Specifically, in the treatment classrooms of Grade 2 and Grade 3, teachers delivered science unit lessons but were unable to implement the subsequent social studies unit. In terms of post-tests, we administered two sets of assessment to Grade 2 students: researcher developed measures of (a) science vocabulary knowledge depth test and (b) transfer tests of science content reading comprehension. None of the assessments were given to Grade 3 students due to the incompletion of the science thematic lessons and assessments.

We adjusted our plans to evaluate fidelity of lesson implementation by conducting teacher surveys instead of classroom observation. In short, due to the COVID-19 impact on data collection, available data allowed us to examine effects of the 12-month Grade 1 to Grade 2 intervention on growth in domain general reading comprehension (research question 1) and impact on spring Grade 2 science vocabulary knowledge and transfer effects on science content comprehension outcomes (research question 2 and 3).

Supplementary Materials S-B

Table S-B.1

List of Informational Science and Social Studies Books Used in Grade 1 Treatment and Control Conditions

Domain	Treatment	Control
Science	• *Ultimate Ocean Rumble* • *Polar Bear vs Grizzly Bear* • *Desert Animal Adaptations* • *Rhino Rescue* • *Killer Whale vs Great White Shark* • *Penguins* • *What was the Ice Age* • *Desert Animal Adaptations*	• *Velociraptor* • *Tacky the Penguin* • *What If You Had Animal Teeth?* • *National Geographic Readers: Great Migrations Whales* • *National Geographic Readers: Sharks!* • *Emperor Penguins* • *Penny the Rude Penguin* • *Bee* • *Owls- National Geographic Kids* • *Lions - National Geographic Kids* • *National Geographic Kids - Sharks* • *Night and Day* • *What Makes Day and Night* • *Why Do We Need Soil?* • *Earth's Landforms and Bodies of Water* • *Is All Soil the Same?* • *Hawaii Volcanoes National Parks (Reading A to Z)* • *Hummingbirds (Reading A to Z)* • *Blue Whales: Giant Mammals (Reading A to Z)*
Social studies	• *You Wouldn't Want To Be A Viking Explorer* • *Keep On!: The Story of Matthew Henson Co-Discoverer of the North Pole* • *You Wouldn't Want To Be Lewis and Clark* • *National Geographic Readers: Sacagawea* • *You Wouldn't Want to Be a Polar Explorer* • *The Inuit Thought of It*	• *Meet My Neighbor, The Dentist* • *Goods and Services Around Town* • *All Kinds of Homes (RAZ Kids)* • *George Washington and the General's Dog* • *George Washington Carver* • *Epic Books My Itty Bitty Bio on Rosa Parks* • *Epic Books My Itty Bitty Bio on Benjamin Bannekar* • *Epic Books My Itty Bitty Bio on Howard Lewis Lattimer* • *Epic Books My Itty Bitty Bio on Jackie Robinson*

	• *President's Day* • *A Day in the Life of a Garbage Collector*

Table S-B.2

List of Thematically Related Informational Print Books (N = 33) Read During Grade 1 Summer

Title	Author	Lexile	Level
Grade-level matched texts (choice of five books)			
Who Would Win: Ultimate Ocean Rumble	Jerry Pallotta	680L	n/a
Who Would Win: Lion vs. Tiger	Jerry Pallotta	640L	n/a
Who Would Win? Tyrannosaurus Rex vs. Velociraptor	Jerry Pallotta	660L	n/a
Who Would Win: Killer Whale vs. Great White Shark	Jerry Pallotta	620L	n/a
Who Would Win: Hammerhead vs. Bull Shark	Jerry Pallotta	780L	n/a
Who Would Win: Komodo Dragon vs. King Cobra	Jerry Pallotta	680L	n/a
Who Would Win: Polar Bear vs. Grizzly Bear	Jerry Pallotta	720L	n/a
Who Would Win: Whale vs. Giant Squid	Jerry Pallotta	700L	n/a
Who Would Win: Tarantula vs. Scorpion	Jerry Pallotta	570L	n/a
What if You Had Animal Feet?	Sandra Markle	AD950L	n/a
What if You Had Animal Eyes?	Sandra Markle	AD710L	n/a
What if You Had T. Rex Teeth?	Sandra Markle	AD630L	n/a
What if You Had Animal Ears?	Sandra Markle	AD630L	n/a
Reader-level matched texts (choice of five books)			
National Geographic Readers: Saving Animal Babies	Amy Shields	520L	med
I Like Bugs	Margaret Wise Brown	BR 20L	low
National Geographic Readers: Weird Sea Creatures	Laura Marsh	NC 630L	med
National Geographic Readers: Giraffes	Laura Marsh	420L	low
National Geographic Readers: Cheetahs	Laura Marsh	580L	high
Tigers at Twilight (Magic Tree House, No. 19)	Mary Pope Osborne	510L	high
Fly Guy Presents: Sharks (Scholastic Reader, Level 2)	Tedd Arnold	540L	high
Tyrannosaurus Rex (StoryBots) (Step into Reading)	Storybots	350L	low
Big Shark, Little Shark (Step into Reading)	Anna Membrino	150L	low
What do you do with a tail like this?	Steve Jenkins	510L	med
Butterflies (Penguin Young Readers, Level 2)	Emily Neye	460L	med

National Geographic Readers: Swing Sloth!: Explore the Rain Forest	Susan Neuman	100L	low
National Geographic Readers: Monkeys	Anne Schreiber	590L	med
National Geographic Readers: Sharks! (Science Reader Level 2)	Anne Schreiber	520L	high
National Geographic Readers: Great Migrations Amazing Animal Journeys	Laura Marsh	810L	highest
National Geographic Readers: Sacagawea (Readers Bios)	Kitson Jazynka	850L	highest
National Geographic Kids Chapters: Rhino Rescue: And More True Stories of Saving Animals (NGK Chapters)	Clare Hodgson Meeker	780L	highest
What Was the Ice Age? (Chapters 1-3)	Nico Medina	840L	highest
Keep On! The Story of Matthew Henson	Deborah Hopkinson	AD1080L	highest
Penguins and Antarctica: A Nonfiction Companion to Magic Tree House Merlin Mission #12: Eve of the Emperor Penguin	Mary Pope Osborne	700L	highest

Note. n/a = non-applicable

Table S-B.3

List of Thematically Related Digital Books (N = 6) Used for the MORE@Home App During Grade 1 Summer

WOOP Book Title
How I WOOPED Life: The True Story of Basketball Superstar Jamarion Styles
How I WOOPED Life: The True Story of Supreme Court Superstar Sonia Sotomayor
How I WOOPED Life: The True Story of NFL Superstar and Math Whiz John Urschel
How I WOOPED Life: The True Story of NASA Superstar Katherine Johnson
How I WOOPED the 2008 Presidential Election: The story of how President Obama beat the odds and won North Carolina
How I WOOPED Life: The True Story of Harriet Tubman

Note. WOOP = Wish, Outcome, Obstacle, and Plan. Both treatment- and control- Grade 1 students' survey asked: *On the MORE @ Home app, there were 6 digital books you could read. About how many of these digital books did you read over the summer?*

Supplementary Materials S-C

Table S-C.1

Results of Four-Level Piecewise Linear Growth Curve Modeling with Added Teacher-Level Random Effects for MAP Reading Comprehension, Vocabulary, and Reading Total (Reading Comprehension and Reading Vocabulary) Outcomes (N = 2,275)

	Model A: MAP Reading Comprehension	Model B: MAP Vocabulary	Model C: MAP Reading Total
Fixed Effects			
Intercept	170.74 (1.48)***	169.04 (1.40)***	169.41 (1.34)***
Time 1	5.62 (.39)***	5.98 (.43)***	5.74 (.29)***
Time 2	2.75 (.39)***	4.56 (.50)***	4.66 (.33)***
Treatment	-1.81 (.65)**	-.99 (.59)†	-.94 (.60)
Treatment × Time 1	.95 (.52)†	-.144 (.57)	-.03 (.38)
Treatment × Time 2	1.94 (.53)***	1.04 (.67)	.77 (.44)†
Variance Components			
Level 1 (Time level)	70.09	76.94	35.07
Level 2 (Student level)			
Intercept	41.11	43.91	31.15
Time Slope (Continuous)	2.29	21.93	8.83
Covariance	.31	-13.17	1.22
Level 3 (Teacher level)	1.02	1.08	0.47
Level 4 (School level)	1.38	.73	1.98
Goodness of fit			
Deviance (-2LL)	48416.99	50006.21	45972.88
Akaike information criterion	48466.99	50056.21	46022.88
Bayesian information criterion	48636.41	50225.86	46192.53

Note. Student demographic characteristics (i.e., gender, race/ethnicity, grade, limited English proficiency/individual education plan status, socioeconomic status), Measures of Academic Progress (MAP) mathematics pretest scores, and school randomization blocks were included in the analyses as covariates but suppressed in Table S-C.1.
†$p < .10$, **$p < .01$, ***$p < .001$.

Table S-C.2

Results of Three-Level Hierarchical Linear Modeling for the Main Treatment Effects on Science Vocabulary Knowledge Depth and Science Content Reading Comprehension (N = 2,156)

	Coefficient (*SE*)						
	Science content reading comprehension				Science vocabulary knowledge depth		
Source	Overall	Near transfer	Mid transfer	Far transfer	Overall	Taught	Untaught
Fixed effects							
Intercept	.04 (.16)	-.01 (.18)	-.06 (.17)	.19 (.17)	-.19 (.17)	-.256 (.17)	-.06 (.17)
Treatment	.16 (.06)**	.28 (.06)***	.15 (.06)**	.03 (.06)	.03 (.06)	-.03 (.06)	.11 (.06)†
Variance components							
Level 1 (Student level)	.54	.75	.65	.67	.6	.64	.67
Level 2 (Teacher level)	.04	.01	.05	.06	.05	.05	.03
Level 3 (School level)	.02	.02	.01	.00	.01	.01	.02

Note. Student demographic characteristics (i.e., gender, race/ethnicity, grade, limited English proficiency/individual education plan status, socioeconomic status), Measures of Academic Progress (MAP) reading and mathematics pretest score, and school randomization blocks were included in the analyses as covariates but suppressed in Table S-C.2.

†$p < .10$, **$p < .01$, ***$p < .001$.

Supplementary Materials S-D

Detailed information on Grade 2 Science Content Comprehension Passages and Items, IRT Analysis, and Factor Analysis

S-D.1. Grade 2 Science Content Comprehension Passages and Questions

Teacher Directions to Read Aloud: "*Today, you will read this passage about paleontologists and the ammonites. Now you will read the passage silently by yourself and then answer the questions. There are 3 answer choices for each question but only one correct answer. Circle the answer that you think is correct. After question 7, stop and wait for your teacher to read the directions. You may begin.*"

S-D.1.1. Near transfer passage: Paleontologists and the Ammonites

1

There have been five mass extinctions in the earth's history. What happened during the fifth mass extinction nearly 66 million years ago? What species was gone forever? Dinosaurs. What about fish in the oceans?

2

Paleontologists found a body fossil of a shellfish that became extinct long ago. The shellfish is called an ammonite. Ammonites were predators that lived in the ocean. They were related to other ocean predators like the octopus.

3

Ammonites survived for a long time. They had a strong shell that protected them from large fish. They were also good hunters.

4

But there are no more ammonites living today. What happened? Paleontologists have used fossil evidence to understand what happened in the past. Even 66 million years ago. One theory is that a huge rock from outer space killed many plants and animals. Another theory is that ammonites had trouble finding food before the huge rock hit the earth.

5

Body fossils help paleontologists understand what happened to living things in the past. Even when no one was around to see the animals.

Questions

1. According to the passage, who studies the fossils of ammonites?
   a. paleontologists
   b. predators
   c. hunters

2. This passage mainly tells about
   a. the physical characteristics of ammonites
   b. the theories of the ammonite's extinction
   c. the dinosaur extinction 66 million years ago

3. Some paleontologists think
   a. Ammonites became extinct before dinosaurs
   b. Many animals die during a mass extinction
   c. A big rock hit the earth and killed all the ammonites

4. In paragraph 1, what does <u>mass extinction</u> mean?
   a. A long time ago
   b. Disappeared forever
   c. All over the world

5. A strong shell is an example of a living animal's
   a. Physical characteristic
   b. Fossil evidence
   c. Hunter behavior

6. The author of this article wants you to learn
   a. How paleontologists work together
   b. How paleontologists study fossils
   c. How paleontologists look at animals

7. What is one theory about why the ammonites went extinct?
   a. The ammonites were eaten by octopuses
   b. People slowly killed off the ammonites
   c. The ammonites had a hard time finding food

<u>STOP. *Wait for your teacher to read the directions before answering the questions.*</u>

S-D.1.2. Mid-transfer passage: Archaeologists and the city of Pompeii

1

Nearly 2,000 years ago, a volcano eruption destroyed an old city. It was called

Pompei. The eruption of a mountain was brutal. It lasted for more than 24 hours. By the end, Pompeii was covered in ashes. Many people died. The city was totally destroyed. It was like a mass extinction.

2

But archaeologists discovered Pompeii again in 1748. They found fossilized humans—not dinosaur fossils but human fossils. A team of archaeologists found fossilized buildings and human and animal skeletons.

3

Pompei is like a city frozen in time. But scientists are not studying the fossilized bones of dinosaurs. Instead, they are studying the buildings and other artifacts from Pompei. They also want to know what people were like.

4

One theory about Pompeii is that there were rich and poor people. Archaeologists have found fossilized big homes and small homes.

5

Yes, Pompei was fossilized. So archaeologists have clues about life in Pompeii. Even the kinds of people who lived there.

Questions

1. According to the passage, who studies fossils of buildings destroyed by a volcano?
   a. people in Pompeii
   b. rich people
   c. archaeologists

2. This passage mainly tells about
   a. how archaeologists study the old city of Pompeii
   b. what an archaeologist's job is like
   c. how a volcano destroyed a city called Pompeii

3. In paragraph 4, <u>theory</u> can also be called:
   a. An idea

b. A solution
c. An answer

4. Archaeologists have studied the fossilized buildings and skeletons of Pompeii. Fossilized means:
   a. Saved as it was years ago
   b. New thing to help scientists
   c. Physical characteristics

5. How can we imagine what Pompeii was like a long time ago?
   a. By studying many different books about the city
   b. By studying the people currently living in Pompeii
   c. By studying fossilized buildings and skeletons

6. Which type of fossil would help an archaeologist learn about the past?
   a. Fossilized homes from a city
   b. Fossilized rocks from a volcano
   c. Fossilized bones from a dinosaur

7. How are the jobs of paleontologists and archaeologists the same?
   a. They both look for living animals
   b. They both dig for animal fossils
   c. They both use evidence to solve mysteries

S-D.1.3. Far Transfer Passage: Genealogists and our descendants

1

Some people want to know more about their ancestors. They want to know about their relatives who lived a long time ago. Often there are no living family members who know about those relatives. So sometimes they hire a genealogist.

2

A genealogist finds clues that help you to learn about your ancestors. You can learn about where your ancestors came from. You can learn about your ancestors' jobs. You can learn about what they looked like. Sometimes you can learn stories about them.

3

Some people want to prove that an ancestor was in an important event like a war. Some people want to know if they are a descendant of a famous person.

4

Other people just want to know their family lineage. Knowing about their ancestors can help them to understand themselves better.

5

Sometimes the genealogist finds clues using the computer. Sometimes she uses a library. She might talk to individuals in other countries. She might travel to other places to get clues.

Questions

1. According to the passage, who uses clues to find out about ancestors?
   a. detectives
   b. famous people
   c. genealogists

2. What is this passage mainly about?
   a. Where ancestors come from
   b. Why genealogists use the computer
   c. What genealogists do

3. What do genealogists have to do first?
   a. Talk to people
   b. Find clues
   c. Travel to other places

4. In paragraph 3, descendant can also be called:
   a. An important war hero
   b. Someone who is born later
   c. A famous person who is alive today

5. Why does the author think genealogists are important?
   a. They are able to help famous people plan their jobs
   b. They talk to people in other countries and cities
   c. They help people find relatives from a long time ago

6. What is one reason people want to find ancestors?
   a. To become famous like their ancestors
   b. To prove their relative was an important person
   c. To use the computer to find out facts

7. What is similar about the work that paleontologists, archaeologists, and genealogists do?
   a. They learn about things from the past
   b. They learn about ancestors' jobs
   c. They learn about living things

S-D.2. Item Response Theory Analysis Results

S-D.2.1. Results based on Classical Test Theory Approach

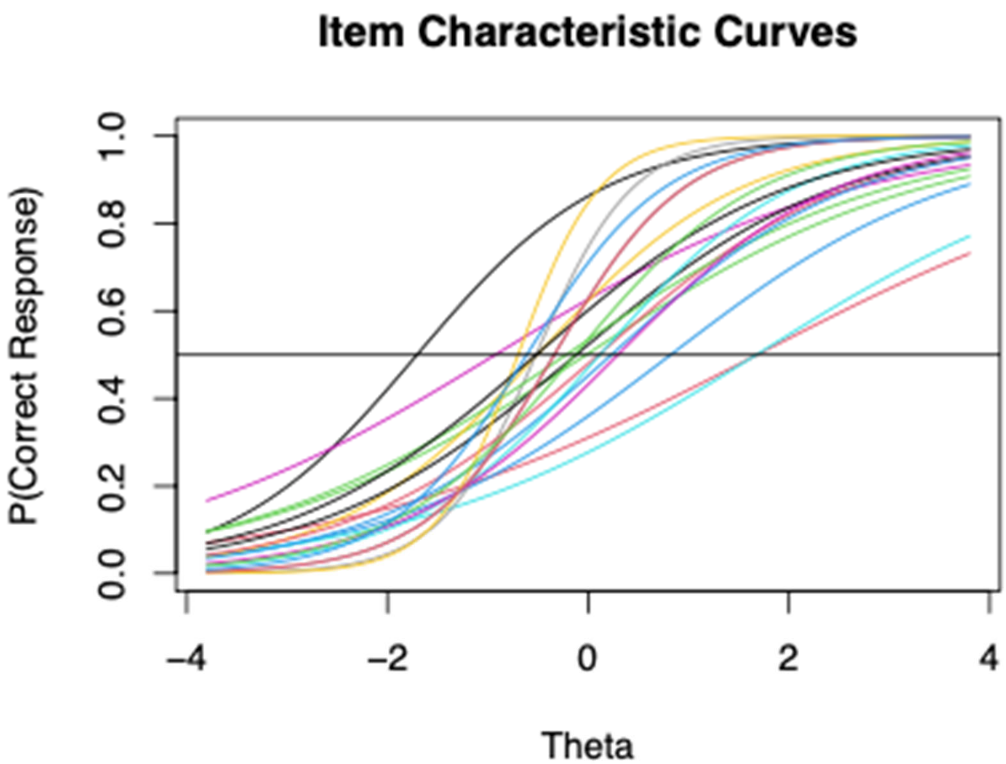
Item	Prop. Correct	Omitted-item alpha	Item-rest Correlation
Near 1	.82	.77	.33
Near 2	.32	.78	.17
Near 3	.50	.78	.26
Near 4	.37	.78	.28
Near 5	.29	.78	.22
Near 6	.62	.78	.23
Near 7	.60	.77	.38
Mid 1	.65	.76	.53
Mid 2	.52	.77	.30
Mid 3	.48	.77	.31
Mid 4	.53	.78	.25
Mid 5	.46	.77	.32
Mid 6	.47	.77	.38
Mid 7	.44	.77	.33
Far 1	.71	.77	.56
Far 2	.59	.76	.46
Far 3	.59	.77	.29
Far 5	.59	.76	.46
Far 6	.53	.77	.37
Far 7	.65	.76	.44

Note. The far-transfer passage item 4 was excluded due to a negative discrimination parameter.

S-D.2.2. Results based on Item Response Theory Approach

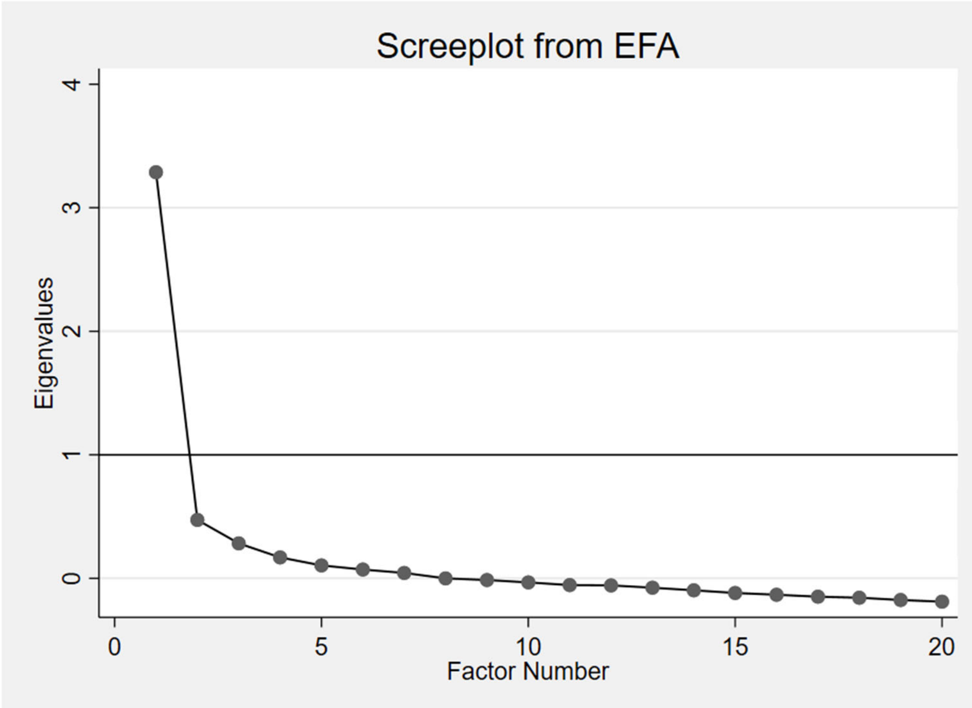
Item	Difficulty	Discrimination
Near 1	-1.71	1.08
Near 2	1.68	0.47
Near 3	-0.02	0.60
Near 4	0.82	0.70
Near 5	1.68	0.57
Near 6	-0.93	0.56
Near 7	-0.51	0.99
Mid 1	-0.52	2.08
Mid 2	-0.12	0.77

Mid 3	0.10	0.80
Mid 4	-0.22	0.62
Mid 5	0.23	0.83
Mid 6	0.12	1.03
Mid 7	0.30	0.92
Far 1	-0.70	2.46
Far 2	-0.34	1.55
Far 3	-0.52	0.80
Far 5	-0.34	1.55
Far 6	-0.13	1.09
Far 7	-0.61	1.47



S-D.2.3. Results from Exploratory Factor Analysis

Factor 1
Eigenvalue 3.29
Proportion 103.16%


Screeplot from EFA
Eigenvalues
0
1
2
3
4
0
5
10
15
20
Factor Number

Supplementary Materials S-E

S-E.1. Grade 2 Science Vocabulary Knowledge Depth Assessment

Practice Question 1: Circle two words that go with the word <u>student</u>.

book fox classroom boat

Practice Question 2: Circle two words that go with the word <u>afraid</u>.

fear blink dinner scare

1. Circle two words that go with <u>carnivore.</u>

fruit care meat prey

2. Circle two words that go with <u>extinct.</u>

gone forever living existing lost

3. Circle two words that go with <u>fossil.</u>

utensil bone folks footprint

4. Circle two words that go with <u>hypothesis.</u>

thought poem guess hippo

5. Circle two words that go with <u>brutal.</u>

harsh brave cruel friendly

6. Circle two words that go with evidence.

choosing proving showing feeling

7. Circle two words that go with theory.

fairy theme argument idea

8. Circle two words that go with hunter.

handout catcher search conclusion

9. Circle two words that go with organism.

creature animal order practice

10. Circle two words that go with trait.

characteristic truth exam feature

11. Circle two words that go with paleontologist.

paint history dinosaur pair

12. Circle two words that go with reptile.

mobile lizard report cold-blooded

Table S-E.2.
Vocabulary Knowledge Depth Measure Scoring System

Score	Item selection and correct/incorrect case
4	Student circled two correct answers
3	Student selected three items: two correct answers and one incorrect answer. Or, student selected one item: one correct answer.
2	Student selected two items: one correct and one incorrect answer. Or, student selected four items: two correct and two incorrect answers.
1	Student selected one item that was an incorrect answer. Or, student selected three items: one correct and two incorrect answers.
0	Student selected two items that were incorrect answers.

Note. We used item-response-theory to fit graded response models that account for the multiple response options. These analyses replicated the results reported in the original study (Kim, Burkhauser, 2021). We used the 0-4 point scale in the paper because it allows us to report treatment-control differences in meaningful units (i.e., number of words learned).

Supplementary Materials S-F

Multilevel Model to Examine Moderating Effects of Initial Domain-General Reading Comprehension

We examined whether students' initial domain-general reading comprehension scores (measured by MAP) moderated the impact of the treatment effect on Grade 2 students' posttest outcomes (i.e., science content reading comprehension and vocabulary knowledge depth). The multilevel model was specified as follows:
$Y_{ij} = \gamma_{00} + \gamma_{01}MORE_j + \gamma_{10}PRETEST_i + \gamma_{11}MORE_j \times PRETEST_i + \sum_{p=2}^{12} \gamma_{p0}\, COV_i + \sum_{q=2}^{8} \gamma_{0q} BLOCK_{jq} + \varepsilon_{ij} + \zeta_{0j}$

$\varepsilon_{ij} \sim N(0, \sigma^2)$, $\zeta_{0j} \sim N(0, \tau_{00})$.

In this model, Y_{ij} is the posttest score of student i in school j. The parameters γ_{00}, γ_{01}, γ_{10}, and γ_{11} represent the intercept, the main treatment effect, the MAP pretest reading score effect, and the cross-level interaction effect between treatment and the MAP pretest reading scores, respectively. The parameters γ_{p0} and γ_{0q} represent the effects of demographic characteristics covariates ($p = 2, \ldots, 12$) and randomization blocks ($q = 2, \ldots, 8$), respectively. The model included the level-1 and -2 residuals, ε_{ij} and ζ_{0j},respectively, assumed to be normally distributed with a mean of zero and uncorrelated with covariates. The results in Table S-F revealed that the interaction effects between treatment and pretest reading scores on domain-specific reading comprehension and vocabulary knowledge depth outcomes were weak but statistically significant for Near Transfer and Taught Vocabulary.

Table S-F. *Results of Treatment-by-Pretest Interaction Treatment Effects on Grade 2 Students' (*N *= 2,156) Science Content Reading Comprehension and Science Vocabulary Knowledge Depth*

	Coefficient (SE)						
	Science content reading comprehension				Science vocabulary knowledge depth		
Source	Overall	Near transfer	Mid transfer	Far transfer	Overall	Taught	Untaught
Fixed effects							
Intercept, γ_{00}	.19 (.16)	.12 (.17)	.03 (.17)	.29 (.17)†	-.08 (.16)	-.19 (.17)	.08 (.17)
Treatment, γ_{01}	.18 (.06)**	.22 (.06)***	.17 (.06)**	.04 (.06)	.04 (.06)	-.01 (.06)	.11 (.06)†
Pretest[a], γ_{10}	.43 (.04)***	.33 (.05)***	.33 (.04)***	.36 (.04)***	.39 (.04)***	.34 (.04)***	.39 (.04)***
Treatment x Pretest[a], γ_{11}	.01 (.04)	.09 (.04)*	.003 (.04)	.01 (.04)	.07 (.04)†	.09 (.04)*	.03 (.04)
Variance components							
Level 1, ε_{ij}	.59	.78	.69	.72	.656	.69	.71
Level 2, ζ_{0j}	.02	.02	.02	.02	.021	.02	.02

Note. Student demographic characteristics (i.e., gender, race/ethnicity, grade, limited English proficiency/individual education plan status, socioeconomic status), Measures of Academic Progress (MAP) mathematics pretest scores and school randomization blocks were included in the analyses as covariates but suppressed in Table.

[a]MAP reading pretest (baseline).

†$p < .10$, *$p < .05$, **$p < .01$, ***$p < .001$

Supplementary Materials S-G

Multiple Imputation for Missing Data in Assessing Treatment Effects on Science Content Reading Comprehension and Vocabulary Knowledge Depth

Missing values existed for the Measures of Academic Progress (MAP) baseline MAP mathematics (0.8%) pretest scores. With the assumption that data were missing at random (MAR), in which the probability of missing values was systematically related to the observed but not the unobserved data, we used multiple imputation by simulating 10 data sets with plausible values in place of missing observations as a sensitivity check for the main effect models (research question 2). The procedure was performed using a Markov Chain Monte Carlo (MCMC) algorithm with a multivariate normal distribution using the *mi impute mvn* command in Stata. We did not impute posttest scores.

Results from the multiple imputation analysis of the main treatment effects (see Table S-G) revealed no substantive changes in the interpretation. Furthermore, results were robust to the alternative pretest specification in interaction models, which used only the linear term of baseline MAP reading scores rather than cubic forms of both reading and mathematics MAP scores.

Table S-G. *Sensitivity Analysis Results of Main Treatment Effects on Science Content Reading Comprehension and Science Vocabulary Knowledge Depth Using Multiple Imputation for Missing Data (*N *= 2,174)*

	Coefficient (SE)						
	Science content reading comprehension				Science vocabulary knowledge depth		
Source	Overall	Near transfer	Mid transfer	Far transfer	Overall	Taught	Untaught
Fixed effects							
Intercept	.09 (.16)	-.01 (.17)	-.03 (.17)	.24 (.17)	-.15 (.16)	-.23 (.17)	-.02 (.17)
Treatment	.17 (.06)**	.23 (.06)***	.16 (.06)**	.04 (.06)	.05 (.06)	-.01 (.06)	.12 (.06)†
Variance components							
Level 1	.57	.76	.68	.72	.64	.68	.69
Level 2	.02	.02	.02	.02	.02	.02	.02

Note. Student demographic characteristics (i.e., gender, race/ethnicity, grade, limited English proficiency/individual education plan status, socioeconomic status), Measures of Academic Progress (MAP) reading and mathematics pretest scores, and school randomization blocks were included in the analyses as covariates but suppressed in Table.
†p < .01, *p < .05, **p <.01, ***p < .001.