Improving Reading Comprehension, Science Domain Knowledge, and Reading Engagement through a First-Grade Content Literacy Intervention

James S. Kim

Mary A. Burkhauser

Laura Mesite

Catherine Asher

Jackie Eunjung Relyea

Harvard University, Graduate School of Education

Jill Fitzgerald

The University of North Carolina at Chapel Hill

Jeff Elmore

MetaMetrics

Forthcoming in the *Journal of Educational Psychology*

Author Note

This research was supported by the Cheng Yu Tung Research Innovation Fund the Institute of Education Sciences, U.S. Department of Education through grant R305B150010. The opinions expressed are those of the authors and do not represent the views of the funders.

Correspondence concerning this article should be addressed to James S. Kim, Harvard University, Graduate School of Education, Cambridge, MA 02138. E-mail: james_kim@harvard.edu

Abstract

This study investigated the effectiveness of the Model of Reading Engagement (MORE), a content literacy intervention, on first graders' science domain knowledge, reading engagement, and reading comprehension. The MORE intervention emphasizes the role of domain knowledge and reading engagement in supporting reading comprehension. MORE lessons included a 10-day thematic unit that provided a framework for students to connect new learning to a meaningful schema (i.e., Arctic animal survival) and to pursue mastery goals for acquiring domain knowledge. A total of 38 first-grade classrooms (N = 674 students) within 10 elementary schools were randomly assigned to (a) MORE at school (MS), (b) MORE at home, (MS-H), in which the MS condition included at-home reading, or (c) typical instruction. Since there were minimal differences in procedures between the MS and MS-H conditions, the main analyses combined the two treatment groups. Findings from hierarchical linear models revealed that the MORE intervention had a positive and significant effect on science domain knowledge, as measured by vocabulary knowledge depth (ES = .30), listening comprehension (ES = .40), and argumentative writing (ES = .24). The MORE intervention effects on reading engagement as measured by situational interest, reading motivation, and task orientations were not statistically significant. However, the intervention had a significant, positive effect on a distal measure of reading comprehension (ES = .11), and there was no evidence of treatment-by-aptitude interaction effects. Content literacy can facilitate first graders' acquisition of science domain knowledge and reading comprehension without contributing to Matthew effects.

Keywords: content literacy intervention, science domain knowledge, reading comprehension, reading engagement, randomized controlled trial

Educational Impact and Implications Statement

In the present study, classroom teachers taught first-grade children about science knowledge while they conducted literacy instruction. Grounding literacy instruction in science content is called content literacy instruction. The aim of content literacy instruction is to help young children acquire conceptually related vocabulary while learning science (and history) content. Results indicate that content literacy instruction can improve first-graders' science domain knowledge and reading comprehension outcomes. Furthermore, there were no negative or adverse effects on first graders' reading engagement or basic literacy skills. The study suggests that content literacy instruction can improve the rigor, quality, and effectiveness of whole class literacy instruction in the early elementary grades.

Improving Reading Comprehension, Science Domain Knowledge, and Reading Engagement through a First-Grade Content Literacy Intervention

Despite the increased emphasis on reading complex and challenging texts following the publication of the Common Core State Standards (CCSS; National Governors Association, 2010), too many U.S. school children struggle to read and understand informational texts with high knowledge demands. According to descriptive findings from the Early Childhood Longitudinal Survey (Reardon, Valentino, & Shores, 2012), fewer than 5% of U.S. first graders can understand informational texts requiring prior knowledge of science and social studies content. Why do so many primary-grade children struggle to read complex texts?

Limitations on children's opportunities to acquire domain knowledge in content area subjects may play a critical role (Banilower et al., 2013; Biemiller, 2005). One obstacle to developing students' content knowledge is the overemphasis on basic academic skills (Pearson, Moje, & Greenleaf, 2010; The New Teacher Project, 2018). Particularly in the domain of primary-grade literacy instruction, teachers are held accountable for their contribution to students' acquisition of easily assessed code-related skills and literal comprehension of short narrative texts on standardized tests (Duke & Block, 2012; Kane, 2014). As a result, there are virtually no incentives for primary-grade teachers to focus on content literacy instruction. In response to these challenges, numerous researchers have suggested that embedding science content into literacy instruction may be an ideal and pragmatic approach for building students' domain knowledge, while also increasing reading engagement and, ultimately, reading comprehension (Guthrie, McRae, & Klauda, 2007; Mantzicopoulos, Patrick, & Samarapungavan, 2013; Pearson & Billman, 2016).

Over the past decade, researchers have developed and tested the efficacy of several content literacy interventions, for example: Content Area Literacy Instruction (CALI; Connor et al., 2017), In-Depth Expanded Applications of Science (IDEAS; Romance & Vitale, 2001), and Concept-Oriented Reading Instruction (CORI; Guthrie et al., 2004; Guthrie & Klauda, 2014). Although each intervention includes unique components, they share the common goal of helping students integrate newly learned concepts with prior domain knowledge to build coherent text representations (Cromley & Azevedo, 2007; Kintsch, 1988; Kintsch, 2009). Content literacy interventions have demonstrated more success in the upper-elementary and middle grades than in the early grades (Douglas & Albro, 2014), leaving open the question of how best to enhance the effectiveness of first-grade content literacy instruction.

Theoretical and Conceptual Framework for First-Grade Content Literacy

Using content literacy instruction to build students' domain knowledge in the early grades is not a novel approach to improving students' reading comprehension. Nearly 20 years ago, the National Research Council report, *Preventing Reading Difficulties in Young Children* (Snow, Burns, & Griffin, 1998), argued that integrating rich science and social studies content into early-grade literacy instruction was essential to preparing children to read complex texts in the later grades. More recently, the CCSS encouraged educators to infuse "domain-specific nonfiction titles" into literacy instruction as a first and critical step toward building children's domain knowledge and, ultimately, their reading comprehension (National Governors Association, 2010). And yet, there is a lack of evidence as to the effectiveness of such an approach, particularly on distal measures of reading comprehension. Guided by the aim of integrating science content into first-grade literacy instruction, our goal was to design and test the efficacy of a content literacy intervention that emphasizes the contributions of domain knowledge acquisition and engagement to reading comprehension.

Domain Knowledge and Reading Comprehension

Put simply, domain knowledge is a measure of how much a person knows about a particular school subject (e.g., science, history, economics), while topic knowledge is how much one knows about a particular topic within a given domain (e.g., Arctic ecosystems, 20^{th} Century explorers) (Alexander, 2003). It is well established that domain knowledge contributes to a student's ability to comprehend a text (e.g., Chiesi, Spilich, & Voss, 1979; Cromley & Azevedo, 2007; Recht & Leslie, 1988; Spilich, Vesonder, Chiesi, & Voss, 1979). In the construction-integration model of reading, Kintsch (2009) underscores the role of domain knowledge in distinguishing how novices and experts learn from text. During the construction phase, both novices and experts must remember the words and sentences in the text, hold information about the sentences in working memory, and construct an explicit textbase. Although novices are not as good as experts in performing these tasks, "there is really no qualitative difference in what they do" (Kintsch, 2009, p. 226). During the integration phase, however, there are important qualitative differences in how novice and expert readers use domain knowledge to form a situation model implied by text. Expert readers tap into retrieval structures that connect information in working memory automatically with related information in long-term memory (Kintsch, 1988; Perfetti & Stafura, 2014). Prior domain knowledge helps readers—even young children with some particular expertise—make gap-filling inferences that are critical to learning from text in science and social studies (Cervetti, Wright, & Hwang, 2016; Hirsch, 2016; Williams et al., 2016).

Research also indicates that compared with their less-knowledgeable peers, poor readers with high-domain specific knowledge can more efficiently recall text and produce superior summaries; they also demonstrate stronger motivation for reading challenging text than even good readers with low-domain specific knowledge (Brown, Bransford, Ferrara, & Campione, 1983; Recht & Leslie; 1988, Marr & Gormley, 1982; Schneider, 1985; Schneider, Körkel, & Weinert, 1989; Walker, 1987). Although decoding, oral vocabulary knowledge, and listening comprehension are critical determinants of reading comprehension (Gough & Tunmer, 1986; Hoover & Tunmer, 2018), there is growing evidence that domain-specific knowledge in science, social studies, and the humanities (Kintsch, 2005; Talwar, Tighe, & Greenberg, 2018; Recht & Leslie, 1988; Schneider, Körkel, & Weinert, 1989), as well as the construct of academic language (Snow, 2010), are also critical and can be changed through high-quality content literacy instruction (Snow, Laurence, & White, 2009; Vaughn et al., 2013).

Science Content Literacy Instruction and Reading Engagement

Not only does high-quality content literacy instruction facilitate students' domain knowledge acquisition, it can also facilitate students' engagement with complex text (Guthrie et al., 2004). Reading engagement is an umbrella term that encompasses cognitive, motivational, and behavioral characteristics of readers (Alexander, Jetton, & Kulikowich, 1995; Fredericks, Blumenfeld, & Paris, 2004; Guthrie & Wigfield, 2000; Linnenbrink-Garcia & Patall, 2016; Linnenbrink & Pintrich, 2003; Sinatra, Heddy, & Lombardi, 2015). In this study, we emphasize three manifestations of reading engagement: situational interest, reading motivation, and task orientations. In many ways, content literacy texts and tasks provide an ideal context for developing engaged readers with the interest, self-competence beliefs and tasks values, and behaviors to acquire knowledge from complex text.

Attending to students' reading engagement is important given the strong association between engagement and reading comprehension outcomes (Guthrie, Wigfield, & You, 2012). Research suggests that while children enter school with relatively high reading interest, self-competence beliefs, and task values (Wigfield et al., 2015), reading engagement tends to decline as children move through elementary school (Wigfield & Eccles, 2000). Furthermore, there is reason to believe that even early elementary students could experience decreased engagement if asked to read complex science texts without instructional supports for reading interest, self-competency beliefs, and task values. For one, large gaps in general science knowledge exist at kindergarten entry (Morgan, Farkas, Hillemeier, & Maczuga, 2016) and put some students at a disadvantage when asked to engage with texts on science topics. Also, accountability policies may mean that even young children are aware of their reading performance (Wigfield, Gladstone, & Turci, 2016). How might a science content literacy intervention support students' reading engagement?

First, science content literacy emphasizes the use of informational texts that cohere around a single theme and are designed to spark students' situational interest (Cervetti et al., 2016; Guthrie et al., 2006). Interest is the psychological state of "being engaged, engrossed, or entirely taken up with some activity" (Dewey, 1913, p. 17), and has been conceptualized as either a temporary state or an enduring trait. In particular, situational interest is "spontaneous, transitory, and environmentally activated" whereas individual interest is "of enduring personal value and activated internally" (Schraw, Flowerday, & Lehman, 2001, p. 211). Both the structural and content features of text contribute to situational interest. Structural features include texts that are visually appealing, coherently organized, and full of novel and surprising content (Hidi, 1990; Schiefele, 1999). Content features such as life themes can also foster situational interest (Hidi, 1990). Situational interest is an important resource for educators whose students do not have pre-existing interest in science or social studies content (Ainley, Hidi, & Berndorff, 2001; Hidi & Harackiewicz, 2000). Interesting texts can spark young children's situational interest, sustaining positive affect and the persistence to learn from complex informational texts (Ainley et al., 2001; Kulikowich & Hepner, 2018).

Second, content literacy instruction emphasizes mastery goals that foster children's motivation to learn from texts. The motivation to read is facilitated by an individual child's mastery goals, self-competence beliefs, and situational interests (Conradi, Jang, & McKenna, 2014). In intervention programs like CORI and IDEAS, teachers create classroom structures that emphasize thematic teaching, knowledge goals, and conceptual development (Guthrie et al., 2006; Romance & Vitale, 2001). Together, these classroom structures help students connect new knowledge with previous knowledge and perceive progress in their understanding (Meece, Anderman, & Anderman, 2006). Classroom goal structures that foster a mastery goal orientation encourage students to focus on acquiring content knowledge from interesting texts, to enjoy the intrinsic value of the task, and to develop the persistence needed to continue learning from challenging texts (Guthrie et al., 2006; Hulleman, Durik, Schweigert, & Harackiewicz, 2008; Conradi et al., 2014). A reader's self-competence beliefs and task values are thought to be malleable factors that are sensitive to the kinds of texts and tasks afforded children during content literacy instruction (Guthrie, Wigfield, & You, 2012; Unrau et al., 2018).

Third, content literacy instruction emphasizes the role of children's task orientations, including behaviors such as attention, persistence, and concentration (Fredricks, Blumenfeld, & Paris, 2004; Guo, Connor, Tompkins, & Morrison, 2011; Lepola et al., 2016). The concept of task orientations emphasizes a child's tendency to accept and overcome challenging learning tasks (Lepola et al., 2016). Because content literacy instruction is designed to help students achieve challenging mastery goals for learning, acquire deep conceptual understanding, draw inferences from text, and use reading and writing as tools for understanding of disciplinary knowledge (Ainley et al., 2001; Graham et al., 2017; Lepola et al., 2005), it affords numerous opportunities for children to develop and exercise attentional and behavioral control. In other words, task orientations are an observed child behavior that sustain learners' engagement with challenging literacy tasks and foster reading comprehension gains, above and beyond decoding and language skills (Cartwright & Guajardo, 2015; Lepola et al., 2005; van de Sande, Segers, & Verhoven, 2013).

Making Content Literacy Effective in First Grade

Content literacy instruction affords first graders opportunities to become engaged readers who can acquire domain knowledge from complex informational texts. At the same time, however, content literacy may present certain challenges for young children. First graders may lack the automatic word reading skills and fluency to focus attention on higher-order text comprehension processes (Williams et al., 2016), have limited knowledge of domain specific vocabulary needed to understand expository text (Kintsch, 1998, 2005; Neuman, Newman, & Dwyer, 2011), and/or have working memory capacity constraints that make reading and writing about content topics difficult (Berninger & Abbott, 2010). Furthermore, content literacy instruction typically emphasizes knowledge acquisition over instruction in basic word reading skills, leaving open the possibility that there are opportunity costs. In other words, there may be unintended adverse effects on student outcomes if early grade content literacy instruction de-emphasizes basic literacy instruction.

In addition, there is emerging evidence that content literacy instruction and academic language interventions may contribute to Matthew effects (Stanovich, 1986) in which treatment effects are larger for higher-performing than lower-performing children (Coyne et al., 2010; Penno, Wilkinson, & Moore, 2002). To date, studies in early and middle-grade literacy suggest that initially higher-performing students enjoy larger achievement gains than lower-performing students (Connor et al., 2017; Faggella-Luby, Graner, Deshler, & Drew, 2012; Herman et al., 2015). In content literacy programs that do not show evidence of Matthew effects, there is a strong emphasis on teacher-managed instruction that is explicit, direct, and strategic (Connor et al., 2009; Pressley, 1998; Swanson & Hoskyn, 1998; Williams et al., 2016) and helps students organize and integrate new content into existing knowledge structures.

Research Supporting the MORE Instructional Components

In developing the Model of Reading Engagement (MORE), we first identified existing models of content literacy instruction that emphasized the contribution of domain knowledge acquisition and reading engagement to text comprehension (Alexander, 2018; Alexander et al., 1995; Cromley & Azevedo, 2007; Guthrie & Wigfield, 2000). We were also influenced by intervention programs that represent varied instantiations of content literacy instruction in the domain of science (Connor et al., 2017; Guthrie et al., 2004; Mantzicopoulos, Patrick, & Samarapungavan, 2013; Romance & Vitale, 2001).

A core component of content literacy instruction is the use of thematic units that provide an overall intellectual framework that helps students connect new learning to a meaningful schema, or knowledge structure, and pursue mastery goals for acquiring domain knowledge. We designed MORE at school as a 10-day thematic unit on the topic of Artic animal survival. Teachers integrated three practices to facilitate domain knowledge acquisition (conceptually-related science texts, concept mapping, argumentative writing) with two practices to facilitate reading engagement (read-alouds and discussion, collaborative research). In addition, we tested a supplemental component (MORE at school plus home) involving student choice of three informational books to read at home with their parents (Mantzicopoulos, Patrick, & Samarapungavan, 2013). We provide a brief summary of the research pillars for each instructional component below.

MORE Components to Foster Domain Knowledge

Conceptually related science texts that provide repeated exposures to semantically related vocabulary. There is clear evidence that recurrent exposure to target words in semantically meaningful contexts and texts is critical to fostering students' vocabulary knowledge depth (Snow, Lawrence, & White, 2009). Cervetti and colleagues (2016) have suggested that conceptually-related science texts "build substantive understanding of a set of concepts through repetition, elaboration, and examples provided across the texts" (p. 746). Their study indicated that fourth-grade students who read a series of conceptually-related science texts about birds for three consecutive days had greater knowledge of both concept words and transfer to general academic words (*ES* = .49) compared to control students who read conceptually-unrelated texts (Cervetti et al., 2016). Ultimately, depth of vocabulary may serve as a proxy for having a rich network of domain knowledge that helps students read and learn from text (McKeown, Deane, Scott, Krovetz, & Lawless, 2017; Neuman et al., 2011). Selecting coherent texts in thematically organized lessons may help students (a) develop conceptual understanding of semantically related words (Cervettti et al., 2016; Fitzgerald, Elmore, Kung, & Stenner, 2017; McKeown et al., 2017; Read, 2004), (b) integrate words into text comprehension processes about domain specific texts (Bos & Anders, 1990; Hirsch, 2016; Perfetti, 2007; Perfetti & Stafura, 2014), and (c) build domain knowledge needed to generate relevant ideas in reading and writing (Graham et al., 2017). In the MORE lessons, we chose informational texts related to animals in the Artic ecosystem because they afforded multiple exposures to target vocabulary words related to the topic of animal survival. We chose this topic because it aligned with the science standards in the state where we conducted this study (as described in the Methods).

Concept mapping to make connections among related vocabulary. A large body of research indicates that children who can extract new ideas from text and connect those ideas to prior knowledge show an advantage in learning from complex texts (Brown, Roediger, & McDaniel, 2014). Meaningful learning (Ausubel, 1968) occurs when children can "hook" new information into previously learned content. A concept map is a useful visual tool for organizing and representing the structure of domain knowledge so that children's learning of new knowledge is connected to what they already know (Novak & Gowin, 1984; Novak, 1990). The ability to instantiate a concept map can serve as a strong indicator of the quality and depth of students' vocabulary knowledge (Anderson & Freebody, 1981). A concept map, or a knowledge map, is a graphical tool for organizing conceptual knowledge (Nesbit & Adesope, 2006). Concept maps can further facilitate learning if they include pictures that are paired with the words and help learners process visual and verbal information (Paivio, 1986).

Importantly, concept mapping may help to reduce Matthew effects in literacy outcomes. They are visual aids with brief labels, pictures, and node-link-node syntax to represent ideas to build a meaningful schema of texts. Such schemas may already exist for knowledgeable students and proficient readers. Therefore, they may be more helpful for low-ability readers and less-knowledgeable students who are just beginning to organize and make connections among related vocabulary. For example, a meta-analysis of concept mapping (Nesbit & Adesope, 2006) revealed significantly large effects for students with lower domain knowledge and verbal ability (*ES* = .40) than higher-skilled students (*ES* = .13). In the MORE lessons, teachers introduced concept mapping as a tool for helping children organize, remember, and use taught words in daily oral language, reading, and writing activities.

Strategic support for argumentative writing. Content literacy instruction provides students with opportunities to use reading and writing as tools for extending their understanding of science (and social studies) content. A recent meta-analysis of reading and writing during content literacy revealed modest to large effects on both student reading and writing outcomes in grades 3 to 8, but no extant evidence on effectiveness for first graders (Graham et al., 2017). Given this research gap, it is critical to understand if young children who are developing basic decoding and encoding skills, as well as executive functions to plan and coordinate their writing, can similarly benefit from writing about what they read (Ahmed, Wagner, & Lopez, 2014).

Because young children have working memory constraints and difficulty regulating the writing process (McCutchen, 2000), we drew upon research on the self-regulated strategy development model of writing to support argumentative writing instruction (Boscolo & Gelati, 2007; Ferretti & Lewis, 2013; Graham & Harris, 1989). In particular, the TREE strategy directs students to include a **T**opic sentence, provide **R**easons for their opinion, **E**xamine each reason from the audience's perspective, and provide an **E**nding. The TREE strategy has shown strong evidence of improving student's argumentative writing and self-efficacy (Graham & Harris, 2005). Meta-analytic evidence also indicates that such strategy instruction is equally effective in helping students with learning disabilities and general education students exert the attention, concentration, and focus to plan and write effectively (Graham, Harris, & Santangelo, 2015).

The role of writing in learning, specifically including the role of writing to learn in the science discipline, has been well established (e.g., Britton, Burgess, Martin, McLeod, & Rosen, 1975; Keys, 1999; Rivard, 1994). A strong positive relationship exists between knowledge in general (and topic knowledge in particular) and writing performance (e.g., Albin, Benton, & Khramtsova, 1996; Langer, 1984; Olinghouse, Graham, & Gillespie, 2015). Moreover, students' domain knowledge and argument structure knowledge have been shown to be strongly positively correlated (e.g., Olinghouse et al., 2015).

Composing in scientific genres tends to promote instantiation of science knowledge, and it promotes reflection and production of new knowledge (e.g., Keys, 1999). The primary reason that writing aids learning is that language does not only describe pre-existing mental thought and concept structures. Language use can also support creation of such mental structures (e.g., Halliday & Martin, 1993). For instance, argument structure has been considered a common structural form in science (e.g., Kuhn, 1993; Bell, 2004). As children learn to write an argument about a science topic, they are guided by mental search and retrieval of relevant content from long-term memory. As they learn about, and use, the argument structure while composing (as they acquire argument discourse structure knowledge), the mental instantiation of the topic knowledge may be structurally transformed and enriched through associated details. For these reasons, writing argument structure for learning science topics was incorporated into MORE lessons.

MORE Components to Foster Reading Engagement

Interactive read-alouds and discussion. A key aim of the MORE intervention was to embed oral language activities in motivating texts. Teacher read-alouds of informational texts can also nurture young children's topic interest and task values (Marinak, Malloy, Gambrell, &

Mazzoni, 2015) and persistence to learn from complex texts (Ainley et al., 2001; Kulikowich & Hepner, 2018). Existing research indicates that interactive read-alouds and discussion are a critical ingredient of personalized literacy interventions that tailor instruction based on children's code and language skills (Connor et al., 2009; Mol, Bus, & de Jong, 2009). In other words, interactive read-alouds and discussion have robust main effects on young children's reading development.

Peer-mediated, collaborative research. To further enhance engagement in complex science texts and discussion of science content, MORE lessons included a large number of "open" tasks—i.e., tasks for which there was not one right answer (Parsons, Malloy, Parsons, & Burrowbridge, 2015), as well as opportunities for student choice and collaboration around these tasks. For example, students chose one Artic animal (snowy owls, artic foxes, lemmings, narwhals) to study in greater depth as part of a collaborative research group. Students worked with their peers to conduct research on their animals and address the question – how does the animal survive in its habitat? Students also discussed and wrote about the question: Could *you* survive in the Arctic?

Research groups were designed to foster engaging tasks that emphasized peer-mediated collaboration, challenging sentence and paragraph-level writing, and student-directed activities that were sustained for three or more days (Parsons et al., 2015; Guthrie et al., 2006). Collaborative social structures foster students' feelings of belongingness and enhance task orientations, that is, students' ability to focus, concentrate, and overcome challenging tasks like using textual evidence in writing (Lepola et al., 2016).

Student choice of conceptually related information books to read at home. To further support readers' motivation, we adapted a home reading component in a science content literacy program (Mantzicopoulos, Patrick, & Samarapungavan, 2013) that involved giving children a choice of three informational books. Students were allowed to choose three books and parents were encouraged to engage in shared reading activities at home.

The Present Study

In developing the instructional components in the MORE lessons, our goal was to move as far away as possible from typical instructional practices in first-grade reading. To do so, we interleaved instructional components that were designed to supports first graders' science domain knowledge, reading engagement, and reading comprehension. We ask four research questions:

1. What is the impact of MORE on proximal measures of first graders' science domain knowledge?
2. What is the impact of MORE on proximal measures of first graders' reading engagement?
3. What is the impact of MORE on distal measures of first graders' reading comprehension and basic literacy skills?
4. Is there evidence of treatment-by-student characteristic (i.e., initial reading comprehension ability) interaction effects on posttest outcomes?

Methods

Research Design and Participants

We designed this study as a cluster randomized controlled trial (RCT) in which first-grade classrooms nested within participating schools were randomly assigned to experimental conditions. Our chief goal was to generate internally and externally valid causal estimates of MORE on proximal and distal student outcomes. In this study, the inference population included the district's 94 K-5 schools that were organized by 10 geographic regions called learning communities. To maximize external validity, we used stratified probability sampling to select schools for a RCT that would allow us to generalize findings to our defined inference population (Hedges, 2013). Within each learning community, all K-5 schools were assigned a random number and schools with the highest random number were invited to participate in the study. Ten schools of the fourteen initially contacted agreed to participate in the RCT, leaving 84 non-RCT schools. To evaluate the representativeness of the target population and generalizability of our findings, we sampled these non-RCT schools in addition to RCT schools and compared demographic characteristics of the samples from both RCT and non-RCT schools. To enhance internal validity, we randomly assigned 38 first-grade classroom teachers nested with the 10 schools to (a) MORE at school (MS), in which teachers taught 10 MORE lessons on Arctic animal survival, (b) MORE at school plus home condition (MS-H), in which students received three free books to take home in addition to the MS lessons, or (c) typical instruction (TI), in which teachers taught their normal literacy program. Students were administered a posttest after a one-week spring break to assess implementation and impact of the MS-H condition.

Because teachers in the experimental conditions replaced TI with MORE lessons, we inferred that the core components rather than additional time contributed to any posttest differences. In other words, our goal was to avoid confounding the treatment with additional instructional time by requiring all teachers to teach either MORE or TI lessons during the three-week implementation window. Thus, any differences in student outcomes can be linked to differences in the quality rather than the quantity of literacy instruction during Tier 1 instruction.

Table 1 displays demographic characteristics of students by treatment condition in the RCT sample and the non-RCT sample. Importantly, there were no statistically significant differences on each of the four measures of student demographics (i.e., gender, race/ethnicity, English language learner status, special education status) by condition. There was also no significant difference between children in the RCT sample and non-RCT sample on each of the demographic measures. The similarity between the RCT and non-RCT sample on observable student characteristics enhances the generalizability of our findings to our inference population.

First-grade classroom teachers in our study were all certified, had a range of teaching experience (M = 10.24 years as a classroom teacher; median = 7 years; range 1-22 years), and were expected to follow the district's scope and sequence for first-grade reading. Typical instruction in first grade was based on a balanced literacy program including word study, independent reading, guided reading, and writing activities in small group, teacher-directed instruction, and independent reading. To implement the balanced literacy program, teachers were provided Pinnell and Fountas's *Continuum of Literacy* (2007), *Words Their Way: Word Study for Phonics, Vocabulary, and Spelling Instruction* (Bear, Templeton, Johnston, & Invernizzi, 2012), and Reading A-Z texts that include a variety of narrative and informational texts. All teachers in the state where we conducted this study were also required to administer mCLASS: Reading 3D and Text Reading Comprehension (TRC) tasks that provided data on students' nonsense word reading fluency, oral reading fluency, and text reading levels. During the three-week MORE implementation window, the district literacy calendar had TI teachers leading a fiction unit in their readers workshop focused on retelling, asking and answering questions, and identifying who is telling the story at various points in time. In their writing workshop, TI teachers were leading a unit on persuasive writing.

Procedures

Study preregistration and combining treatment conditions. The study design was pre-registered in a registry for RCTs prior to program implementation (Kim, 2018). Hence, the details on the methods and data analytic plan were pre-specified prior to program implementation and the receipt of posttest student data. In our original design, we included a home component (i.e., MS-H) in the MORE-at-school (i.e., MS) condition. There were minimal differences in the procedures and activities in the two conditions. Students in the MS-H component chose three books to read and were assigned a homework activity, but only 1% of the students completed and returned their homework assignments (completed comprehension questions and read with their parent). In addition, both MS and MS-H conditions were not significantly different from each other in pretest reading comprehension, $F(1, 317) = .17, p > .05$, and basic literacy skills, $F(1, 423) = .87, p > .05$, as well as posttest outcomes, $F(1, 401) = .84, p > .05$, and, $F(1, 429) = 2.60, p > .05$, respectively. Because students in the two experimental conditions received essentially similar treatments and because there were no differences in pre and posttest outcomes by condition, we simplified our main analysis and presentation of the results by combining the two conditions (MS and MS-H).

MORE Intervention Description

Professional development. Prior to program implementation, MORE teachers participated in a two-hour afterschool professional development workshop to learn about the MORE program theory and to review the MORE lesson materials. All lessons were fully scripted to provide teachers with a clear example of what the lesson could sound like. Teachers were told, however, that they could "use their own teacher voice," as long as they taught all of the lesson's core components. Teachers also received ongoing support from their school literacy facilitators during the implementation period. The research team provided two personal, 30-minute planning calls to each school literacy facilitator—one at implementation outset and another toward the end of the implementation period. The goal of these calls was to prepare literacy facilitators to support teachers as they rolled out the MORE lessons and administered student assessments, respectfully. Additionally, research team members visited each school at least once during the implementation period to distribute instructional materials, support the data collection process, discuss MORE teachers' questions and concerns, and address challenges.

MORE theme and schedule. MORE lessons consisted of one unit on the life science topic of Arctic animal survival. The unit was 10 lessons long and each lesson was designed to take about 60 minutes. Teachers were given flexibility as to exactly when within a 3-week implementation window to teach the lessons in spring 2018 (March to April). MORE teachers followed a 10-day lesson sequence in which the five practices were interleaved, as described in Figure 1.

Procedures for selecting domain specific vocabulary. We used an iterative procedure to identify domain specific vocabulary in the MORE lessons. First, we anchored the words to the state's Grade 1 Science standards for ecosystems (1.I.1: Understand characteristics of various environments and behaviors of humans that enable plants and animals to survive) and the Next Generation Science Standards (National Research Council, 2012) for life science (Core idea LS2.A: Interdependent relationships in ecosystems). We conducted a content analysis of each standards documents as well as our lesson texts to identify the related vocabulary: *adapt/adaptation, advantage, behavior, endangered, extinct, habitat, physical feature, shelter, species, survive.*

Second, we cross-validated these words against content standards that predated the state standards and Next Generation Science Standards (NGSS) to ensure that the words were relatively stable features of U.S. school curricula over time (Hirsch, 2016). To do this, we examined whether the words appeared in Marzano's (2004) analysis of primary instructional concepts—i.e., domain-specific vocabulary—appearing in national science content standards from 1993 to 2000 (American Association for the Advancement of Science [Project 2016], 1993; Council for Basic Education, 1998; Kendall & Marzano, 2000; National Research Council, 1996; National Science Teachers Association, 1993).

Third, we cross-validated our words against an automated concept network generation procedure using data from the top four best-selling science textbook programs in grades 1-5. Specifically, we created a concept network for each target concept (i.e., domain specific MORE vocabulary) containing the target word and additional associated words also appearing in the lesson texts. Each word was represented as a node with weighted connections between nodes indicating the degree of similarity. For example, for the target concept ‘adaptation’, the top five most closely associated words were: reproduce, vertebrate, camouflage, survive, and mimic. Reviewing the concept maps for all target concepts enabled us to identify five polysemous academic words (i.e., diversity, resource, complexity, potential, unique) that appeared in our lesson texts and that were related to target concepts, but were not directly taught in the MORE lessons, particularly in the concept mapping activity. We chose five words that were unlikely to be known by first graders since the age of acquisition was above 8.5 years (unique = 8.5, potential = 9.61, resource = 10.00, diversity = 10.79, complexity = 12.16). We included these five words as far transfer items on the vocabulary depth measure.

Daily lesson activities. In lessons 1 and 2, teachers introduced the first informational text with the target science words and instructed children how to organize the words into a concept map. The goal of the first two MORE lessons was for students to know who would win in a fight: a polar bear or a grizzly bear. The teacher conducted an interactive read-aloud of *Who Would Win? Polar Bear vs. Grizzly Bear* (Pallotta, 2010). Students were also explicitly taught the following key science concepts: *survive*, *physical feature*, *behavior*, and *advantage*. For each concept, the teacher made a connection to the interactive read-aloud and gave students a child-friendly definition (e.g., a *behavior* is something that an animal does; see Appendix A). Students were asked to say the concept word with the teacher, trace the letters to spell the concept, and copy the definition in their MORE research notebook along with characteristics and/or examples associated with the concept (e.g., dig a den, wait quietly). Finally, the teacher led students through a brief activity that asked them to actively work with the new concept (e.g., "stand up when you hear me say a polar bear behavior"). Concepts were organized in a large "class concept map" that was given a prominent position in the classroom. Appendix B displays a MORE concept map for this unit.

In lessons 3 through 5, teachers used new informational texts connected to the theme of Arctic animal survival during interactive read-alouds, provided repeated exposures to target words, and introduced the argumentative writing activity. Thus, the chief goal of the remaining lessons was for students to leverage their emerging domain and topic knowledge in becoming Arctic animal experts. MORE teachers continued to build students' topic knowledge on the Arctic—and specifically polar bears in the Arctic—through interactive read-alouds of books (*Where do Polar Bears Live? Polar Bears and the Arctic*). In each lesson, the teacher either introduced a new key concept to students or engaged them in an activity to keep them working with the concepts, for example, having groups decide where to place an image on the class concept map or having a group sort words and pictures on a concept map. In these lessons, the teacher also introduced an argumentative writing strategy called "A-TREE" (Graham & Harris, 2005) and students used A-TREE to discuss and write an argumentative response to the open-ended question: *Could you survive in the Arctic?*

In lessons 6 through 10, teachers continued to implement the core MORE practices and also added collaborative research activities. Thus, the goal in lessons 6-9 remained the same but teachers organized students into leveled research groups and, rather than the whole class researching polar bears, each group researched a different Arctic animal: lemming, snowy owl, Arctic fox, or narwhal. Students continued to work with the A-TREE strategy. This time, they used A-TREE to engage in a discussion and write an argumentative response to the open-ended question: *Should people who live in the Arctic be allowed to hunt and kill seals?* Students shared their expertise on the Artic animal they studied, and students in the MORE at school plus home classrooms also chose three books and homework activities to complete during spring vacation.

Fidelity of Implementation

Teachers audio-recorded all of their literacy lessons during the MORE implementation period. We then used these data to assess (a) adherence to the MORE lessons and (b) program differentiation between MORE lessons and TI read-aloud texts and tasks used during literacy instruction.

Adherence to MORE components. To assess teachers' adherence to the MORE lessons, we developed and applied an adherence checklist to recordings of lessons from the beginning (lesson 2), middle (lesson 5), and end (lesson 8) of the unit. Each checklist captured the extent to which teachers enacted the researcher-identified "core components" of the lessons. Core components are those parts of the lessons that are unique and essential to the intervention (Munter, Wilhelm, Cobb, & Cordry, 2014) and/or are backed by research evidence as important to achieving lesson objectives (McMaster et al., 2014). For example, the core components in lesson 5 included: concept mapping, read-aloud and discussion, argumentation, and lesson closing. Each core component was further operationalized by establishing its essential elements, which reflect the theory underlying the component and the teacher's responsibility in leading that component.

Coders rated the presence or absence of 21 essential elements in lesson 2, 22 essential elements in lesson 5, and 21 essential elements in lesson 8 for all 24 treatment classrooms. Adherence was calculated per lesson by summing up the total number of essential elements present in each lesson recording, dividing that number by the total number of essential elements for that lesson, and multiplying by 100 to yield a percentage score. Overall adherence was calculated by summing up the total number of essential elements present across all recordings, dividing that number by the total number of essential elements across all recorded lessons, and multiplying by 100 to yield a percentage score. All lessons were initially coded by one of the authors and then a randomly selected 20% subset of the lesson recordings was independently double-coded by two research team members to determine interrater reliability. Agreement ranged from 92.19% (Cohen's κ = .51) to 96.88% (Cohen's κ = .88). Overall, the mean adherence rate across MORE classrooms was 74%, suggesting that fidelity was acceptable and comparable to other content area literacy interventions (cf. Vaughn et al., 2013; Williams et al., 2016).

Program differentiation in read-aloud texts. To assess differences in text complexity of read-aloud texts used in MORE and TI classrooms, three members of the research team listened to all recordings of TI lessons and coded the title of the books used in the TI read-alouds. Table 2 displays the title list of informational and narrative reads-aloud books used in MORE and TI instruction classrooms. There were five unique titles in the MORE lessons and 20 unique titles in TI lessons. TI read-aloud texts also included more narrative than informational texts.

Next, we used the CCSS's approach for assessing text complexity that included both quantitative and qualitative dimensions (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010). Lexile measures of texts were used as the quantitative measure of complexity. For the qualitative measures of text complexity, we created a 3-point Likert scale assessing the following text features: (a) level of meaning or purpose, (b) text structure, (c) text features/illustrations, (d) language conventionality and clarity, (e) content knowledge demands, (f) cultural knowledge demands, and (g) vocabulary knowledge demands.

There were significant differences on the quantitative measure of text complexity using the Lexile measure. As shown in Figure 2, when only informational books were compared, the average Lexile score for texts used in MORE conditions (M = 673L) was higher than in TI (M = 500L). There were also significant qualitative differences in text complexity. As shown in Figure 3, read-aloud books used in MORE instruction were significantly more challenging than those in the TI condition on the level of meaning or purpose, text structure, text features/illustrations, language conventionality and clarity, content knowledge demands, cultural knowledge demands, and vocabulary knowledge demands.

Program differentiation in literacy tasks. The literacy tasks assigned to students in TI were also assessed for the degree of engagement as compared to the MORE literacy tasks. Using the audio recordings of MORE and TI lessons from the beginning (lesson 2), middle (lesson 5), and end (lesson 8) of the unit, three research assistants first documented the classroom literacy tasks and identified 12 to 42 unique literacy tasks from the MORE and TI lessons, respectively (see Appendix C). Then, we assessed the openness of literacy tasks—that is, the extent to which each literacy task fostered authenticity, collaboration, challenge level, student-directed work, and sustained effort (Parsons et al., 2015). The aforementioned five components of each literacy task were rated on a 3-point Likert scale by a research team member: closed = 1, moderately open = 2, and open task = 3. The use of more open-ended literacy tasks was an indicator of promoting student engagement and student-centered learning than closed tasks (Duke, Purcell-Gates, Hall, & Tower, 2006; Guthrie & Humenick, 2004). Twenty percent of the entire literacy tasks identified were then double coded by the second research team member to determine interrater reliability. Overall agreement was 94.8% (Cohen's $\kappa = .67$).

There were significant differences in the types of literacy tasks in MORE and TI classrooms. Table 3 presents the results of an independent t-test that compared the openness of literacy tasks in MORE and TI lessons. The comparison analysis using revealed that the degree of authenticity, collaboration, challenge, student-directed, and sustainability in the MORE literacy tasks were significantly greater than the TI tasks (all $ps < .001$), such that the MORE literacy tasks were more open and likely to foster student engagement (Parsons et al., 2015) compared to TI tasks.

Student Measures

Reading comprehension. The Measure of Academic Progress (MAP) Primary Grade Reading is a computer-adaptive, early literacy assessment that uses an interval scale, called the Rasch unit (RIT) scale score, to capture student growth in reading comprehension from kindergarten to second grade. The MAP yields a total reading score and subtest scores for each of the four strands that comprise the assessment. The literature and informational strand is designed to assess children's understanding of 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 vocabulary use and functions strand assesses children's ability to determine the meaning of new and unknown words in context, to analyze word parts, and to understand figurative language. The foundational skills strand assesses children's ability to apply phonics skills in decoding words: their ability to isolate, hear, and manipulate sounds within words. Finally, the language and writing strand assess children's understanding of the conventions of English capitalization, punctuation, spelling, and grammar. Performance on the four strands yields an overall RIT score which was used for this analysis as a pretest covariate and posttest outcome measure. The MAP has a reported test-retest reliability from .89 to .96 (Brown & Coughlin, 2007).

Basic literacy skills. The mCLASS Dynamic Indicators of Basic Early Literacy Skills (DIBELS) assesses several early literacy skills from kindergarten through sixth grade. The K-3 DIBELS includes the following subtests: sound fluency, phoneme segmentation fluency, letter naming fluency, nonsense word fluency, oral reading fluency, and retell abilities (Good & Kaminski, 2002). We used a composite score that combines subtest scores for first grader's nonsense word reading fluency (correct letter sounds and whole words read), oral reading fluency, and retell ability. Evidence of predictive validity is based on the moderate correlations with standardized tests of reading comprehension (e.g., $r = .73$ between DIBELS composite and the Group Reading Assessment and Diagnostic Evaluation reading test). Reliability estimates (alternate-form, test-retest, and inter-rater) of the composite ranged from .88 to.98 across grades. Assessments of validity (content, criterion, and discriminant) with other reading assessments for separate reading components and the composite indicated that the results were at appropriate levels (University of Oregon, 2012). The mCLASS DIBELS composite scores have a reported first-grade test-retest reliability above .90.

Science domain knowledge: Vocabulary knowledge depth. We developed a 12-item semantic association task to assess students' vocabulary knowledge depth of taught science words and their ability to identify relations between the target word and other known words (Collins & Loftus, 1975; Schmitt, 2014; Stahl & Fairbanks, 1986; see Appendix D). We adapted the semantic association task (Read, 1998; 2004) for our study to assess first graders' ability to identify semantically related words and their knowledge of how words are networked to each other. The task included seven domain specific words taught in the MORE lessons (*survive, species, behavior, advantage, adaptation, habit, physical feature*) and five associated words that were not directly taught in the MORE lessons (*potential, unique, resource, diversity, complex*). For each target word, there were four-word options in which one to three words were semantically linked to the target word. Students were prompted to "circle all of the words that go with" the target word. Each item was scored 0 to 4. Reliability coefficients were acceptable for the taught words (Cronbach's α = .85) and untaught words (Cronbach's α = .77)

Science domain knowledge: Listening comprehension. Students listened to a non-fiction passage about an ecosystem that was not taught in either treatment or control classes. The passage was on rainforests and was adapted from the *Magic Tree House Fact Tracker* series (Will Osbourne and Mary Pope Osbourne). The 189-word passage had Lexile level 800L and was complex for most first graders. We created four items to assess students' ability to answer a series of inferential questions that included domain specific vocabulary (see Appendix E). The assessment was administered to a whole class and the passage and questions items were all read aloud to students. The reliability (Cronbach's α) for this measure was .35.

Science domain knowledge: Argumentative writing. The writing assessment was designed to capture students' science domain knowledge about rainforests and their knowledge of the argument genre/structure. The directions prompted students to address an open-ended writing prompt: “Should people be allowed to cut down trees in the rainforest?” The directions prompted children to “answer this question by making an argument,” encouraged them to take three minutes to plan or think about what they might say, and reminded them of the components of a good argument (says your opinion, says your reasons, explains your thinking, has a conclusion).

We consulted with existing rubrics for assessing expository and argumentative writing (Graham, Harris, & Hebert, 2011; McNeill, 2011; Wang, Matsumara, & Correnti, 2017) and received feedback from five leading experts who were involved in writing research. Based on our review of prior research and expert feedback, we undertook several steps to assess student writing. First, we transcribed the students’ responses. During this process, clear misspellings were corrected, and if research assistants could not determine the spelling of a word, “XXX” was inserted in its place. The purpose of this process was to reduce presentation bias stemming from poor handwriting skills (Graham et al., 2011) and to focus on scoring three dimensions of argumentative writing. This rubric was adapted from previous research on elementary grade students’ expository writing (McNeill, 2011; Wang et al., 2017).

Students’ domain knowledge as evidenced during argumentative writing was scored on three dimensions: claim, evidence, and ending. A claim was defined as a statement that attempts to answer the original question and assigned a score of 0 (does not make a claim), 1 (takes a position, either yes or no, but does not make clear whether they are arguing for or against), or 2 (provides an appropriate claim that makes clear that they are either for or against people being allowed to cut down trees).

While stating a claim is critical to argumentation, a student's choices about how to support their claim reveal more about their pre-existing knowledge of rainforests and/or their ability to extract relevant knowledge from source text. To more systematically assess the extent to which students extracted knowledge from the source text, we divided the text into 11 "concept units" or discrete pieces of information about rainforests.[1] Evidence use was scored based on the extent to which students' appropriately and sufficiently supported their claim. Students received a score of 1 for evidence if they included any concept unit from the source text but the concept unit did not support their claim. Students received a 2 if they used knowledge that was relevant to support their claim but not found in the source text. Students received a 3 if they used at least one relevant concept unit from the source text to support their claim. They received a 4 if they used more than one relevant concept unit to support their claim. Finally, the ending was scored 0 or 1 (present or not present).

A total writing score was calculated for each student by summing up their claim, evidence, and ending scores; thus, a student's use of evidence was weighted more heavily than either claim or ending. Taken together, all three scores reflect students' science domain knowledge and argument structure knowledge. Relatively high inter-rater agreement (Cohen's $\kappa = .74$ for a total writing scores and Cohen's $\kappa = .90$ for agreement within one score-point) was at a level consistent with intervention studies involving ratings of student writing in the elementary grades (Coker et al., 2016; Collins, Lee, Fox, & Madigan, 2017).

Engagement: Situational interest. Over the 10-day lesson cycle, students completed a 3-item situational interest measure that asked how they felt about the texts used for read-alouds (feeling-related valence). Each item was on a 3-point scale (feelings about the read-aloud book: felt great, ok, boring; feelings about receiving the book as a gift: mad, okay, happy; feelings about self during the lesson: not a good reader, ok reader, very good reader). For the MORE students, the title of the texts focused on the read-aloud books following lesson 3 (*Who Would Win: Polar Bear vs. Grizzly Bear*), lesson 5 (*Where do Polar Bears Live*), and lesson 8 (the book chosen for the Arctic animal research group on snowy owls, Arctic foxes, narwhals, or lemmings). A sample of situational interest measure for the MORE students is provided in Appendix F. For the TI condition, students were asked, "When your teacher read a book or article out loud to you today, did you think it was Great, Okay, or Boring?"

After lessons 3, 5, and 8, students in both conditions completed item 1 (i.e., how they felt about the read-aloud book), item 2 (i.e., how they would feel if they received the book as a gift), and item 3 (i.e., how they felt during the lesson). All items were read aloud to students. Combining each item across the three lessons yielded reliabilities (Cronbach's α) of .59, .58, and .62, respectively.

Engagement: Reading motivation. The Me and My Reading Profile (MMRP; Marinak et al., 2015) is a K-2 student-self report measure of reading motivation and includes 20 items (see Appendix G). Reliabilities for the 5-item self-concept scale (Cronbach's α = .86), the 10-item valuing of reading scale (Cronbach's α = .87), and the 5-item literacy out loud scale (Cronbach's α = .87) were adequate and validated previous factor analytic results highlighting the existence of these three subscales of motivation (Marinak et al., 2015). The MMRP was administered to an entire class simultaneously and all items were all read aloud to students.

Engagement: Task orientations. Teachers assessed students' task orientations using a 5-item measure used in a previous study of first-grade literacy development (Lepola et al., 2005). After the implementation of the MORE lessons, teachers rated the following student behaviors: (a) concentration on the task, (b) showing persistent effort when facing difficulties (i.e., not giving up easily), (c) becoming absorbed in the given task, and (d) being eager to do tasks that exceed one's competence. Each item was scored on a 1-5 scale (1 = *the behavior does not occur at all*, 2 = *very seldom*, 3 = *sometimes*, 4 = *does occur often*, 5 = *does occur very often*). Reliabilities (Cronbach's α) for the first-grade task-orientation measure was .85.

Data Analysis

Given the clustered nature of the data (students nested within classrooms and schools), we used hierarchical linear models (HLMs; Raudenbush & Bryk, 2002) to examine the MORE intervention effects on student outcomes. We first estimated a fully unconditional three-level model with random intercepts and random residual errors for each outcome to compute the intraclass correlation coefficients (ICC) by partitioning the total variance in dependent variables into three components: between students within classrooms (Level 1), between classrooms within schools (Level 2), and between schools (Level 3). In fitting the subsequent conditional HLMs, we fitted three-level models when school-to-school variation accounted for a relatively large proportion of the total variability in dependent variables (Level-3 ICC above the recommended threshold of .05; Kenny, Kashy, & Cook, 2006) to account for the school-level dependency in the data. However, for the dependent variables that were explained by negligible school-level variance (Level-3 ICC below .05), we conducted two-level HLMs including students at Level 1 and classrooms at Level 2, ignoring school-level effect.

The composite specification of the full two- and three-level HLMs is expressed as follows:

Two-level model: $Y_{ij} = \beta_{00} + \beta_{01}(MORE)_j + \beta_{10}(PRE)_{ij} + \beta_{11}(PRE \times MORE)_{ij} + \sum_{p=2}^{8} \beta_{p0}\,(COV)_{pj}$ $+ e_{ij} + r_{0j}$,

$$e_{ij} \sim \mathrm{N}(0, \sigma_{ij}^2),\ r_{0j} \sim \mathrm{N}(0, \tau_{00}),$$

Three-level model: $Y_{ijk} = \gamma_{000} + \gamma_{010}(MORE)_{jk} + \gamma_{100}(PRE)_{ijk} + \gamma_{110}(PRE \times MORE)_{ijk} +$
$\sum_{p=2}^{8} \gamma_{p00}\,(COV)_{pjk} + \gamma_{001}(TITLE\_I)_k + e_{ijk} + r_{0jk} + \mu_{00k}$,

$$e_{ijk} \sim \mathrm{N}(0, \sigma^2_{ijk}),\ r_{0jk} \sim \mathrm{N}(0, \tau_\pi),\ \mu_{00k} \sim \mathrm{N}(0, \tau_\beta),$$

where Y represents the respective post-intervention outcome score and the subscripts $i, j,$ and k correspond to students, classrooms, and schools, respectively. Parameters β_{00} and γ_{000} refer to the average student outcomes across all classrooms and schools, respectively; β_{10} and γ_{100} are the main effects of classroom mean reading pretest score on posttest outcome. The MAP reading pretest scores were included in the models of the MAP reading posttest, science content knowledge, and reading engagement, while the DIBELS pretest scores were used in the model of the DIBELS posttest. Predictor COV_{pij} was a vector of student-level covariates—gender, race/ethnicity (i.e., Asian, Hispanic, White, and other), English-language learner status, and special education status—and β_{p0} and γ_{p00} were the effects of corresponding student-level covariates; β_{01} and γ_{010} were the main effects of MORE intervention treatment; and β_{11} and γ_{110} were the cross-level interaction effects involving MORE intervention with reading pretest scores (centered at classroom-level group means), only included in the model for research question 4; γ_{001} refers to the difference in average post-intervention outcome between Title I and non-Title I schools; and residuals e_{ij} and e_{ijk}, r_{0j} and r_{0jk}, and μ_{00k} are Level-1, -2, and -3 random effects, respectively, assumed to be normally distributed with a mean of 0 and variance. All analyses were performed using Stata 15.0 statistical software (StataCorp, 2017).

Finally, we computed an effect size (Hedges, 2007) by taking the parameter estimates for the MORE treatment variable, β_{01} and γ_{010}, respectively, and dividing each estimate by the unadjusted pooled within-group standard deviation. The effect size metric captures the difference between MORE and control students in standard deviation units and facilitates comparison of the magnitude of the estimated treatment effect to other content area literacy interventions.

Missing Data

Missing values existed in pretest reading scores (ranging from 5.5% for DIBELS and 28.2% for MAP reading, respectively) and across posttest reading scores (ranging from 4.3% for DIBELS to 30.9% for MAP reading, respectively). Little's Missing-Completely-At-Random (MCAR) test revealed that missing values were not MCAR ($\chi^2 = 1106.99$, $df = 792$, $p < .001$). With the assumption that data were missing at random (MAR), in which the propensity of missing values was systematically related to the observed but not the unobserved data, we used multiple imputation by simulating 20 data sets with plausible values in place of missing observations. The procedure was performed using a Markov Chain Monte Carlo algorithm with a multivariate normal distribution, for data augmentation (Schafer, 1997), using the *mi impute mvn* command in Stata statistical software (StataCorp, 2017). We also conducted sensitivity analyses to assess the robustness of the findings to different missing data approaches and assumptions—means substitution (see Appendices I and J) and listwise deletion (Appendices K and L).

Results

Descriptive and Correlational Analyses

Table 4 presents (pairwise correlation matrix and) descriptive statistics of pretest and posttest measures by treatment conditions. Analyses of baseline equivalence on reading comprehension pretests revealed no significant baseline differences between the TI ($M = 173.76$, $SD = 15.97$) and MORE ($M = 171.92$, $SD = 15.18$), $t(482) = 1.21$, $p = .23$) groups on the MAP reading pretest or the DIBELS pretest, $t(635) = 0.21$, $p = .84$. Importantly, the correlational analyses revealed consistently stronger correlations between each reading comprehension posttest and measures of science domain knowledge than measures of reading engagement. To build on these descriptive and correlational analyses, we fit a series of multilevel models to estimate the causal effects of MORE on both proximal and distal student outcomes.

Hierarchical Linear Modeling Analyses

Preliminary analysis. The first model to fit across dependent variables (i.e., reading comprehension, basic literacy skills, science domain knowledge, and reading engagement) involved estimating a fully unconditional model with random intercepts to examine whether systemic variation existed within classrooms and schools and between schools. The estimate of ICC indicated that the largest amount of total variance in all outcome variables resided at Level 1 (.63 to .76) and a smaller portion at Level 2 (.13 to .27) and 3 (.00 to .30), which suggests that a significant proportion of total variance in all student outcomes were explained by between-student within-classroom differences rather than within-school or between-school clusters. In particular, the moderate level of variance existed between schools across dependent variables with the exception of one of the reading engagement models—reading motivation—that had little school-to-school variability (ICC = .00). We, therefore, fitted a two-level conditional model for reading motivation without specifying the Level-3 predictor and three-level conditional models for the reading outcomes, science domain knowledge, and task orientation variables.

Research question 1: Main effect of MORE intervention on proximal measures of science domain knowledge. Estimates of the MORE intervention effects on the proximal measures of science domain knowledge—vocabulary knowledge depth, listening comprehension, and argumentative writing—are presented in the left panel of Table 5. First, students in the MORE condition significantly outperformed the TI condition on vocabulary knowledge depth ($\hat{\gamma}_{010}$ = 1.61, *SE* = .67, $p < .05$, *ES* = .30), controlling for student pretests, student demographic characteristics, and school Title I status. These positive effects, however, were driven by gains on students' knowledge of explicitly taught words (*ES* = .56) rather than far transfer words that were not explicitly taught (*ES* = -.06) but part of the concept network. In addition, MORE students had a higher listening comprehension scores than students in the TI students, yielding a moderate effect size ecosystems ($\hat{\gamma}_{010}$ = .45, *SE* = .18 $p < .05$; *ES* = .40). Finally, relative to students in the TI condition, MORE students demonstrated higher performance on the argumentative writing outcome ($\hat{\gamma}_{010}$ = .40, *SE* = .20 $p < .05$, *ES* = .24). Overall, the results indicate that MORE had positive effects on first graders' vocabulary knowledge depth of taught words that were part of the MORE concept network, listening comprehension, and argumentative writing.

Research question 2: Main effect of MORE intervention on proximal measures of reading engagement. The MORE intervention effects on reading engagement measures—situational interest, reading motivation, and task orientations—are reported in the middle panel of Table 5. The MORE condition did not demonstrate a statistically significant effect on the three reading engagement measures at the posttest, controlling for all covariates (all *ps* >.05). We did not find a significant treatment main effect on the three sub-dimensions of reading motivation (i.e., self-concept, valuing of reading, and literacy out loud; all *ps* >.05). Thus, the results imply that first graders in both conditions performed similarly on measures of situational interest, reading motivation, and task orientations.

Research question 3: Main effect of MORE intervention on distal measures of reading comprehension outcomes. As shown in the right panel of Table 5, a significant difference between the MORE and TI conditions, controlling for all covariates, was found in MAP reading posttest, but not in DIBELS posttest. Students in the MORE condition scored significantly higher on MAP reading posttest than students in the TI condition ($\hat{\gamma}_{010}$ = 1.75, *SE* = .83 *p* < .05, *ES* = .11). However, no treatment effect on DIBELS was observed at the posttest ($\hat{\gamma}_{010}$ = 3.70, *SE* = 4.08, *p* > .05). The results indicate that MORE improved children's reading comprehension and did not have adverse effects on the basic literacy skills measured by DIBELS, including nonsense word reading fluency and oral reading fluency.

Research question 4: Treatment-by-student interaction effects on posttest outcomes. Table 6 presents the estimates of treatment-by-student interaction effects on all posttest outcomes. Concerning science domain knowledge, there were significant interactions between treatment and MAP reading pretest on vocabulary knowledge depth of total vocabulary items ($\hat{\gamma}_{110}$ = .07, *SE* = .03, *p* < .05) and taught words ($\hat{\gamma}_{010}$ = .05, *SE* = .02, *p* < .05), such that the MORE intervention effect for students with intially higher MAP reading scores was greater than the MORE intervention effect for students with lower MAP reading pretest scores. However, the interaction effects between the MORE condition and reading pretest did not show a statistically significant effect on listening comprehension, argumentative writing, the reading engagement measures, and the reading comprehension and basic literacy skills posttest measures (all *ps* > .05), such that the treatment effects did not differ significantly by pretest reading scores.

Sensitivity analyses. The sensitivity analyses based on three different procedures for handling missing data—(a) multiple imputation with multivariate normal distribution, (b) sample mean substitution (Appendices I and J), and (c) listwise deletion (Appendices K and L) yielded similar findings. The results based on models that use aforementioned procedures replicate the confirmatory analyses reported here in Table 5 and demonstrate positive main effects on science domain knowledge and reading comprehension outcomes.

Discussion

Using a within-school cluster randomized controlled design involving 38 first-grade classrooms, we compared the effectiveness of a 10-day lesson content literacy intervention to typical instruction during the 2-hour English language arts (ELA) block. First graders in MORE classrooms were exposed to more challenging texts focused on the science topic of Arctic animal survival. Findings indicate that first graders in MORE classrooms enjoyed gains in science domain knowledge and reading comprehension and performed as well as control students on measures of basic literacy and reading engagement. These findings support the theoretical proposition that domain knowledge plays a critical role in a student's reading comprehension ability (Alexander, 2018; Kintsch, 1998; Cromley & Azvedo, 2007).

Increasing young children's exposure to informational text through engaging content literacy instruction has been advocated for at least two decades and most recently by the CCSS (National Governors Association, 2010). Content literacy instruction has been documented to be effective for improving outcomes for upper elementary and middle-grade students (Connor et al, 2017; Guthrie et al., 2004; Vaughan et al., 2017; Wang & Herman, 2006). To our knowledge, however, few experimental studies have demonstrated the effectiveness of a whole class (i.e., Tier I) content literacy instruction for *first-graders*. In addition, the present study is the first to conduct a science content literacy intervention at scale with regular classroom teachers (rather than interventionists trained by researchers) and in schools that represent the inference population. Generalization of the results to authentic conditions matters because novel literacy interventions are often studied in high-quality implementations and are therefore difficult to replicate in real-world school and classroom contexts (Cronbach et al., 1980). Finally, this study underscores the value of targeting and measuring improvement in proximal domain knowledge outcomes such as vocabulary knowledge depth and argumentative writing. These outcomes are difficult to measure and frequently overlooked by researchers and practitioners but are sensitive changes in the quality of texts and tasks used during Tier I literacy instruction.

Evidence for the Effectiveness of MORE

In all likelihood, there are a number of related instructional factors in MORE that led to gains in comprehension. In many respects, the MORE intervention bears similarities to other content literacy programs that emphasize mastery goals for acquiring science domain knowledge (Guthrie et al., 2004; Romance & Vitale, 2001), complex and conceptually related informational texts (Cervetti et al., 2016; Duke et al., 2003), explicit instruction in domain specific science vocabulary with concept maps (Nesbit & Adesope, 2006), daily integrated reading and writing activities (Graham et al., 2017), and student choice of collaborative research topics (Guthrie et al., 2004; Souvignier & Mokhlesgerami, 2006; Vaughn et al., 2013). Nonetheless, a key malleable factor that differentiated instruction during MORE and typical English language arts lessons was the use of challenging texts with relatively higher readability levels, more complex text structure, and higher knowledge demands that cohere around a single topic. The experience of reading challenging texts afforded students opportunities to acquire more formal academic language, more abstract concepts, and more sophisticated sentences in daily reading, writing, and listening activities (Lupo, Tortorelli, Invernizzi, Ryoo, & Strong, 2019).

Interleaving the several instructional practices into a thematic unit may also foster a language-rich environment that facilitates children's domain knowledge acquisition and reading comprehension. For example, Perfetti (2007) has argued that individual differences in children's reading ability can be traced to "variation in literacy and language experiences" and "practice in reading and writing, and engagement with concepts and their language forms" may be critical to supporting reading comprehension development (p. 380). The absence of child-by-treatment interaction effects on reading comprehension and basic literacy skills suggests that MORE intervention did not lead to Matthew effects (Stanovich, 1986) in which stronger readers reap greater gains in reading comprehension than weaker readers. In sum, the positive main effect of MORE on students' reading comprehension replicates existing models of content literacy instruction in the early grades (Connor et al., 2017; Williams et al., 2016). At the same time, there was evidence of Matthew effects on proximal vocabulary knowledge depth measures. These mixed findings should encourage replication studies that explore whether content literacy interventions mitigate or exacerbate *long-term* gaps between low- and high-performing readers.

Our study also extends research by highlighting key instructional components in first-grade content literacy instruction. There were positive treatment effects on all three posttest measures of science domain knowledge. To begin, using conceptually related science texts and providing children with repeated exposures to related domain specific vocabulary may help to foster vocabulary knowledge depth (Fitzgerald et al., 2017). The concept of vocabulary knowledge depth implies that each word in our lexicon has connections to other words (McKeown et al., 2017). Put another way, students' knowledge of an individual word is "merely the exposed tip of the conceptual iceberg" (Anderson & Freebody, 1981, p. 82). Furthermore, concept mapping may facilitate connections among related words; that is, knowing that the words habitat, physical feature, behavior, and adaptation are related to the topic of animal survival may serve as a proxy for students' domain knowledge (Elleman, Lindo, Morphy, & Compton, 2009), a critical determinant of reading comprehension (Kintsch, 2009). We did not find evidence of incidental acquisition of general academic words that were not directly taught in the lessons or included in the classroom concept map. Although some research indicates that reading conceptually related science texts may support the acquisition of general academic words in fourth-grade (Cervetti et al., 2016), such transfer in the early grades may depend on longer and more intensive program implementations that help children develop a richer network of conceptually related words (Stahl & Nagy, 2006).

The listening comprehension and argumentative writing tasks captured students' domain knowledge acquisition in two additional ways. In other words, positive results on the listening comprehension and argumentative writing task provide further evidence that the MORE instructional strategies helped children learn science domain content. However, the effect on listening comprehension should be interpreted with caution given the small number of items (n = 4) on the scale and low reliability. Nevertheless, the listening comprehension results imply that acquisition of domain specific words may help children instantiate a general schema (i.e., ecosystems) that provides mental hooks for hanging and organizing newly learned concepts in academic subjects (Barnett & Ceci, 2002; Kintsch, 2009). In addition, strategic support for argumentative writing may have helped students draw upon textual evidence to support claims in science (Ferretti & Lewis, 2013) and build their knowledge of argument structure (Bell, 2004; Ferretti & Graham, 2019; Olinghouse, Graham, & Gillespie, 2015). In particular, the argumentative writing prompt used in this study (i.e., *Should people be allowed to cut down trees in the rainforest?*) assessed students' ability to draw upon their domain knowledge of a general schema that was not directly taught in the MORE lessons, to plan and organize a response, and to use textual evidence. In other words, first graders had to demonstrate science domain knowledge and argument structure knowledge (i.e., having a claim, evidence, and ending). Importantly, the limited attention to writing in first grade (Coker et al., 2016) and the positive effects on argumentative writing suggest the importance of providing first graders daily opportunities to acquire domain knowledge and to build their knowledge of argument structure (Bereiter & Scardamalia, 1987; Graham & Harris, 2005; Olinghouse, Graham, & Gillespie, 2015).

In contrast to the positive effects on domain knowledge measures, we found no significant effects on proximal measures of students' situational interest, reading motivation, and task orientations. In some ways, these null findings are consistent with developmental research indicating that first graders typically have strong interest in learning, strong self-competence beliefs and task values, and are behaviorally engaged in classroom instruction (McKenna, Kear, & Ellsworth, 1990; Nicholls, 1979; Wigfield, Eccles, Schiefele, Roeser, & Davis-Kean, 2006). Reviewing the developmental research on motivation, Wigfield and Tonks (2004) have suggested that first grade represents an "optimistic beginning" (p. 263). Our results are generally consistent with this finding since over 70% of first graders in this study reported the highest levels of reader self-competence and valuing of reading and there was no posttest difference on the reading motivation outcomes between children in the MORE and typical instruction classrooms.

At the same time, existing longitudinal research indicates that task orientations in the early grades become an increasingly more important predictor of reading comprehension as children progress through school (Lepola et al., 2016). In fact, our bivariate analyses in Table 4 reveal a moderate correlation between task orientations and posttest reading comprehension (Pearson's $r = .49$) comparable in strength to the correlation between vocabulary depth and reading comprehension (Pearson's $r = .48$). On the basis of our correlational findings and prior research, future research might specifically explore whether proximal gains in vocabulary knowledge depth and task orientations underlie gains in reading comprehension outcomes.

Limitations and Future Research

There are several study limitations that should motivate future research. First, we did not demonstrate sustained effects following a more intensive content literacy intervention. To date, virtually all content literacy interventions have examined only short-term effects on student outcomes measured immediately after the program period. There is limited evidence on whether content literacy programs that are designed to serve as Tier 1 instruction can foster long-term gains in students' domain knowledge and reading comprehension.

Second, while the primary goal of this experimental study was to provide new insights about the effects of first-grade content literacy instruction on the proximal and distal outcomes, further study is needed to determine the relations between proximal outcomes (i.e., science domain knowledge and reading engagement) and their associations with distal outcomes (i.e., reading comprehension and basic literacy skills). Exploration of the inter-relations between proximal and distal outcomes using different statistical analytic approaches (e.g., multilevel structural equation modeling) would uncover potential mechanisms through which the content literacy intervention improved students' reading comprehension outcomes. In addition, the development of more valid and reliable proximal measures (e.g., listening comprehension and situational interest) would allow for a more robust analysis of the link between proximal and distal outcomes. Because children's decoding and language skills are both critical determinants of reading comprehension, future replications should examine whether all subcomponents of the MAP (both code-related and language comprehension tasks) are sensitive to the MORE intervention.

Third, there are open questions about the implementation of content literacy instruction. For example, it is unclear whether shorter or longer implementations are more effective since programs tend to yield comparable effects on far transfer reading comprehension outcomes regardless of the duration of program implementation (Guthrie et al., 2006; Vaughn et al., 2013, 2017). It is also unclear whether content literacy instruction is more effective if it covers two domains–social studies and science–rather than a single domain (Connor et al., 2017; Guthrie et al., 2004; Williams et al., 2016). Future research might explore whether longer implementation of content literacy and inclusion of other domains (e.g., history) enhance program effectiveness without compromising fidelity of implementation.

Finally, the novel findings from this study merit replication. Given the limited research based on first-grade content literacy interventions and the unique theory of change guiding the MORE intervention, it is critical to understand whether the effects from this study can be reproduced in a follow-up experiment. Replication is critical to establishing the reliability of the findings in the current study and addressing key study limitations, including the need for more reliable instruments for capturing first graders' listening comprehension. Most critically, future replication attempts need to explore whether a multi-year implementation of MORE beyond first-grade can produce far transfer to reading comprehension gains in third-grade and beyond (National Governors Association, 2010).

Conclusion

In conclusion, an emerging body of experimental research has shown that content literacy instruction can build domain knowledge in science (and social studies) particularly in the upper elementary and middle school grades. Furthermore, research suggests that many critical components of upper-elementary and middle-grade content literacy interventions, such as the integration of domain knowledge building and engagement practices (Guthrie et al., 2004; Romance & Vitale, 2001; Vaughn et al., 2013), can also work effectively in the early grades. The present study extends prior findings, suggesting that critical components of content literacy can be replicated in first grade as children begin the long and difficult climb to becoming proficient and engaged readers.

At the same time, it is important to bear in mind that MORE was implemented as a supplemental content literacy program during a brief implementation period without adequate attention to code-related instruction. Thus, a key question is how code-focused instruction should be integrated into content literacy instruction for beginning readers. Existing models based on the simple view of reading indicate that a parallel code- and language-focused program may benefit first graders (Hirsch, 2016; White, Grissmer, Altenhofen, & Larson, 2013). Other literacy models that integrate and individualize code-focused instruction into Tier I instruction have demonstrated robust positive impact on reading comprehension (Connor et al., 2009). Typical instruction in the district context for this study incorporated balanced literacy practices which organizes general reading instruction into discrete blocks of time focused on word reading, reading and writing, and small group guided reading (Fountas & Pinnell, 2010). Clearly, a comprehensive first-grade literacy program would need to emphasize both code- and language-focused instruction and include a wider set of texts and tasks. Our study underscores that within the context of a diverse array of comprehensive literacy programs that more emphasis on domain knowledge building may nonetheless enhance first graders' ability to read for understanding.

References

Ahmed, Y., Wagner, R. K., & Lopez, D. (2014). Developmental relations between reading and writing at the word, sentence and text levels: A latent change score analysis. *Journal of Educational Psychology, 106*, 419-434.

Ainley, M., Hidi, S., & Berndorff, D. (2001). Interesting, learning, and the psychological processes that mediate their relationship. *Journal of Educational Psychology, 94*, 545-561.

Albin, M., Benton, S., & Khramtsova, I. (1996). Individual differences in interest and narrative writing. *Contemporary Educational Psychology, 21,* 305-324.

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

Alexander, P. A. (2018). Into the future: A prospective look at the Model of Domain Learning. In H. Fives, & D. L. Dinsmore (Eds.), *The Model of Domain Learning: Understanding the development of expertise* (pp. 195–214). New York, NY: Routledge

Alexander, P. A., Jetton, T. L., & Kulikowich, J. M. (1995). Interrelationship of knowledge, interest, and recall: Assessing a Model of Domain Learning. *Journal of Educational Psychology, 87*(4), 559–575. doi:10.1037/0022-0663.87.4.559

American Association for the Advancement of Science (Project 2016). (1993). *Benchmarks for science literacy*. New York: Oxford University Press.

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.

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

Banilower, E. R., Smith, P. S., Weiss, I. R., Malzahn, K. A., Campbell, K. M., & Weis, A. M. (2013). Report of the 2012 National Survey of Science and Mathematics Education. Chapel Hill, NC: Horizon Research, Inc.

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.

Bear, D. R., Templeton, S., Johnston, F., & Invernizzi, M. (2012). *Words their way: Word study for phonics, vocabulary, and spelling instruction*. Upper Saddle River, NJ: Pearson Education.

Bell, P. (2004). Promoting students' argument construction and collaborative debate in the science classroom. In M. C. Linn, E. A. Davis, and P. Bell (Eds.), *Internet environments for science education*. Mahwah, NJ: Erlbaum.

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale, NJ: Erlbaum.

Berninger, V.W., & Abbott, R.D. (2010). Listening comprehension, oral expression, reading comprehension, and written expression: Related yet unique language systems in grades 1, 3, 5, and 7. *Journal of Educational Psychology*, *102*(3), 635–651.

Biemiller, A. (2005). Size and sequence in vocabulary development: Implications for choosing words for primary grade vocabulary instruction. In A. Hiebert & M. Kamil, (Eds.), *Teaching and learning vocabulary: Bringing research to practice* (pp. 223–242). Mahwah, NJ: Erlbaum

Bos, C., & Anders, P. (1990). Effects of interactive vocabulary instruction on the vocabulary learning and reading comprehension of junior-high learning disabled students. *Learning Disability Quarterly*, *13*(1), 31–42.

Boscolo, P., & Gelati, C. (2007). Best practices in promoting motivation in writing. In S. Graham, C. A. MacArthur, & J. Fitzgerald (Eds.), *Best practices in writing instruction* (pp. 202–221). New York: Guilford P

Britton, J., Burgess, T., Martin, N., McLeod, A., & Rosen, H. (1975). *The development of writing abilities (11-18)*. London: Macmillan.

Brown, A. L., Bransford, J. D., Ferrara, R. A., & Campione, J. C. (1983). Learning, remembering, and understanding. In J. H. Flavell & E. M. Markman (Eds.), *Handbook of child psychology: Vol. 3: Cognitive development* (pp. 77-166). New York, NY: Wiley.

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

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

Cartwright, K.B., & Guajardo, N.R. (2015). The role of hot and cool executive functions in prereader comprehension. In A. DeBruin-Parecki, A. van Kleeck, & S. Gear (Eds.), *Developing early comprehension: Laying the foundation for reading success* (pp. 151–177). Baltimore, MD: Paul H. Brookes

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

Chiesi, H. L., Spilich, G. J., & Voss, J. F. (1979). Acquisition of domain-related information in relation to high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior, 18*, 257-273.

Coker, D. L., Farley-Ripple, E., Jackson, A. F., Wen, H., MacAruthur, & Jennings, A. S. (2016). Writing instruction in first grade: An observational study. *Reading and Writing, 29*, 793-832.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*, 407-428.

Collins, J. L., Lee, J., Fox, J. D., & Madigan, T. P. (2017). Bringing together reading and writing: An experimental study of writing intensive reading comprehension in low-performing urban elementary schools. *Reading Research Quarterly, 52*, 311-332.

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.

Connor, C. M., Piasta, S. B., Fishman, B., Glasney, S., Schatschneider, C., Crow, E., Underwood, P., Morrison, F. J. (2009). Individual student instruction precisely: Effects of child X instruction interactions on first graders' literacy development. *Child Development, 80*, 77-100.

Conradi, K., Jang, B. G., & McKenna, M. C. (2014). Motivation terminology in reading research: A conceptual review. *Educational Psychology Review, 26*, 127-164.

Council for Basic Education. (1998). *Standards for excellence in education*. Washington, DC: Author.

Coyne, M. D., McCoach, D. B., Loftus, S., Zipoli, R., Ruby, M. Crevecoeur, Y. C., & Kapp, S. (2010). Direct and extended vocabulary instruction in kindergarten: Investigating transfer effects. *Journal of Research on Educational Effectiveness, 3,* 93-120.

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., Ambron, S. R., Dornbusch, S. M., Hess, R. D., Hornik, R. C., Phillips, D. C., Walker, D. F., & Weiner, S. S. (1980). *Toward reform of program evaluation: Aims, methods, and institutional arrangements.* San Francisco: Jossey-Bass.

Dewey, J. (1913). *Interest and effort in education.* Cambridge, MA: Riverside Press.

Douglas, K. M., & Albro, E. R. (2014). The progress and promise of the Reading for Understanding Research Initiative. *Educational Psychology Review, 26*, 341-355.

Duke, N. K., & Block, M. K. (2012). Improving reading in the primary grades. *The Future of Children, 22*(2), 55-72.

Duke, N. K., Martineau, J. P., Frank, K. A., & Bennett-Armistead, V. S. (2003). *3.6 minutes per day: What happens when we include more informational text in first-grade classrooms?*

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

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

Faggella-Luby, M. N., Graner, P. S., Deshler, D. D., & Drew, S. V. (2012). Building a house on sand: Why disciplinary literacy is not sufficient to replace general strategies for adolescent learners who struggle. *Topics in Language Disorders, 32*, 69-84.

Ferretti, R. P., & Graham, S. (2019). Argumentative writing: Theory, assessment, and instruction. *Reading and Writing, 32*, 1345-1357.

Ferreti, R. P., & Lewis, W. E. (2013). Best practices in teaching argumentative writing. In S. Graham, C. A. MacArthur, & J. Fitzgerald (Eds.), *Best practices in writing instruction* (pp. 113-140). New York, NY: Guilford Publications.

Fitzgerald, W. J., Elmore, J., Kung, M., & Stenner, A. J. (2017). The conceptual complexity of vocabulary in elementary-grades core science program textbooks. *Reading Research Quarterly, 52*(4), 417-442.

Fountas, I. C., & Pinnell, G. S. (2010). The Fountas & Pinnell Leveled Book List, K-8+, 2010-12 edition.

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.

Good, R. H., & Kaminski, R. A. (2002). *Dynamic indicators of basic early literacy skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement. Available online at: http://dibels.uoregon.edu

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. Remedial and Special Education, 7(1), 6–10.

Graham, S., & Harris, K. J. (2005). *Writing better: Effective strategies for teaching students with learning difficulties*. New York: Brooks.

Graham, S., & Harris, K. R. (1989). Improving learning disabled students' skills at composing essays: Self-instructional strategy training. *Exceptional Children, 56*, 201–214.

Graham, S., Bollinger, A., Olson, C. B., D'Aoust, C., MacArthur, C., McCutchen, D., & Olinghouse, N. (2012). *Teaching elementary school students to be effective writers: A practice guide*. NCEE 2012-4058. What Works Clearinghouse.

Graham, S., Harris, K. R., & Hebert, M. (2011). It is more than just the message: Analysis of presentation effects in scoring writing. *Focus on Exceptional Children, 44*(4), 1–12.

Graham, S., Harris, K., R., & Santangelo, T. (2015). Research-based writing practices and the Common Core. *Elementary School Journal, 115*, 498-522.

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

Guo, Y., Connor, C. M., Tompkins, V., & Morrison, F. J. (2011). Classroom quality and student engagement: Contributions to third-grade reading skills. *Frontiers in Psychology, 2*, 1-10.

Guthrie, J. T., & Humenick, N. M. (2004). Motivating students to read: Evidence for classroom practices that increase reading motivation and achievement. In P. McCardle & V. Chhabra (Eds.), *The voice of evidence in reading research* (pp. 329-354). Baltimore, MD: Brookes.

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*, 237-250.

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 engagement* (pp. 601-634). Springer, Boston, MA.

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. *Journal of Educational Research, 99*, 232–245.

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). Mahwah, NJ: Erlbaum.

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

Halliday, M. A. K., & Martin, J. R. (1993). *Writing science: Literacy and discursive power.* Pittsburgh, PA: University of Pittsburgh Press.

Hedges, L. V. (2007). Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Studies, 32,* 341–370. http://dx.doi.org/10.3102/1076998606298043

Hedges, L. V. (2013). Recommendations for practice: Justifying claims of generalizability. *Educational Psychology Review* 25(3): 331–37.

Herman, J. L., Epstein, S., Leon, S., Dai, Y., La Torre Matrundola, D., Reber, S., & Choi, K. (2015). *The implementation and effects of the Literacy Design Collaborative (LDC): Early findings in sixth-grade Advanced Reading courses* (CRESST Report 846). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Hidi, S. (1990). Interest and its contribution as a mental resource for learning. *Review of Educational Research, 60*(4), 549–571. doi:10.3102/00346543060004549

Hidi, S., & Harackiewicz, J. (2000). Motivating the academically unmotivated: A critical issue for the 21st century. *Review of Educational Research, 70,* 151–179.

Hirsch, E. D. (2016). *Why Knowledge Matters: Rescuing Our Children from Failed Educational Theories*. Harvard Education Press. 8 Story Street First Floor, Cambridge, MA 02138.

Hoover, W. A., & Tunmer, W. E. (2018). The simple view of reading: Three assessments of its adequacy. *Remedial and Special Education, 39*, 304-312.

Hulleman, C. S., Durik, A.M., Schweigert, S., & Harackiewicz, J. M. (2008). Task values, achievement goals, and interest: An Integrative analysis. *Journal of Educational Psychology, 100,* 398-416.

Kane, T. J. (2014). "Do value-added estimates identify causal effects of teachers and schools? Washington, DC: Brookings Institution. Retrieved from: https://www.brookings.edu/research/do-value-added-estimates-identify-causal-effects-of-teachers-and-schools/

Kendall, J. S., & Marzano, R. J., (McRel). (2000). *Content knowledge: A compendium of standards and benchmarks for K-12 education* (3rd ed.). Alexandria, VA: Association for the Supervision and Curriculum Development.

Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis*. New York: Guilford Press.

Keys, C. W. (1999). Revitalizing instruction in scientific genres: Connecting knowledge production with writing to learn in science. *Science Education, 83,* 115-130.

Kim, J. (2018). A randomized controlled trial to replicate the effects of a Model of Reading Engagement (MORE) on first-graders' reading engagement, concept lnowledge, expository writing, and reading comprehension ." AEA RCT Registry. https://doi.org/10.1257/rct.2746-1.0.

Kintsch, W. (1988). The use of knowledge in discourse processing: A construction-integration model. *Psychological Review, 95*, 163–182.

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

Kintsch, W. (2005) An overview of top-down and bottom-up effects in comprehension: The CI Perspective. *Discourse Processes, 39*, 125-128

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.

Kuhn, D. (1993). Science as argument: Implications for teaching and learning scientific thinking. *Science Education, 77,* 319-337.

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: Understanding the development of expertise* (pp. 3-18). New York: Routledge.

Langer, J. (1984). The effects of available information on responses to school writing tasks. *Research in the Teaching of English, 18,* 27-44.

Lepola, J., Lynch, J., Kiuru, N., Laakkonen, E., & Niemi, P. (2016). Early oral language comprehension, task orientation, and foundational reading skills as predictors of grade 3 reading comprehension. *Reading Research Quarterly, 51*, 373-390.

Lepola, J., Poskiparta, E., Laakkonen, E., & Niemi, P. (2005). Development of and relationship between phonological and motivational processes and naming speed in predicting word recognition in grade 1. *Scientific Studies of Reading, 9*, 367-399.

Linnenbrink-Garcia, L., & Patall, E. A. (2016). Motivation. In L. Corno & E. Anderman (Eds.), *Handbook of educational psychology* (pp. 91–103). New York, NY: Routledge.

Linnenbrink, E. A., and Pintrich, P. R. (2003). The role of self-efficacy beliefs in student engagement and learning in the classroom. *Reading and Writing Quarterly, 19*, 119–137.

Lovett, M., Frijters, J. C., Wolf, M., Steinbach, K. A., Sevcik, R. A., & Morris, R. D. (2017). Early intervention for children at risk for reading disabilities: The impact of grade at intervention and individual differences on intervention outcomes. *Journal of Educational Psychology, 109*, 889-914.

Lupo, S. M., Tortorelli, L., Invernizzi, M., Ryoo, J. H., & Strong, J. Z. (2019). An exploration of text difficulty and knowledge support on adolescents' comprehension. *Reading Research Quarterly, 54*(4), 457-479. https://doi-org.prox.lib.ncsu.edu/10.1002/rrq.247

Mantzicopoulos, P., Patrick, H., & Samarapungavan, A. (2013). Science literacy in school and home contexts. *Cognition and Instruction, 31*, 62-119.

Marinak, B. A., Malloy, J. B., Gambrell, L. B., & Mazzoni, S. A. (2015). Me and my reading profile: A tool for assessing early reading motivation. *Reading Teacher, 69*, 51-62.

Marr, M. B., & Gormley, K. (1982). Children's recall of familiar and unfamiliar text. *Reading Research Quarterly, 18,* 89-104.

Marzano, R. J. (2004). *Building background knowledge for academic achievement: Research on what works in schools*. Alexandria, VA: Association for Supervision & Curriculum Development.

McCutchen, D. (2000). Knowledge, processing, and working memory: Implications for a theory of writing. *Educational Psychologist, 35*, 13-23.

McKenna, M. C., Kear, D. J., & Ellsworth, R. A. (1990). Children's attitudes toward reading: A national survey. *Reading Research Quarterly, 30*, 934-956.

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.

McMaster, K. L., Jung, P.G., Brandes, D., Pinto, V., Fuchs, D., Kearns, D., … Yen, L. (2014). Customizing a research-based reading practice. *The Reading Teacher*, *68*(3), 173–183.

McNeill, K. L. (2011). Elementary students' views of explanation, argumentation, and evidence, and their abilities to construct arguments over the school year. *Journal of Research in Science Teaching, 48*(7), 793-823.

Meece, J., Anderman, E., & Anderman, L. (2006). Classroom goal structure, student motivation, and academic achievement. *Annual Review of Psychology, 57*, 487–503.

Mol, S. E., Bus, A. G., & de Jong, M. T. (2009). Interactive book reading in early education: A tool to stimulate print knowledge as well as oral language. *Review of Educational Research, 79*, 979–1007.

Morgan, P. L., Farkas, G., Hillemeier, M. M., & Maczuga, S. (2016). Science achievement gaps begin very early, persist, and are largely explained by modifiable factors. *Educational Researcher*, *45*(1), 18-35.

Munter, C., Wilhelm, A.G., Cobb, P., & Cordray, D.S. (2014). Assessing fidelity of implementation of an unprescribed, diagnostic mathematics intervention. *Journal of Research on Educational Effectiveness*, *7*(1), 83–113. doi:10.1080/19345747.2013.809177

National Governors Association. (2010). *Common Core State Standards (English language arts/literacy standards)*. Washington, DC: National Governors Association Center for Best Practices, & Council of Chief State School Officers.

National Research Council. (1996). *National science education standards*. Washington, DC: National Academy Press.

National Research Council. (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Committee on a Conceptual Framework for New K-12 Science Education Standards. Board on Science Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

National Science Teachers Association. (1993). *Scope, sequence, and coordination of secondary school science. Vol. 1. The content core*. Washington, DC: Author.

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.

Nicholls, J. G. (1979). Development of perception of own attainment and causal attributions for success and failure in reading. *Journal of Educational Psychology,* 71,94-99.

Novak, J. D. (1990). Concept mapping: A useful tool for science education. *Journal of Research in Science Teaching, 27,* 937–949.

Novak, J. D., & Gowin, D. B. (1984). *Learning how to learn.* New York: Cambridge University Press.

Olinghouse, N. G., Graham, S., & Gillespie, A. (2015). The relationship of discourse and topic knowledge to fifth graders' writing performance. *Journal of Educational Psychology, 107,* 391-406.

Paivio, A. (1986). *Mental representations: A dual coding approach.* Oxford, UK: Oxford University Press.

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., & Billman, A.K. (2016). Reading to Learn Science; A Right that Extends to Every Reader—Expert or Novice. In Zehlia Babaci–Wilhite (Ed.), *Human Rights in Language and STEM Education* (pp. 17–35). The Netherlands: Sense Publishers.

Penno, J. F., Wilkinson, I. A. G., & Moore, D. W. (2002). Vocabulary acquisition from teacher explanation and repeated listening to stories: Do they overcome the Matthew effect? *Journal of Educational Psychology, 94*, 23-33.

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.

Pinnell, G. S., & Fountas, I. C. (2007). The continuum of literacy, grades K-8: Behaviors and understandings to notice, teach, and support. Portsmouth, NH: Heinemann.

Pressley, M. (1998). *Reading instruction that works: The case for balanced teaching*. New York, NY: Guilford Press

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). Mahwah, NJ: 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: John Benjamins.

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

Recht, D. R., & Leslie, L. (1988). Effect of prior knowledge on good and poor readers' memory of text. *Journal of Educational Psychology, 80,* 16-20.

Rivard, L. P. (1994). A review of writing to learn in science: Implications for practice and research. *Journal of Research in Science Teaching, 31,* 969-983.

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.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. London: Chapman & Hall.

Schiefele, U. (1999) Interest and learning from text. *Scientific Studies of Reading, 3*, 257-279

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

Schneider, W. (1985). Developmental trends in the metamemory behavior relationship: An integrative review. In D. L, Forrest-Pressley, G. E. MacKinnon, & T. G. Waller (Eds.), *Metacognition, cognition, and human performance* (Vol. 1, pp. 57-109). Orlando, FL: Academic Press

Schneider, W., Körkel, J., & Weinert, F. E. (1989). Domain-specific knowledge and memory performance: A comparison of high- and low-aptitude children. *Journal of Educational Psychology, 81*, 306-312.

Schraw, G., Flowerday, T., & Lehman, S. (2001). Increasing situational interest in the classroom. *Educational Psychology Review, 13*, 211-224.

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.

Snow, C. E. (2010). Academic language and the challenge of reading for learning about science. *Science, 328,* 450–452. http://dx.doi.org/10 .1126/science.1182597

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*, 325-244.

Souvignier, E., & Mokhlesgerami, J. (2006). Using self-regulation as a framework for implementing strategy instruction to foster reading comprehension. *Learning and Instruction, 16*, 57–71.

Spilich, G. J., Vesonder, G. T., Chiesi, H. L., & Voss, J. F. (1979). Text processing of domain-related information for individuals with high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior, 18*, 275-290.

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

Stahl, S.A., & Nagy, W.E. (2006). *Teaching word meanings*. Mahwah, NJ: Erlbaum.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21,* 360–401.

StataCorp. (2017). *Stata statistical software: Release 15*. College Station, TX: StataCorp LLC.

Swanson, H. L., & Hoskyn, M. (1998). Experimental intervention research on students with learning disabilities: A meta-analysis of treatment outcomes. *Review of Educational Research, 68,* 277–321.

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*, 351-366.

The New Teacher Project. (2018). *The opportunity myth*. New York, NY. Retrieved from https://tntp.org/publications/view/student-experiences/the-opportunity-myth

University of Oregon (2012). Understanding the research behind DIBELS 8th (Technical Report 1204). Eugene, OR: Center on Teaching and Learning.

Unrau, N. J., Rueda, R., Son, R., Polanin, J. R., Lundeen, R. J., & Muraszewski, A. K. (2018). Can reading self-efficacy be modified? A Meta-analysis of the impact of interventions on reading self-efficacy. *Review of Educational Research, 88, 167–204*.

van de Sande, E., Segers, E., & Verhoeven, L. (2013). How phonological awareness mediates the relation between self-control and word decoding. *Learning and Individual Differences, 26*, 112-118.

Vaughn, S., Martinez, L. R., Wanzek, J., Roberts, G., Swanson, E., & Fall, A. (2017). Improving content knowledge and comprehension for English language learners: Findings from a randomized control trial. *Journal of Educational Psychology, 109*, 22-34.

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.

Walker, C. H. (1987). Relative importance of domain knowledge and overall aptitude on acquisition of domain-related information. *Cognition and Instruction, 4,* 25-42.

Wang, J., & Herman, J. (2006). Evaluation of Seeds of Science/Roots of Reading Project: Shoreline Science and Terrarium Investigations (CSE Tech. Rep. No. 676). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing & Center for the Study of Evaluation (CSE).

Wang, E., Matsumura, L. C., & Correnti, R. (2017). Making a CASE: Improving use of text evidence in students' writing. *The Reading Teacher*, *70*(4), 479-484.

White, T. G., Grissmer, D. W., Altenhofen, S., & Larson, J. (2013, September). Lessons learned in conducting a lottery-based study of Core Knowledge Charter Schools. Poster presented at the Society for Research on Educational Effectiveness (SREE) conference, Washington, DC, USA

Wigfield, A., & Tonks, S. (2004). The development of motivation for reading and how it is influenced by CORI. In J. T. Guthrie, A. Wigfield, & K. C. Perencevich (Eds.), *Motivating reading comprehension: Concept-oriented reading instruction* (pp. 249-272). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Wigfield, A., Eccles, J. S., Fredricks, J. A., Simpkins, S., Roeser, R. W., & Schiefele, U. (2015). Development of achievement motivation and engagement. In M. E. Lamb & R. M. Lerner (Eds.), *Handbook of child psychology and developmental science* (Vol. 3, pp. 657-700). Hoboken, New Jersey: John Wiley & Sons, Inc

Wigfield, A., Eccles, J. S., Schiefele, U., Roeser, R. W., & Davis-Kean, P. (2006). Development of achievement motivation. In W. Damon, R. M. Lerner, & N. Eisenberg (Eds.), *Handbook of child psychology: Social, emotional and personality development* (Vol. 3, pp. 933-1002). Hoboken, New Jersey: John Wiley & Sons, Inc.

Wigfield, A., Gladstone, J. R., & Turci, L. (2016). Beyond cognition: Reading motivation and reading comprehension. *Child development perspectives*, *10*(3), 190-195.

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*, 1061-1077.

Literature Cited: Children's Books Used in the MORE Intervention for Interactive Read-Alouds

Berkes, M. (2008). *Over in the Arctic: Where the cold wind blows.* Nevada City, CA: Dawn Publications.

Osborne, M. P. (1998). *Polar Bears Past Bedtime.* New York, NY: Random House Books for Young Readers.

Osborne, M. P., & Byce, N. P. (2007). *Polar Bears and the Arctic*. New York, NY: Random House Books for Young Readers.

Pallotta, J. (2015). *Who would win? Polar bear vs. grizzly bear.* New York, NY: Scholastic Inc.

Thomson, S. L. (2009). *Where do polar bears live?* New York, NY: HarperCollins.

Footnote

[1]The 11 concept units from the source text were as follows: (1) Rainforest are valuable; (2) Rainforests are being destroyed quickly. People/families are cutting them down; (3) There are lots of people on Earth; (4) People need to cut down trees to clear land for roads to travel; (5) People need to cut down trees to clear land for houses for shelter from weather; (6) People need to cut down trees to clear land for growing food; (7) Unique/rare/that don't live anywhere else plants and/or animals live in the rainforest; (8) Cutting down trees leads to animals being destroyed, killed, endangered and/or extinct; (9) Woolly and/or spider monkeys live in the rainforest; (10) Monkeys are an example of an animal that is endangered and might be extinct one day; and, (11) This is bad because of the interdependence between animals and plants in the rainforest.

Table 1

Demographic Comparisons of Students in the Baseline Randomized Controlled Trial (RCT) Full Sample and by Treatment Condition and the Non-RCT Sample

	RCT Sample[a]					Non-RCT Sample	
	Total (N = 674)		MS (n = 213)	MS-H (n = 237)	TI (n = 224)	Total (N = 10,412)	
	N	%	n	n	n	n	%
Gender							
Female	339	51	99	122	118	5,090	49
Male	323	49	110	113	100	5,322	51
Race/ethnicity							
African American	217	33	69	78	70	3,599	35
Asian	60	9	19	25	16	855	8
Hispanic	221	33	65	81	75	2,674	26
White	141	21	47	41	53	2,917	28
Other	23	4	9	10	4	367	4
ELL status							
Non-ELLs	526	79	174	186	166	8,472	81
ELLs	136	21	35	49	52	1,940	19
Special education status							
No	617	93	191	221	205	9,523	91
Yes	45	7	18	14	13	889	9

Note. TI = Typical Instruction. MS = Model of Reading Engagement (MORE) at school. MS-H = MORE at school plus home. ELL = English-language learners.

[a]Of the RCT sample (N = 674), 662 students' demographic data were available.

Table 2

Informational and Narrative Books Used in Read-Alouds in MORE and TI Conditions

MORE	TI
Informational books	Informational books
Who Would Win?	*A Landforms Adventure*
Polar Bear vs. Grizzly Bear	*Sharks!*
Where Do Polar Bears Live?	*A Tree is a Plant*
Polar Bears and the Arctic	*Penguins!*
Narrative books	*Arctic Hares*
Over in the Arctic: Where the Cold Wind Blows	*Wild Baby Animals*
Polar Bears Past Bedtime	Narrative books
	The Little Three Pigs
	When Sophie Gets Angry--Really, Really Angry
	Sheila Rae the Brave
	The True Story of the Three Little Pigs by A.Wolf
	Knuffle Bunny: A Cautionary Tale
	Llama Llama Mad at Mama
	Officer Buckle and Gloria
	The Searcher and Old Tree
	The Little Red Hen: An Old Fable
	Mr. Putter & Tabby Drop the Ball
	Just Me And My Dad
	Hey, Little Ant
	Bonk and the Big Splash
	Arthur's Bad-News Day

Note. MORE = Model of Reading Engagement. TI = Typical Instruction

Table 3

Means, Standard Deviations, Independent t-test Results of the Openness of the Literacy Tasks in MORE and TI Lessons

	MORE		TI		
Openness criterion	*M*	*SD*	*M*	*SD*	*t*-test
Authenticity	2.25	.87	1.33	.48	4.82***
Collaboration	2.17	.94	1.10	.30	6.48***
Challenge level	2.42	.79	1.48	.59	4.48***
Student-directed work	2.83	.39	1.81	.67	5.03***
Sustained effort	2.83	.58	1.17	.49	9.99***

Note. Rating scale: 1 = close, 2 = moderately open, 3 = open.
MORE = Model of Reading Engagement. TI = Typical instruction.
†$p < .10$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Table 4

Pairwise Correlation Matrix[a] and Descriptive Statistics for MORE and TI Condition

	1	2	3	4	5	6	7	8	9	10	11	12	*n* (TI)	*M*	*SD*	Min.	Max.
Reading comprehension																	
1. MAP pretest	-	**.86**	**.74**	**.74**	**.53**	**.39**	**.48**	**.35**	**.48**	.03	**.20**	**.56**	165	173.76	15.97	129	221
2. MAP posttest	**.86**	-	**.74**	**.74**	**.53**	**.40**	**.47**	**.36**	**.46**	.10	**.22**	**.53**	204	175.88	17.13	125	219
3. DIBELS pretest	**.70**	**.68**	-	**.93**	**.48**	**.33**	**.44**	**.24**	**.42**	**.16**	**.28**	**.45**	212	191.19	112.27	0	450
4. DIBELS posttest	**.67**	**.68**	**.91**	-	**.45**	**.31**	**.42**	**.24**	**.40**	**.22**	**.35**	**.49**	214	186.26	93.54	0	377
Science domain knowledge																	
5. Vocabulary: Total	**.48**	**.47**	**.40**	**.37**	-	**.80**	**.82**	**.38**	**.47**	**.15**	**.27**	**.36**	205	30.07	4.51	18	42
6. Vocabulary: Taught	**.45**	**.46**	**.36**	**.34**	**.90**	-	**.32**	**.34**	**.39**	.11	**.24**	**.18**	205	17.71	2.72	11	24
7. Vocabulary: Untaught	**.37**	**.36**	**.33**	**.30**	**.85**	**.51**	-	**.28**	**.38**	.14	**.20**	**.39**	205	12.36	2.84	5	19
8. Listening comprehension	**.30**	**.33**	**.25**	**.24**	**.34**	**.27**	**.36**	-	**.34**	.01	.06	**.17**	203	1.56	1.06	0	4
9. Argumentative writing	**.36**	**.37**	**.34**	**.35**	**.37**	**.33**	**.32**	**.27**	-	.14	**.18**	**.27**	186	3.13	1.70	0	7
Reading engagement																	
10. Situational interest	-.06	-.01	.02	.00	.09	.11	.04	**.11**	**.14**	-	**.41**	**.26**	212	2.68	.36	1	3
11. Reading motivation	**.13**	.09	**.11**	.09	**.20**	**.20**	**.12**	.07	**.12**	**.37**	-	**.23**	208	49.33	8.75	6	60
12. Task orientations	**.49**	**.54**	**.49**	**.49**	**.40**	**.38**	**.31**	**.33**	**.38**	**.16**	.08	-	197	3.75	1.05	1	5
n (MORE)	319	403	425	431	415	415	413	416	401	366	417	417					
M	171.92	176.59	189.37	188.05	31.67	19.56	12.18	1.95	3.35	2.67	50.22	3.63					
SD	15.81	15.58	101.26	85.78	6.27	3.83	3.11	1.18	1.70	.37	7.94	1.11					
Min.	118	122	0	0	3	1	2	0	0	1	3	1					
Max.	206	216	478	362	43	26	20	4	7	3	60	5					

Note. MORE = Model of Reading Engagement. TI = Typical Instruction. MAP = Measure of Academic Progress. DIBELS = Dynamic Indicators of Basic Early Literacy Skills.

[a]Correlation values for MORE and TI condition appear below and above the diagonal, respectively. Significant correlations at the .05 level (two-tailed) are displayed in **bold**.

Table 5

Results of Hierarchical Linear Models Predicting the Main Effect of MORE on Science Domain Knowledge, Reading Engagement, and Reading Outcome

	Coefficient (*SE*)									
	Science domain knowledge					Reading engagement			Reading comprehension outcomes	
	Vocabulary knowledge depth									
Source	Total items	Taught words	Untaught words	Listening comprehension	Argumentative writing	Situational interest	Reading motivation	Task orientations	Reading comprehension (MAP)	Basic literacy skills (DIBELS)
Fixed effect										
Intercept, β_{00}[a], γ_{000}[b]	4.02 (3.01)	4.45 (1.89)*	.69 (1.64)	-1.58 (.64)*	-3.65 (.98)***	2.50 (.22)***	34.69 (4.30)***	-2.34 (.51)***	34.46 (4.96)***	45.16 (6.29)***
MORE, β_{01}, γ_{010}	1.61 (.67)*	1.85 (.39)***	-.18 (.37)	.45 (.18)*	.40 (.20)*	-.004 (.04)	1.00 (.68)	-.12 (.18)	1.75 (.83)*	3.70 (4.08)
Pretest, β_{10}, γ_{000}	.15 (.02)***	.08 (.01)***	.07 (.01)***	.02 (.003)***	.04 (.01)***	.001 (.001)	.09 (.02)***	.04 (.003)***	.83 (.03)***	.75 (.02)***
Male, β_{20}, γ_{200}	.68 (.39)†	.31 (.25)	.36 (.22)†	-.06 (.08)	.01 (.13)	-.11 (.03)***	-2.42 (.64)***	-.17 (.07)*	-1.43 (.71)*	-7.32 (2.72)**
Race/ethnicity[c]										
Asian, β_{30}, γ_{300}	2.00 (.79)*	1.40 (.51)**	.65 (.43)	.34 (.16)*	.44 (.25)†	.09 (.06)	1.89 (1.26)	.43 (.14)**	4.78 (1.37)***	12.10 (5.61)*
Hispanic, β_{40}, γ_{400}	1.06 (.59)†	.79 (.37)*	.40 (.33)	.14 (.12)	.23 (.19)	.03 (.05)	.62 (.95)	.20 (.11)†	1.53 (1.06)	5.16 (4.06)
White, β_{50}, γ_{500}	.48 (.64)	.45 (.41)	.13 (.35)	.26 (.14)†	.29 (.21)	-.04 (.05)	-2.34 (1.00)*	.15 (.11)	2.31 (1.14)*	4.13 (4.38)
Other, β_{60}, γ_{600}	.53 (1.11)	.38 (.71)	.13 (.61)	-.29 (.23)	.01 (.36)	.001 (.09)	.02 (1.74)	-.09 (.20)	-.12 (1.89)	-3.42 (7.87)
ELL, β_{70}, γ_{700}	-1.14 (.63)†	-1.01 (.40)*	-.20 (.34)	-.21 (.13)	-.08 (.20)	-.03 (.05)	.60 (1.01)	-.07 (.11)	-3.51 (1.16)**	-6.08 (4.27)
Special ed, β_{80}, γ_{800}	-.36 (.82)	-.45 (.52)	.25 (.46)	-.12 (.17)	-.71 (.30)*	-.01 (.06)	-.14 (1.33)	-.10 (.15)	.33 (1.45)	-21.59 (5.58)***
Title I, γ_{001}	-.66 (.94)	-.24 (.53)	-.55 (.50)	-.10 (.19)	-.14 (.24)	.02 (.06)		.01 (.18)	-1.69 (1.35)	-1.79 (4.55)
Random effect										
Level 1	4.69 (.14)	3.01 (.09)	2.58 (.08)	.98 (.03)	1.48 (.05)	.35 (.01)	7.84 (.22)	.81 (.03)	7.85 (.26)	33.54 (1.05)
Level 2	1.53 (.32)	.84 (.20)	.84 (.18)	.45 (.10)	.44 (.10)	.08 (.02)	1.20 (.50)	.45 (.08)	1.02 (.75)	8.39 (2.32)
Level 3	.83 (.44)	.44 (.28)	.41 (.26)	.00 (.00)	.11 (.22)	.04 (.03)		.00 (.00)	1.32 (.71)	.00 (.00)
N	620	620	618	619	587	578	625	614	607	645

Note. MORE = Model of Reading Engagement. MAP = Measure of Academic Progress. ELL = English-language learners. Special ed = special education.

[a]Symbols correspond to the estimates in the two-level model (i.e., reading motivation).

[b]Symbols correspond to the estimates in the three-level models.

[c]The reference category is African American.

†$p < .10$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Table 6

Results of Hierarchical Linear Models Predicting the Treatment-by-Student Interaction Effects on Science Domain Knowledge, Reading Engagement, and Reading Outcome

	Coefficient (*SE*)									
	Science domain knowledge					Reading engagement			Reading outcome	
	Vocabulary knowledge depth									
Source	Total items	Taught words	Untaught words	Listening comprehension	Argumentative writing	Situational interest	Reading motivation	Task orientations	Reading comprehension (MAP)	Basic literacy skills (DIBELS)
Fixed effect										
Intercept, β_{00}[a], γ_{000}[b]	11.70 (4.63)*	10.20 (2.91)**	1.54 (2.49)	-.98 (1.01)	-4.52 (1.51)**	2.21 (.32)***	28.88 (6.67)***	-2.21 (.84)**	24.67 (7.54)**	48.46 (7.26)***
MORE, β_{01}, γ_{010}	1.65 (.68)*	1.88 (.39)***	-.18 (.37)	.45 (.18)*	.40 (.20)*	-.01 (.04)	1.00 (.68)	-.12 (.18)	1.72 (.86)*	3.79 (4.14)
Pretest, β_{10}, γ_{000}	.10 (.03)***	.04 (.02)*	.06 (.01)***	.01 (.01)*	.04 (.01)***	.003 (.002)	.13 (.04)**	.03 (.005)***	.88 (.04)***	.74 (.02)***
MORE × Pretest, β_{11}, γ_{110}	.07 (.03)*	.05 (.02)*	.01 (.02)	.01 (.01)	-.01 (.01)	-.003 (.002)	-.05 (.04)	.001 (.01)	-.09 (.05)†	.03 (.03)
Male, β_{20}, γ_{200}	.70 (.39)†	.33 (.25)	.36 (.22)†	-.06 (.08)	.01 (.13)	-.11 (.03)***	-2.44 (.64)***	-.17 (.07)*	-1.46 (.71)*	-7.33 (2.72)**
Race/ethnicity[c]										
Asian, β_{30}, γ_{300}	1.95 (.78)*	1.36 (.51)**	.65 (.43)	.33 (.16)*	.44 (.25)†	.09 (.06)	1.93 (1.26)	.43 (.14)**	4.85 (1.36)***	11.79 (5.61)*
Hispanic, β_{40}, γ_{400}	1.06 (.59)†	.79 (.37)	.40 (.33)	.14 (.12)	.23 (.19)	.03 (.04)	.61 (.95)	.20 (.11)†	1.54 (1.05)	5.05 (4.06)
White, β_{50}, γ_{500}	.47 (.64)	.43 (.41)	.13 (.35)	.26 (.14)†	.30 (.21)	-.04 (.05)	-2.32 (.99)*	.15 (.11)	2.35 (1.13)*	4.05 (4.38)
Other, β_{60}, γ_{600}	.48 (1.10)	.34 (.70)	.12 (.61)	-.30 (.23)	.03 (.36)	.002 (.09)	.12 (1.75)	-.09 (.20)	.01 (1.89)	-3.63 (7.87)
ELL, β_{70}, γ_{700}	-1.25 (.82)*	-1.09 (.40)*	-.21 (.35)	-.22 (.13)†	-.07 (.20)	-.03 (.05)	.67 (1.01)	-.07 (.11)	-3.43 (1.15)**	-6.09 (4.27)
Special ed, β_{80}, γ_{800}	-.41 (.63)	-.49 (.52)	.24 (.46)	-.12 (.17)	-.71 (.30)*	-.01 (.06)	-.08 (1.33)	-.10 (.15)	.38 (1.44)	-21.91 (5.58)***
Title I, γ_{001}	-.63 (.97)	-.21 (.55)	-.54 (.50)	-.10 (.19)	-.14 (.24)	.02 (.05)		.01 (.18)	-1.67 (1.31)	-1.90 (4.60)
Random effect										
Level 1	4.66 (.14)	2.99 (.09)	2.58 (.08)	.98 (.03)	1.48 (.05)	.35 (.01)	7.84 (.22)	.81 (.03)	7.81 (.26)	33.48 (1.28)
Level 2	1.55 (.31)	.83 (.20)	.85 (.18)	.45 (.07)	.43 (.10)	.07 (.02)	1.16 (.50)	.45 (.08)	1.26 (.66)	8.63 (2.52)
Level 3	.88 (.45)	.48 (.27)	.42 (.26)	.00 (.00)	.12 (.21)	.04 (.03)		.00 (.00)	1.18 (.82)	.00 (.00)
N	620	620	618	619	587	578	625	614	607	645

Note. MORE = Model of Reading Engagement. MAP = Measure of Academic Progress. ELL = English-language learners. Special ed = special education.

[a]Symbols correspond to the estimates in the two-level model (i.e., reading motivation).

[b]Symbols correspond to the estimates in the three-level models.

[c]The reference category is African American.

†$p < .10$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

10 Day Lesson Sequence for 1 MORE Unit on Artic Animal Survival

Minutes	1	2	3	4	5	6	7	8	9	10
5	Establish Learning Goals and Interest	Concept Mapping	Establish Learning Goals and Interest	Concept Mapping	Interactive Read Aloud and Discussion	Interactive Read Aloud and Discussion	Collaborative Research	Interactive Read Aloud and Discussion	Interactive Read Aloud and Discussion	Concept Mapping
10										
15	Interactive Read Aloud and Discussion		Interactive Read Aloud and Discussion					Argumentative Writing (A TREE: Should people who live in the Artic be allowed to hunt and kill seals?)	Argumentative Writing (A TREE: Should people who live in the Artic be allowed to hunt and kill seals?)	
20					Concept Mapping	Concept Mapping	Concept Mapping			
25		Interactive Read Aloud and Discussion		Argumentative Writing (A TREE: Could You Survive in the Artic?) Could not survive in the Artic?	Argumentative Writing (A TREE: Could You Survive in the Artic?) Could survive in the Artic?	Collaborative Research				Large Group Discussion
30										
35										
40	Concept Mapping	Concept Mapping	Concept Mapping				Interactive Read Aloud and Discussion			
45										Review Learning Goals
50										
55							Small Group Discussion	Small Group Discussion	Small Group Discussion	
60										

Figure 1. The MORE intervention unit integrates practices to foster domain knowledge acquisition (conceptually related texts to establish learning goals and interest, concept mapping, argumentative writing) and engagement (interactive read aloud and discussion and collaborative research).

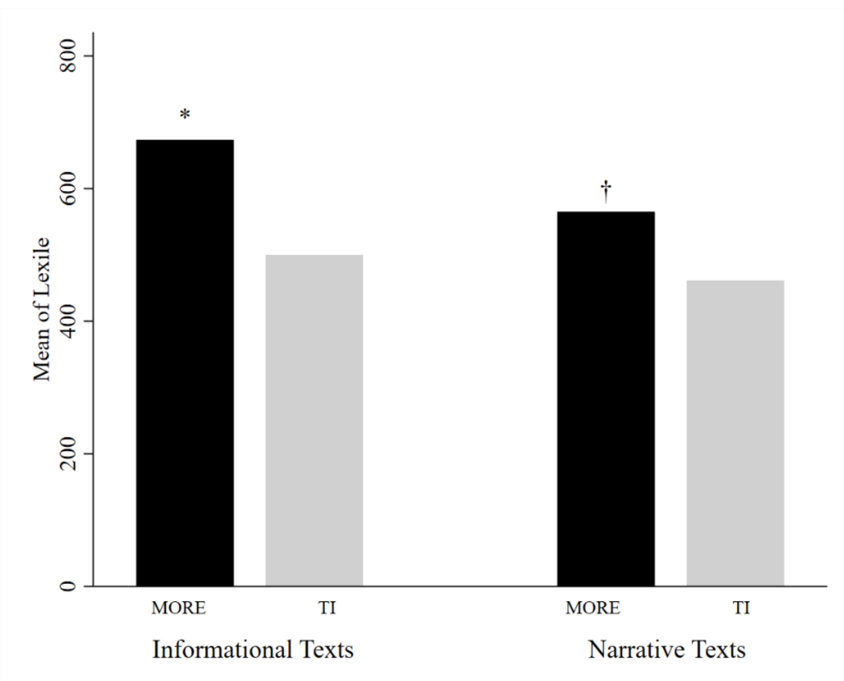


Figure 2. Means of Lexile levels of books (informational and narrative) used in MORE intervention and typical instruction lessons.
Note. †$p < .10$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

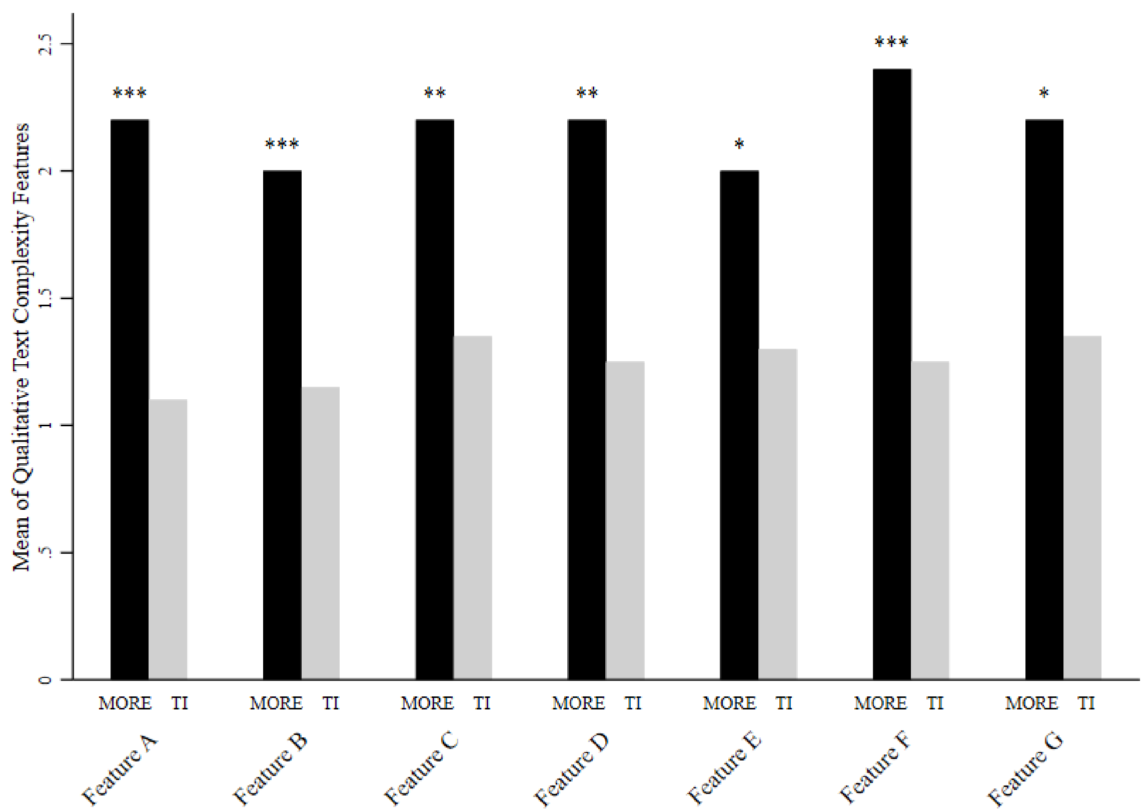


Figure 3. Bar graphs for means of qualitative features of text complexity for read-aloud books used in MORE and TI conditions. *Note*. MORE = Model of Reading Engagement. TI = Typical instruction. Feature A = Level of Meaning or Purpose. Feature B = Text Structure. Feature C = Text Features/Illustrations. Feature D = Language Conventionality and Clarity. Feature E = Knowledge Demands - content/discipline knowledge. Feature F = Knowledge Demands - cultural knowledge. Feature G = Knowledge Demands - vocabulary knowledge (see Appendix H for the description of rating scale).
†$p < .10$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Appendix A

Child-Friendly Definitions for MORE Domain-Specific Vocabulary

Adapt (verb) - To change over a long time so that it is easier to live someplace

Adaptation (noun) - A change in an animal (or plant) that makes it more likely to survive someplace

Advantage (noun) - Something that helps to make someone or something better more likely to win

Behavior (noun) - Something an animal does

Endangered (adjective) - When an animal species is in danger of no longer surviving

Extinct (adjective) - When an animal species no longer exists or survives

Habitat (noun) - The place where an animal makes its home

Physical feature (noun) - Something about the way an animal looks

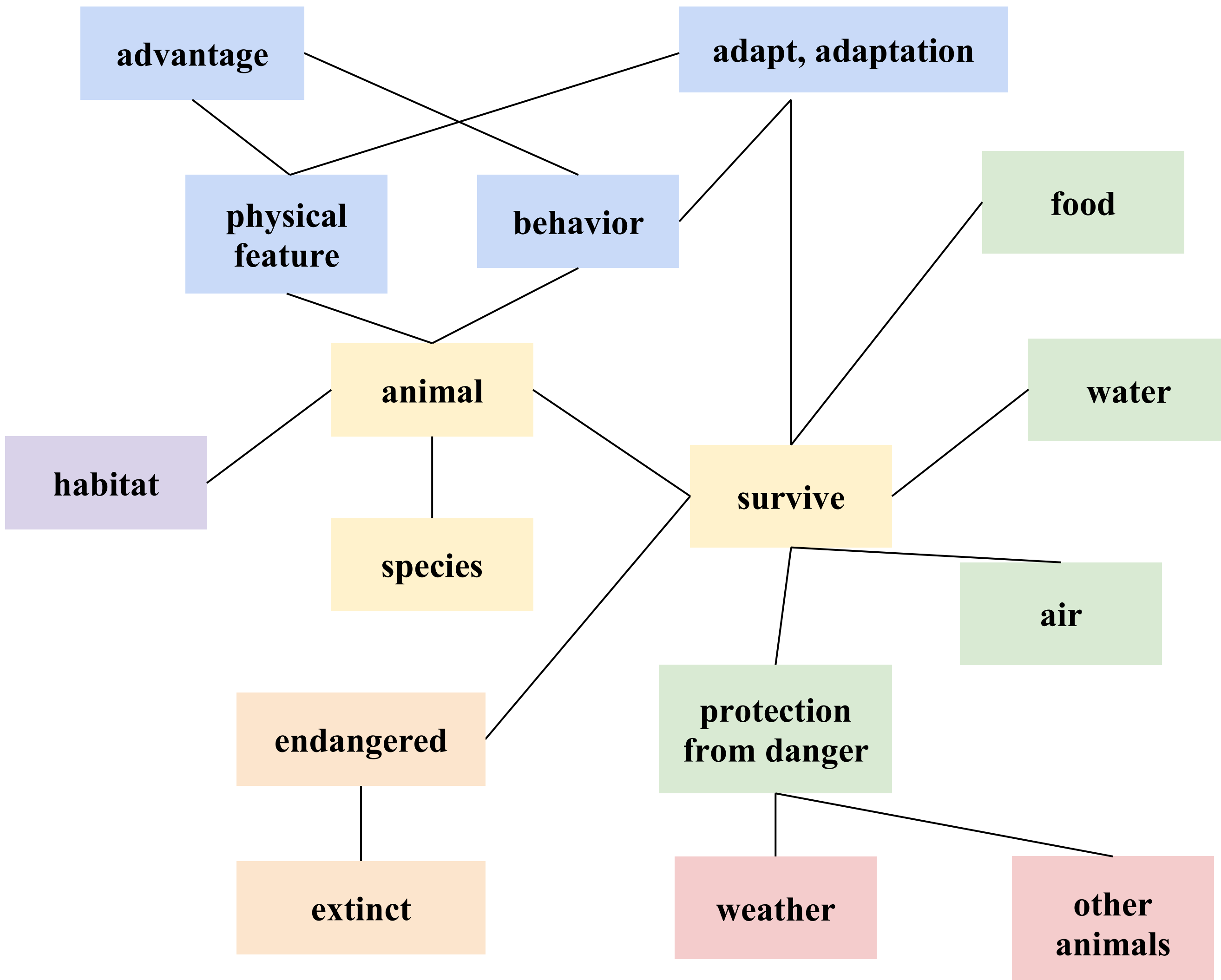
Shelter (noun) - Something that gives protection from things like bad weather or other dangers

Species (noun) - A group of similar animals

Survive (verb) - To stay alive

Appendix B

Concept Map Used in MORE Lessons



Appendix C

Literacy Tasks in MORE and TI Condition Lessons

MORE	TI
Students match the pictures of polar bears and seals to concept vocabulary words on the wall and explain their reasons.	Activate prior knowledge about chipmunk and discuss what students know about chipmunk based on the previous reading.
After teacher’s read-aloud, students engage in whole-class discussion of "Who would win between polar bear and grizzly bear?"	Activate students’ prior knowledge about the food chain and discuss with the examples.
Students read aloud, trace over, and write down a new vocabulary word and its definition.	Students identify and circle the vocabulary words from the book to match to the definition.
Students stand up if they believe what teacher says is a polar bear behavior.	The teacher and students describe the characters' positive or negative feelings.
Students complete the "exit ticket" in the notebook as wrap-up.	Students discuss different facts about animals based on the previous reading with a group.
Whole-class discussion about how people survive in the Arctic.	Repeated reading to enhance fluency and review the vocabulary word learned.
Students identify the main idea of the book and summarize the details.	Students learn about more words in the *kn* family (e.g., know) and use the words in sentence.
Students make an argument by applying “A TREE” strategy and explain reasons using evidence from the text.	The teacher and students discuss interesting facts about animals.
Students share their thoughts and opinions if they change their thoughts.	Students discuss their predictions, observations, and associations to the book.
Role Play: making verbal argument using "A TREE" strategy	Teacher reads questions and students circle answer of their choice on the worksheet.
After teacher’s read-aloud, students engage in whole-class discussion about main information in the story.	Students practice how to read exclamation points along with the teacher.
Students discuss “Should people who live in the Arctic be allowed to hunt and kill seals?” and share their opinions and supporting evidence.	Based on teacher' read-aloud, students discuss why rabbit was important to the story.
	The teacher and students discuss the feeling based on the characters' actions.
	Students describe the characters from the book, using pictures and vocabulary words.

	The teacher and students discuss lesson from the book, and how students feel about the book using thumbs up and down. Teacher asks students literal questions of the book. After choral reading, discuss why community helpers are important to community. Students apply RAP (restate, answer, and prove) strategy for reading comprehension using a book. Activate students' prior knowledge about goats by answer a question: "What is one thing in your memory that stands out about goats?" Students write several sight words (e.g., use and saw) on the board and create a sentence with the word "goat" in it using the sight words. Students analyze characters (e.g., What are characters like on the inside? What type of character traits could you use to describe Trixie?) Students turn and talk with the sentence starters including "I predict that..." or "The problem of the study is..." Students complete a graphic organizer about story setting and event. Students orally reconstruct a story that they have read. Group discussion on life lessons from the fable that students read. Teacher asks students literal questions of the book (e.g., "What has happened so far in the story?"). Teacher asks students inference questions of the book (e.g., "Why would she say no?" or "What lesson does she teach us?"). Students complete a graphic organizer about the lesson from fables. Students watch a video about recycling and then write down what they wonder about recycling.

	Teacher asks students inference questions of the book (e.g., "What are some clues in the last couple pages that tell us that the narrator is telling us the story looking in?") Students write three clues/key words that identify the voice of the story, either the narrator or characters. Students reflect how much they like the story book. Students look for words on a list and highlight with yellow crayon. Students write their realistic stories and teacher reads the students' stories to class. Students listen to a song called "reduce, reuse, recycle" and write down what they wonder about recycling. Students write sentences to answer why Stephanie Hardy wrote the article called "don't trash the earth, reduce, reuse, and recycle." Students individually research about three plants online. Students watch a video clip about seals and answer to teachers' questions about what they watched (e.g., tell something about the puffin.). After reading a book about the Inuit people, students write to describe the pictures associated with the Inuit people. Word study: students read *ai* words with the teacher. Write an acrostic poem about teacher and themselves using adjectives. Practice how to summarize main ideas and retell a story using anchor chart with the teacher and then independently. Students write sentences focusing on punctuations (e.g., spaces and capitals, etc.)

Note. MORE = Model of Reading Engagement. TI = Typical instruction.

Appendix D

Science Domain Knowledge Assessment: Vocabulary Knowledge Depth

1. Circle all of the words that go with the word **<u>potential</u>**.

future	bones	ability	report

2. Circle all of the words that go with the word **<u>unique</u>**.

characteristic	terrible	careful	quite different

3. Circle all of the words that go with the word **<u>survive</u>**.

food	dirt	alive	music

4. Circle all of the words that go with the word **<u>species</u>**.

endangered	alike	furry	not real

5. Circle all of the words that go with the word **<u>behavior</u>**.

fur	hunts	does	grass

6. Circle all of the words that go with the word **<u>resource</u>**.

leader	valuable	useful	falling apart

7. Circle all of the words that go with the word **advantage**.

teaches finds follows helps

8. Circle all of the words that go with the word **diversity**.

large pieces many variety riches

9. Circle all of the words that go with the word **adaptation**.

freeze camouflage scientist change

10. Circle all of the words that go with the word **habitat**.

wing forest place building

11. Circle all of the words that go with the words **physical feature**.

claws looks like sleeps acts like

12. Circle all of the words that go with the word **complex**.

not simple quick upset problem

Appendix E

Science Domain Knowledge Assessment: Listening Comprehension

Rainforests
By Will Osborne and Mary Pope Osborne

People in the Rainforest
Rainforests are one of the Earth's most valuable resources. But the rainforests are being destroyed very quickly. New babies are born every day. There are more and more people living on the Earth. Families are cutting down huge numbers of trees. They're clearing land to build roads so that they can travel from place to place. They're clearing land to build houses for shelter from the wet weather. They're also clearing land to grow crops and raise cattle for their food. Half of the world's rainforests are now gone.

Animals in the Rainforest
The rainforest is home to unique plants and animals that don't live anywhere else. When a rainforest is destroyed, these plants and animals are destroyed with it. Some rainforest animals are becoming very rare. For example, there were once thousands of woolly spider monkeys. Now there are only a few hundred. This is bad news for many plants, flowers, and fruits that need spider monkeys to carry their seeds from place to place. Because of the interdependence between animals and plants in the rainforest, what hurts one organism could hurt many organisms.

1. What does the word "interdependence" mean in the sentence, "Because of the interdependence between animals and plants in the rainforest, what hurts one organism could hurt many organisms."
   a. When two or more things need each other
   b. When two or more things eat the same food
   c. When two or more things look like each other
   d. I don't know
2. Why are people clearing land in the rainforest?
   a. They want to hunt monkeys.
   b. They want to stay safe and dry.

c. They want to destroy the rainforest.
d. They want to protect the rainforest.

3. What will happen to the woolly spider monkey if its habitat is destroyed?
   a. It will become a species.
   b. It will become extinct.
   c. It will become unique.
   d. It will become endangered.
4. Why did the authors include information about woolly spider monkeys?
   a. Because the passage is mostly about spider monkeys.
   b. Because it is evidence that cutting down the rainforest might not be a good idea.
   c. Because they wanted to summarize their thinking.
   d. Because they wanted to tell the reader about this special monkey.

Appendix F

Sample Situational Interest Measure from MORE Lesson 3

Sample question: What year is it?

1.	2.	3.
1998	2012	2018

A. When you read books about your research animal today, did you think it was:

1.	2.	3.
Great	OK	Boring

B. What kind of a reader did you feel like during today's MORE lesson?

1.	2.	3.
Not a good reader	An OK reader	A very good reader

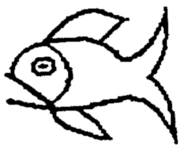
C. How would you feel if someone gave you your research group books for a present?

1.	2.	3.
Mad	OK	Happy

Appendix G

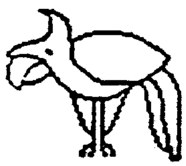
Reading Motivation Measure Adopted from Me and My Reading Profile

(Marinak, Malloy, Gambrell, & Mazzoni, 2015)



What grade are you in?

1.	2.	3.
Kindergarten	**First grade**	**Second grade**

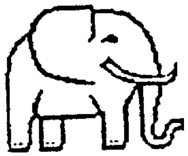


I am a ______________.

1.	2.
Boy	**Girl**

Do you like to read books all by yourself?

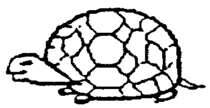
1.	2.	3.
Yes	**It's OK**	**No**



Learning to read is___________________.

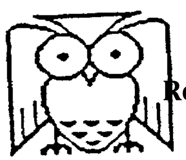
1. **2.** **3.**

Not very important **Sort of important** **Very important**



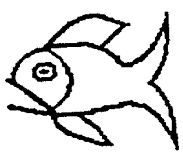
What kind of reader are you?

1.	2.	3.
I am not a good reader	**I am an OK reader**	**I am a very good reader**



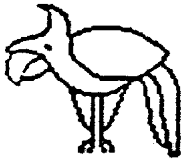
My friends think reading is ____________________.

1.	2.	3.
Really fun	**OK to do**	**No fun**



How do you feel when you read out loud to someone?

1.	2.	3.
Happy	**OK**	**Sad**



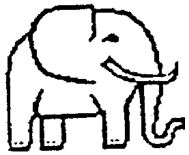
Do you tell your friends about books you read? 1. 2. 3.

Never Sometimes A lot

For me, learning to read is________________.

1. 2. 3.

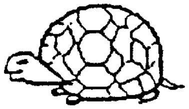
Easy Sort of hard Really hard



When someone reads books out loud to me, I think it is____________.

1. 2. 3.

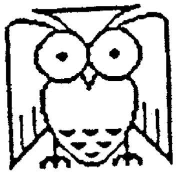
Great OK Boring



Do you like to read books out loud to someone else?

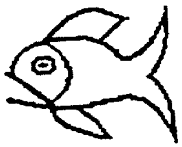
1. 2. 3.

No	It's OK	Yes



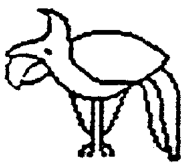
I think libraries are ____________________.

1. A great place to spend time
2. An OK place to spend time
3. A boring place to spend time



How do you feel about reading?

1. I don't like it
2. It's OK
3. I like it a lot



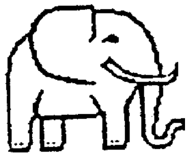
I spend ________________________.

1. None of my time reading books
2. Some of my time reading books
3. A lot of my time reading books

How do you feel when you are in a group talking about books?

1.
2.
3.

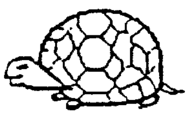
I do not like to talk about my ideas	**I sometimes like to talk about my ideas**	**I always like to talk about my ideas**



How would you feel if someone gave you a book for a present?

1. **2.** **3.**

Mad **OK** **Happy**



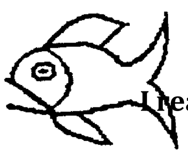
How do you feel about learning to read?

1. **2.** **3.**

I like it a lot **It's OK** **I don't like it**

Do you like to read when you have free time? 1. **2.** **3.**

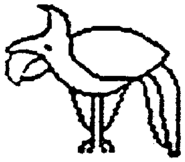
No **It's OK** **Yes**



How do you feel about reading with others? **1.** **2.** **3.**

I really like it **It's OK** **I don't like it**

at all



Do you have "favorite" books?

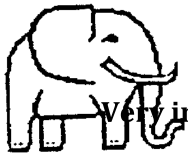
1. **2.** **3.**

Lots **Some** **None**

For me, reading is _____________________________.

1. **2.** **3.**

Really hard **Sort of hard** **Easy**



I think becoming a good reader is ______________________. 1. 2. 3.

Very important **Sort of important** **Not very important**

Appendix H

Qualitative Features of Text Complexity for Read-Aloud Books Used in MORE and TI Lessons and Rating Scale

Description of Qualitative Features and Rating Scale
A. Level of Meaning or Purpose
1. single meaning; purpose explicitly stated
2. more than one meaning; purpose implied but easy to infer
3. multiple layers of meaning; purpose subtle, implied, and hidden
B. Text Structure
1. simple and explicit organization of ideas
2. moderately complex
3. subtle and complex
C. Text Features/Illustrations
1. unnecessary or merely supplemental to understanding the text
2. supplemental to understanding the text
3. sophisticated, essential, and integrated
D. Language Conventionality and Clarity
1. literal, familiar, or conversational
2. less familiar
3. figurative or formal
E. Knowledge Demands - content/discipline knowledge
1. familiar concrete concepts or everyday experience
2. familiar discipline-specific concepts
3. abstract discipline-specific concepts
F. Knowledge Demands - cultural knowledge
1. low intertextuality (limited connections with other text, ideas, or theories)
2. moderate intertextuality
3. high intertextuality (contains multiple references)
G. Knowledge Demands - vocabulary knowledge
1. simple, high-frequency vocabulary
2. familiar academic/domain-specific vocabulary

3. sophisticated, complex academic/domain-specific vocabulary

Note. TI = Typical instruction.

Appendix I

Results of Hierarchical Linear Models Predicting the Main Effect of MORE on Science Domain Knowledge, Reading Engagement, and Reading Outcome Using Sample Mean Substitution to Account for Missing Values

	Coefficient (*SE*)									
	Science domain knowledge					Reading engagement			Reading outcome	
	Vocabulary knowledge depth									
Source	Total items	Taught words	Untaught words	Listening comprehension	Argumentative writing	Situational interest	Reading motivation	Task orientations	Reading comprehension (MAP)	Basic literacy skills (DIBELS)
Fixed effect										
Intercept, β_{00}[a], γ_{000}[b]	1.89 (3.11)	3.59 (1.97)†	-.30 (1.72)	-2.02 (.18)**	-4.29 (1.02)***	2.69 (.23)***	33.67 (4.71)***	-2.12 (.57)***	37.04 (5.76)***	45.32 (6.70)***
MORE, β_{01}, γ_{010}	1.61 (.69)*	1.85 (.40)***	-.18 (.38)	.45 (.003)*	.38 (.20)*	-.01 (.04)	1.00 (.68)	-.15 (.18)	1.51 (1.07)*	4.27 (4.40)
Pretest, β_{10}, γ_{000}	.16 (.02)***	.08 (.01)***	.07 (.01)***	.02 (.08)***	.04 (.01)***	.001 (.001)	.10 (.03)***	.03 (.003)***	.81 (.03)***	.75 (.02)***
Male, β_{20}, γ_{200}	.71 (.39)†	.31 (.25)	.40 (.22)†	-.05 (.16)	.02 (.13)	-.11 (.03)***	-2.39 (.64)***	-.17 (.07)*	-1.59 (.75)*	-7.36 (2.90)*
Race/ethnicity[c]										
Asian, β_{30}, γ_{300}	2.15 (.79)**	1.51 (.50)**	.70 (.43)	.35 (.12)*	.49 (.25)*	.10 (.06)	2.03 (1.26)	.48 (.14)**	5.65 (1.49)***	11.77 (5.93)*
Hispanic, β_{40}, γ_{400}	1.08 (.56)†	.81 (.37)*	.39 (.32)	.13 (.13)	.19 (.19)	.04 (.04)	.69 (.95)	.19 (.11)*	1.18 (1.13)	3.84 (4.29)
White, β_{50}, γ_{500}	.61 (.64)	.53 (.41)	.17 (.35)	.27 (.23)*	.31 (.20)	-.03 (.05)	-2.22 (1.00)*	.20 (.11)†	2.76 (1.23)*	5.31 (4.64)
Other, β_{60}, γ_{600}	.65 (1.10)	.45 (.70)**	.19 (.60)	-.28 (.23)	.02 (.35)	.01 (.09)	.14 (1.75)	-.04 (.21)	.10 (2.09)	-11.40 (8.14)
ELL, β_{70}, γ_{700}	-1.31 (.62)*	-1.12 (.40)**	-.26 (.34)	-.23 (.13)†	-.12 (.20)	-.04 (.05)	.44 (1.00)	-.12 (.11)	-4.36 (1.22)***	-4.67 (4.53)
Special ed, β_{80}, γ_{800}	.79 (.81)	-.68 (.52)	.05 (.45)	-.18 (.17)	-.77 (29)**	-.02 (.06)	-.54 (1.32)	-.23 (.14)	-2.23 (1.56)	-21.28 (5.97)***
Title I, γ_{001}	-.48 (.99)	-.16 (.55)†	-.42 (.53)	.01 (.21)	.06 (.25)	.03 (.05)		.09 (.20)	-.61 (1.73)	-2.45 (5.31)
Random effect										
Level 1	22.18 (1.30)	3.03 (.54)	6.69 (.39)	.96 (.06)	1.48 (.04)	.001 (.002)	61.76 (3.53)	.71 (.04)	82.22 (4.90)	1295.13 (79.59)
Level 2	2.59 (1.04)	.87 (.35)	.79 (.33)	.21 (.06)	.43 (.10)	.01 (.003)	1.22 (1.14)	.22 (.06)	3.98 (2.76)	84.55 (40.63)
Level 3	.57 (.71)	.41 (.22)	.15 (.22)	.00 (.00)	.23 (.14)	.12 (.01)		.00 (.00)	2.39 (2.51)	.00 (.00)
N	620	620	618	619	587	578	625	614	607	645

Note. MORE = Model of Reading Engagement. MAP = Measure of Academic Progress. ELL = English-language learners. Special ed = special education.

[a]Symbols correspond to the estimates in the two-level model (i.e., reading motivation).

[b]Symbols correspond to the estimates in the three-level models.

[c]The reference category is African American.

†$p < .10$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Appendix J

Results of Hierarchical Linear Models Predicting the Treatment-by-Student Interaction Effects on Science Domain Knowledge, Reading Engagement, and Reading Outcome Using Sample Mean Substitution to Account for Missing Values

	Coefficient (*SE*)									
	Science domain knowledge					Reading engagement			Reading outcome	
	Vocabulary knowledge depth			Listening comprehension	Argumentative writing	Situational interest	Reading motivation	Task orientations	Reading comprehension (MAP)	Basic literacy skills (DIBELS)
Source	Total items	Taught words	Untaught words							
Fixed effect										
Intercept, β_{00}[a], γ_{000}[b]	10.02 (4.74)*	10.02 (2.98)**	-.73 (2.63)	-1.05 (1.01)	-5.27 (1.62)**	2.58 (.33)***	28.25 (7.32)***	-1.79 (.84)*	24.31 (7.50)**	48.80 (7.17)***
MORE, β_{01}, γ_{010}	1.77 (.85)*	1.83 (.48)***	.03 (.39)	.49 (.21)*	.30 (.26)	-.03 (.05)	9.18 (8.55)	-.16 (.20)	1.94 (.96)*	3.66 (4.10)
Pretest, β_{10}, γ_{000}	.12 (.03)***	.05 (.02)**	.07 (.01)***	.01 (.01)**	.05 (.01)***	.001 (.002)	.13 (.04)**	.03 (.005)***	.88 (.04)***	.74 (.02)***
MORE × Pretest, β_{11}, γ_{110}	.08 (.03)*	.06 (.02)**	.002 (.02)	.01 (.01)	-.007 (.01)	-.001 (.002)	-.05 (.05)	.003 (.01)	-.09 (.05)†	.03 (.03)
Male, β_{20}, γ_{200}	.43 (.44)	.16 (.28)	.28 (.25)*	.002 (.09)	.07 (.15)	-.13 (.03)***	-2.42 (.64)**	-.23 (.07)**	-1.33 (.73)†	-6.97 (2.73)**
Race/ethnicity[c]										
Asian, β_{30}, γ_{300}	1.74 (.90)*	1.10 (.57)*	.70 (.50)	.37 (.19)*	.45 (.29)†	.07 (.07)	2.05 (1.26)	.39 (.15)**	5.49 (1.48)***	10.89 (5.60)*
Hispanic, β_{40}, γ_{400}	.24 (.68)	.08 (.43)†	.31 (.38)	.17 (.14)	.32 (.22)	-.01 (.05)	.68 (.95)	.11 (.12)	1.86 (1.12)	5.18 (4.05)
White, β_{50}, γ_{500}	.14 (.71)	-.03 (.45)	.29 (.40)	.23 (.15)	.26 (.23)	-.04 (.05)	-2.21 (.99)**	.26 (.12)*	3.17 (1.17)**	3.10 (4.38)
Other, β_{60}, γ_{600}	.79 (1.20)	.43 (.76)	.28 (.67)	-.33 (.25)	.11 (.40)	.01 (.09)	.22 (1.75)	-.20 (.20)	1.28 (1.96)	-5.71 (8.27)
ELL, β_{70}, γ_{700}	-.76 (.73)	-.91 (.46)*	.17 (.41)	-.22 (.15)	-.06 (.24)	-.01 (.05)	.48 (1.00)	-.12 (.12)	-4.17 (1.19)**	-6.21 (4.25)
Special ed, β_{80}, γ_{800}	-.77 (.93)	-.32 (.59)	-.26 (.52)	-.19 (.20)	-.72 (.34)*	.02 (.07)	-.51 (1.32)	-.14 (.15)	.89 (1.49)	-21.71 (5.56)***
Title I, γ_{001}	-.28 (1.01)	-.18 (.61)	-.15 (.57)	-.10 (.21)	.15 (.28)	.05 (.06)		.21 (.24)	-.63 (1.31)	-2.85 (4.57)
Random effect										
Level 1	21.18 (1.43)	8.52 (.58)	6.65 (.45)	.93 (.06)	2.17 (.15)	.11 (.01)	61.69 (3.52)	.57 (.04)	60.91 (4.06)	1114.82 (65.39)
Level 2	2.91 (1.30)	.84 (.41)	.49 (.28)	.19 (.07)	.25 (.12)	.01 (.01)	1.18 (1.12)	.17 (.07)	1.92 (1.85)	71.71 (35.28)
Level 3	.73 (.96)	.19 (.28)	.25 (.26)	.00 (.00)	.01 (.06)	.001 (.003)		.03 (.05)	.89 (1.83)	.00 (.00)
N	620	620	618	619	587	578	625	614	607	645

Note. MORE = Model of Reading Engagement. MAP = Measure of Academic Progress. ELL = English-language learners. Special ed = special education.

[a]Symbols correspond to the estimates in the two-level model (i.e., reading motivation).

[b]Symbols correspond to the estimates in the three-level models.

[c]The reference category is African American.

†$p < .10$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Appendix K

Results of Hierarchical Linear Models Predicting the Main Effect of MORE on Science Domain Knowledge, Reading Engagement, and Reading Outcome Using Listwise Deletion to Account for Missing Values

	Coefficient (*SE*)									
	Science domain knowledge					Reading engagement			Reading outcome	
	Vocabulary knowledge depth			Listening comprehension	Argumentative writing	Situational interest	Reading motivation	Task orientations	Reading comprehension (MAP)	Basic literacy skills (DIBELS)
Source	Total items	Taught words	Untaught words							
Fixed effect										
Intercept, β_{00}[a], γ_{000}[b]	1.46 (3.17)	3.57 (1.99)†	-.97 (1.76)	-2.02 (.66)**	-4.43 (1.04)***	2.69 (.22)***	33.88 (5.04)***	-2.10 (.54)***	34.15 (5.04)***	45.51 (6.22)***
MORE, β_{01}, γ_{010}	1.86 (.82)*	1.91 (.47)***	.03 (.39)	.51 (.20)*	.29 (.26)	-.03 (.05)	.58 (.81)	-.16 (.20)	1.87 (.90)*	3.61 (4.04)
Pretest, β_{10}, γ_{000}	.16 (.02)***	.08 (.01)***	.08 (.01)***	.02 (.004)***	.04 (.01)***	.001 (.001)	.10 (.03)***	.03 (.003)***	.82 (.03)***	.76 (.02)***
Male, β_{20}, γ_{200}	.38 (.45)†	.12 (.28)	.27 (.25)†	-.09 (.09)	.01 (.15)	-.13 (.03)***	-2.66 (.77)***	-.23 (.07)**	-1.28 (.73)†	-6.95 (2.73)**
Race/ethnicity[c]										
Asian, β_{30}, γ_{300}	1.76 (.91)†	1.11 (.57)**	.70 (.50)	.38 (.19)*	.63 (.29)*	.07 (.07)	.66 (1.49)	.39 (.15)**	6.43 (1.48)***	11.15 (5.60)*
Hispanic, β_{40}, γ_{400}	.22 (.69)	.07 (.43)*	.31 (.38)	.06 (.14)	.07 (.22)	-.01 (.05)	-.48 (1.18)	.11 (.12)	1.81 (1.12)	5.29 (4.06)
White, β_{50}, γ_{500}	.14 (.72)	-.02 (.45)	.29 (.40)	.22 (.15)*	.37 (.23)	-.04 (.05)	-2.48 (1.00)*	.26 (.12)*	3.11 (1.18)*	3.20 (4.38)
Other, β_{60}, γ_{600}	.82 (1.21)	.46 (.77)	.29 (.67)	-.32 (.25)	.09 (.39)	.01 (.09)	-.67 (2.03)	-.19 (.20)	1.12 (1.96)	-5.61 (8.28)
ELL, β_{70}, γ_{700}	-.61 (.73)†	-.80 (.46)*	.17 (.41)	-.16 (.15)	.04 (.24)	-.01 (.05)	1.12 (1.25)	-.11 (.12)	-4.24 (1.19)**	-6.23 (4.26)
Special ed, β_{80}, γ_{800}	-.78 (.93)	-.32 (.59)	-.26 (.52)	-.24 (.20)	-.67 (.34)†	.02 (.07)	-.38 (1.62)	-.14 (.15)	.86 (1.50)	-21.37 (5.56)***
Title I, γ_{001}	-.35 (1.07)	-.24 (.59)	-.15 (.57)	-.03 (.20)	.15 (.28)	.05 (.06)		.21 (.24)	-.60 (1.36)	-2.74 (4.51)
Random effect										
Level 1	21.53 (1.46)	8.70 (.59)	6.65 (.45)	.93 (.06)	2.17 (.15)	.11 (.01)	67.35 (4.46)	.57 (.04)	61.53 (4.09)	1118.51 (66.51)
Level 2	2.66 (1.22)	.79 (.40)	.48 (.28)	.18 (.07)	.26 (.12)	.01 (.005)	.53 (.99)	.17 (.07)	1.32 (1.57)	67.49 (35.33)
Level 3	.67 (.89)	.16 (.26)	.25 (.26)	.00 (.00)	.01 (.06)	.001 (.003)		.03 (.05)	1.27 (1.82)	.00 (.00)
N	462	462	459	461	438	453	465	450	478	623

Note. MORE = Model of Reading Engagement. MAP = Measure of Academic Progress. ELL = English-language learners. Special ed = special education.

[a]Symbols correspond to the estimates in the two-level model (i.e., reading motivation).

[b]Symbols correspond to the estimates in the three-level models.

[c]The reference category is African American.

†$p < .10$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Appendix L

Results of Hierarchical Linear Models Predicting the Treatment-by-Student Interaction Effects on Science Domain Knowledge, Reading Engagement, and Reading Outcome Using Listwise Deletion to Account for Missing Values

	Coefficient (*SE*)									
	Science domain knowledge					Reading engagement			Reading outcome	
	Vocabulary knowledge depth									
Source	Total items	Taught words	Untaught words	Listening comprehension	Argumentative writing	Situational interest	Reading motivation	Task orientations	Reading comprehension (MAP)	Basic literacy skills (DIBELS)
Fixed effect										
Intercept, β_{00}[a], γ_{000}[b]	10.02 (4.74)*	10.02 (2.98)**	-.73 (2.63)	-1.05 (1.01)	-5.27 (1.62)**	2.58 (.33)***	28.30 (7.77)***	-1.79 (.79)*	24.31 (7.50)**	48.80 (7.17)***
MORE, β_{01}, γ_{010}	1.77 (.85)*	1.83 (.48)***	.03 (.39)	.49 (.21)*	.30 (.26)	-.03 (.05)	.66 (.82)	-.16 (.20)	1.94 (.96)*	3.66 (4.10)
Pretest, β_{10}, γ_{000}	.12 (.03)***	.05 (.02)**	.07 (.02)***	.01 (.01)**	.05 (.01)***	.001 (.002)	.13 (.04)**	.03 (.004)***	.88 (.04)***	.74 (.02)***
MORE × Pretest, β_{11}, γ_{110}	.08 (.03)*	.06 (.02)**	.002 (.02)	.01 (.01)	-.01 (.01)	-.001 (.002)	-.05 (.05)	.003 (.01)	-.09 (.05)†	.03 (.03)
Male, β_{20}, γ_{200}	.43 (.44)	.16 (.28)	.28 (.25)	-.08 (.09)	.01 (.15)	-.13 (.03)***	-2.69 (.78)**	-.23 (.07)**	-1.32 (.73)†	-6.97 (2.73)*
Race/ethnicity[c]										
Asian, β_{30}, γ_{300}	1.74 (.90)†	1.10 (.57)†	.70 (.50)	.38 (.19)*	.62 (.29)*	.07 (.07)	.70 (1.49)	.39 (.15)**	6.49 (1.48)***	10.89 (5.60)†
Hispanic, β_{40}, γ_{400}	.24 (.68)†	.09 (.43)	.31 (.38)	.06 (.14)	.07 (.22)	-.01 (.05)	-.50 (1.18)	.11 (.12)	1.86 (1.12)†	5.18 (4.05)
White, β_{50}, γ_{500}	.14 (.71)	-.08 (.45)	.29 (.40)	.22 (.15)	.38 (.23)	-.04 (.05)	-2.45 (1.13)*	.26 (.12)*	3.17 (1.17)**	3.10 (4.38)
Other, β_{60}, γ_{600}	.79 (1.20)	.43 (.76)	.28 (.67)	-.33 (.25)	.11 (.40)	.01 (.09)	-.57 (2.03)	-.20 (.20)	1.28 (1.96)	-5.71 (8.27)
ELL, β_{70}, γ_{700}	-.76 (.73)*	-.91 (.46)*	.17 (.41)	-.17 (.15)	.05 (.24)	-.01 (.05)	1.18 (1.25)	-.12 (.12)	-4.17 (1.19)***	-6.21 (4.25)
Special ed, β_{80}, γ_{800}	-.77 (.93)	-.32 (.59)	-.26 (.52)	-.23 (.20)	-.66 (.34)†	.02 (.07)	-.35 (1.62)	-.14 (.15)	.89 (1.49)	-21.71 (5.56)***
Title I, γ_{001}	-.28 (1.01)	-.18 (.61)	-.15 (.57)	-.02 (.21)	.15 (.28)	.05 (.06)		.21 (.24)	-.63 (1.31)	-2.85 (4.57)
Random effect										
Level 1	21.18 (1.43)	8.52 (.58)	6.65 (.45)	.93 (.06)	1.50 (.05)	.11 (.01)	67.26 (.4.46)	.57 (.04)	60.91 (4.06)	1114.82 (65.39)
Level 2	2.91 (1.30)	.84 (.41)	.49 (.28)	.19 (.07)	.40 (.10)	.01 (.03)	.48 (.97)	.17 (.07)	1.92 (1.85)	71.71 (35.28)
Level 3	.73 (.96)	.19 (.28)	.25 (.27)	.00 (.00)	.00 (.00)	.001 (.003)		.03 (.05)	.89 (1.83)	.00 (.00)
N	462	462	459	461	438	453	465	450	478	623

Note. MORE = Model of Reading Engagement. MAP = Measure of Academic Progress. ELL = English-language learners. Special ed = special education.

[a]Symbols correspond to the estimates in the two-level model (i.e., reading motivation).

[b]Symbols correspond to the estimates in the three-level models.

[c]The reference category is African American.

†$p < .10$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.