AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Educational Psychology

ISSN: 0022-0663

2024, Vol. 116, No. 4, 550–568
https://doi.org/10.1037/edu0000868

Improving Second-Grade Reading Comprehension Through a Sustained Content Literacy Intervention: A Mixed-Methods Study Examining the Mediating Role of Domain-Specific Vocabulary

Douglas M. Mosher, Mary A. Burkhauser, and James S. Kim
Graduate School of Education, Harvard University

This mixed-methods study explores the relationship between early elementary students' domain-specific vocabulary knowledge and their ability to comprehend grade-level reading passages on unfamiliar science topics. Specifically, this study used (a) structural equation modeling (SEM) to examine the extent to which students' networks of domain-specific vocabulary knowledge in Grades 1 and 2 mediated the effects of a Tier 1 content-based literacy intervention on domain-specific reading comprehension scores in Grade 2 ($N = 2{,}156$); and (b) quantitative survey and qualitative interview data from teachers ($N = 48$) to surface new themes about teacher vocabulary instruction that might suggest potential explanations for the SEM findings. SEM analysis revealed that students' domain-specific vocabulary knowledge in first and second grade explained 69% of the treatment effect on a domain-specific reading comprehension outcome. Results from the quantitative survey also indicated that treatment group teachers reported providing more incidental exposures to vocabulary than control teachers (effect size = .54), and qualitative analyses revealed that teachers with high incidental exposures tended to provide expanded opportunities for their students to engage with words and to connect words to topics. Findings from this mixed-method study paint a more complete picture of (a) the important role domain-specific vocabulary knowledge plays in facilitating reading comprehension transfer in the domain of science, and (b) what teachers do during vocabulary instruction to promote transfer in domain-specific reading comprehension.

Educational Impact and Implications Statement
This study highlights the importance of teaching networks of domain-specific vocabulary in early elementary school grades. Findings show that domain-specific vocabulary knowledge taught in semantic networks explained over two-thirds of the treatment effect for a recent content literacy intervention in second grade. Interviews with participating teachers highlight the numerous ways teachers make vocabulary accessible for students: ample exposures to vocabulary and teaching words by connecting meanings to related words.

Keywords: vocabulary, reading comprehension, literacy, intervention, transfer

According to the most recent administration of the National Assessment of Educational Progress, the COVID-19 pandemic wiped out nearly a decade worth of growth in academic achievement among U.S. 9-year-olds and exacerbated inequality between low- and high-performing students (U.S. Department of Education, 2022). The drop in elementary-grade reading performance is likely to cascade throughout a student's education, resulting in lower levels of high school graduation, college enrollment, and adult earnings, particularly for low-performing students (Doty et al., 2022; Masten & Cicchetti, 2010).

To accelerate and equalize elementary students' reading comprehension outcomes, there is a clear need for more effective Tier I (whole classroom) curricula that improve young children's ability to read complex nonfiction texts with understanding. Although the recently completed $125M U.S. Department of Education's Reading for Understanding (RfU) initiative was aimed to advance this goal from kindergarten to Grade 12, the early elementary K-3 curricula, which were largely 1-year program implementations, did not improve standardized reading comprehension outcomes. In their synthesis of RfU projects, Pearson et al. (2020) underscored

Douglas M. Mosher https://orcid.org/0000-0001-5722-9218
Mary A. Burkhauser https://orcid.org/0000-0001-9142-7269
James S. Kim https://orcid.org/0000-0002-6415-5496
This research was funded by the Chan Zuckerberg Initiative. The opinions expressed are those of the authors and do not represent the views of the funders.

The authors thank Johanna Tvedt, Barbara Gilbert, and Kimberly Burleson for their help in conducting interviews. The authors also thank Margaret G. McKeown, Joshua B. Gilbert, and Jackie E. Relyea for their advice and helpful comments on this article.

Douglas M. Mosher served as lead for conceptualization, data curation, formal analysis, methodology, supervision, validation, writing–original draft, and writing–review and editing. Mary A. Burkhauser served in a supporting role for conceptualization and writing–review and editing. James S. Kim served as lead for funding acquisition and served in a supporting role for writing–review and editing.

Correspondence concerning this article should be addressed to Douglas M. Mosher, Graduate School of Education, Harvard University, Room Q-409, READS Lab, 50 Church Street, Cambridge, MA 02138, United States. Email: douglasmosher@gmail.com

that "transfer—being able to apply new skills and ideas in settings and on measures that differ from the instructional context—is … and likely always will be—a challenge to achieve" (p. 263). The major aim of this study was to build on the RfU projects by studying the longitudinal implementation of a Tier I content literacy program from Grade 1 to Grade 2 and examining the mediating role of domain-specific vocabulary knowledge on transfer measures of domain-specific reading comprehension.

There is emerging evidence that content literacy interventions that sustain and align science and social studies content and literacy practices across grades are a promising vehicle for promoting transfer of acquired knowledge to related topics (K. L. Alexander et al., 2007; Bailey et al., 2017; Kim et al., 2022; Nagy, 2005; Newmann et al., 2001). In all likelihood, it takes time for students to develop the background and vocabulary knowledge needed to understand new but related topics in science and social studies texts. Substantial correlational research indicates that vocabulary predicts reading comprehension in primary and secondary grades (Anderson & Freebody, 1981; Elleman et al., 2009; Stahl & Fairbanks, 1986), and more recent experimental studies of content literacy interventions have shown that positive impacts on comprehension are largely mediated through improvements in vocabulary (Connor et al., 2017; Language and Reading Research Consortium et al., 2019, 2022). In fact, the language-focused classroom intervention Let's Know, an RfU curriculum, found that vocabulary was the most important mediating pathway for improved reading comprehension; no other indirect pathways were significant (Language and Reading Research Consortium et al., 2019). However, content literacy interventions that go beyond a single year are rare and mixed-methods studies that illuminate what students learn and what teachers do are even rarer. There is a need to better understand how students' acquisition of domain-specific vocabulary knowledge in science and social studies across the early grades facilitates reading comprehension and the kinds of teacher practices that may aid in facilitating reading comprehension transfer.

The major aim of this study was to build on findings from the RfU projects by studying the longitudinal implementation of a Tier I content literacy program from first to second grade and examining the mediating role of domain-specific vocabulary knowledge on transfer measures of domain-specific reading comprehension. Thus, using data from a cohort of students that participated in the Model of Reading Engagement (MORE) intervention as first and second graders, the present mixed-methods study was guided by both confirmatory aims (i.e., using structural equation modeling (SEM) to test the theory of change for the intervention of how depth of vocabulary over time mediates the treatment effect on domain-specific reading comprehension shown in Figure 1) as well as exploratory aims (i.e., examining vocabulary practices of teachers and expanding our conceptual model to surface ideas about potential instructional mechanisms that help students deepen their depth of vocabulary knowledge). To date, few experimental studies of Tier I content literacy instruction that emphasize interleaving domain-specific vocabulary words into interactive read alouds, discussion, reading, and writing activities have examined whether and to what extent effects on reading comprehension are mediated by improvement in domain-specific vocabulary. To shed light on this area of study, we employ a novel mixed-methods study to paint a more complete picture of (a) the important role domain-specific vocabulary knowledge plays in facilitating reading comprehension transfer in the domain of science, and (b) what teachers do during vocabulary instruction within the context of a sustained content literacy intervention that was designed to promote transfer in reading comprehension.

Theoretical Foundation of Content Literacy Intervention

Knowledge Is Organized in Schemas

Schemas are an essential component for the instantiation of knowledge. A schema can be described as "an abstract knowledge structure" (Anderson & Pearson, 1984, p. 259) that "provides much of the basis for comprehending, learning, and remembering the ideas in stories and texts" (Anderson, 2013, p. 476). Put differently, schemas are the invisible mental structures on which learners hang their knowledge. Indeed, schemas help students organize and store key information in the surrounding world as schemas are malleable and constantly evolving.

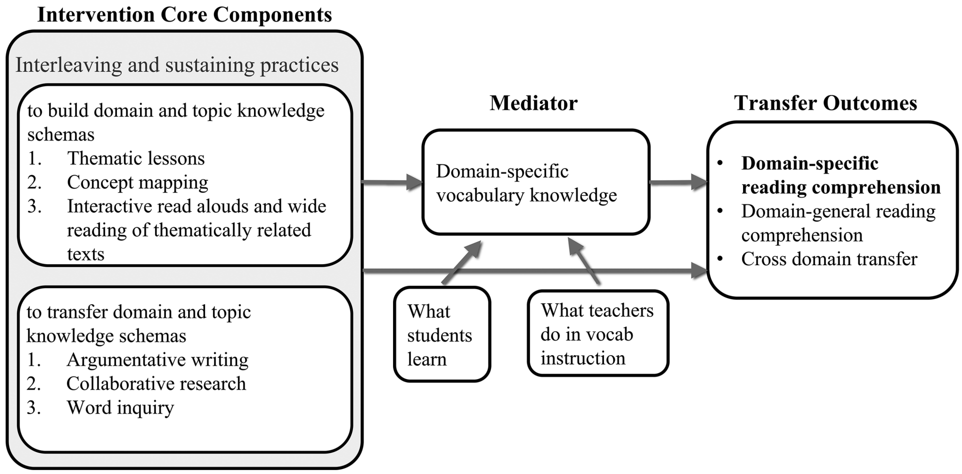
General schemas are broad in nature and encompass general concepts. Within a general schema lies a variety of topic schemas, each of which contain topic-specific information. As topic schemas are acquired and instantiated, the general schema within a given domain is strengthened (Kimball & Holyoak, 2000). For example, as students learn about and develop topic schemas around animal survival and dinosaur mass extinctions, they simultaneously develop and expand on a more general schema focusing on the concept of survival.

Networks of Vocabulary Knowledge Are the Visible Components of Schemas

In their knowledge hypothesis, Anderson and Freebody (1981) posit that greater vocabulary knowledge is an indication that individuals have greater knowledge to support their understanding of a text. Thus, words function as a proxy for knowledge, which can be organized via schemas where "individual word meanings are merely the exposed tip of the conceptual iceberg" (p. 82). While observing an individual's schema presents challenges given its opaqueness, we can observe the kernels of knowledge that form a schema—networks of vocabulary that are semantically related.

Emerging evidence underscores the benefits of teaching vocabulary using semantic networks (Fitzgerald et al., 2020; Hadley et al., 2019; Neuman et al., 2011). In essence, vocabulary instruction using semantic associations aims to help students identify links between groups of words (Read, 2004), thereby enabling children to move beyond rigid, easily memorizable definitions. According to the lexical quality hypothesis (Perfetti, 2007), children need flexible representations of word meanings and those with high-quality representations are better able to access semantic meanings while reading. For example, rather than introducing paleontologist, fossil, dinosaur, extinct, and theory in isolation, teaching students the ways these domain-specific words are related provides students with various representations of word meanings while also helping them understand how words and concepts are interconnected. Thus, as teachers provide high-quality representations of domain-specific word meanings in networks, students are better able to form a topic schema on how paleontologists study dinosaur fossils to explore theories about dinosaur mass extinctions. By teaching words in conceptually related categories and identifying the common semantic features that define each category, students are able

Figure 1
Intervention Theory of Change



to build and access semantic networks more efficiently and hang those words and concepts onto a schema. Thus, networks of vocabulary are the visible elements of a schema.

As students acquire domain-specific vocabulary around a given topic, they are building their domain knowledge—their overall knowledge within a given domain or field (P. A. Alexander, 2003). Where domain knowledge is general and emphasizes breadth of knowledge across a range of topics, topic knowledge underscores depth within a given topic. Students participating in the Grade 1 MORE science unit learned about animal survival in the arctic, which aimed to expand their topic knowledge of polar animals while broadly contributing to their overall domain knowledge in science (P. A. Alexander, 2003; Kim et al., 2020). Domain-specific words like habitat convey conceptual information about the domain of science as well as the topic of animal survival in the arctic and are essential to building both domain and topic knowledge. Indeed, knowledge building and vocabulary acquisition via semantic networks are tightly coupled (Collins & Quillian, 1969; Goldman et al., 1996; Kendeou & O'Brien, 2016) and both have an effect on comprehension (Cabell & Hwang, 2020). Building knowledge through the acquisition of domain-specific vocabulary is essential to formulating robust schemas and comprehending texts.

Schema Instantiation Facilitates Comprehension

Schema instantiation is essential to integrating new knowledge with prior knowledge. In the construction-integration model (Kintsch, 1988, 1998, 2013), readers must construct a situation model that combines literal understanding of the text in conjunction with their prior knowledge. Constructing an adequate situation model is particularly difficult when reading instructional texts because it requires readers to move past the passive processing of a textbase "and actively make meaning of the text utilizing their existing knowledge" (Kintsch & Kintsch, 2005, p. 76). As students build both general and topic schemas through the acquisition of conceptually related domain-specific vocabulary, they are subsequently able to store this knowledge in their long-term memory (Kintsch, 2009), thereby freeing up cognitive space to better comprehend texts. Thus, as learners become "experts" in certain domains and topics, they can access pertinent information fluently, allowing greater cognitive bandwidth to identify new relevant information (Bransford et al., 2000; Kintsch, 2009). More specifically, "words have great diagnostic value for the reader" in that reading a word can instantly evoke a relevant schema, thereby aiding comprehension (Anderson & Pearson, 1984, p. 261). Indeed, word meanings are essential and are the centerpiece in the reading systems framework. As children read, they encounter words that they know, and each of those words evokes certain semantic memories and connections that are automatic, thereby enabling information to be accessed and updated efficiently and effortlessly (Perfetti & Stafura, 2014). Furthermore, even if a specific word is not mentioned, its semantic connection to related words that are mentioned will resonate with the reader and evoke relevant knowledge (Myers & O'Brien, 1998; O'Reilly et al., 2019).

Robust Schemas Foster Comprehension Transfer

As students develop robust schemas, they are able to leverage networks of concepts when encountering related topics. Here, schemas serve "as mediators of analogical transfer" where new concepts are linked with a more established analog to facilitate a more analogous understanding (Gick & Holyoak, 1983, p. 9). Organizing information in networks (i.e., schemas via vocabulary networks) better enables students to transfer this knowledge to related topics (Bransford et al., 2000; Kendeou & O'Brien, 2016). Put differently, as learners develop their expertise in specific domains, they are more likely to be able to draw on and apply existing domain knowledge to novel tasks "that vary in surface characteristics from previously encountered situations" (Kimball & Holyoak, 2000, p. 118). In the most recent second-grade version of the MORE intervention, which forms the basis of the present study, results indicated significant transfer effects for students who received the intervention (Kim et al., 2022). Yet the extent to which the hypothesized mediator from the study—domain-specific vocabulary knowledge—mediated the treatment effect has yet to be examined.

In recent content literacy and vocabulary interventions, there are often significant effects on proximal measures, but null effects on distal measures (Apthorp et al., 2012; Connor et al., 2018; Cromley & Azevedo, 2007; Kim et al., 2021). Unsurprisingly, near-transfer measures are often closely aligned with intervention content, and far-transfer measures are often standardized assessments that have little in common with the intervention. A recent meta-analysis on the impact of vocabulary instruction on passage comprehension found that vocabulary interventions had large effect sizes (ES) on custom measures, most often proximal, and small ES on standardized distal measures (Wright & Cervetti, 2017). Furthermore, recent findings from the Let's Know intervention—a key program from the RfU initiative—showed that there were no significant effects on language comprehension, nor did vocabulary knowledge have a significant indirect effect on language comprehension outcomes (Language and Reading Research Consortium et al., 2022). Building science and social studies topic knowledge takes time in the early grades (Cabell & Hwang, 2020) and most studies examine effects after only 1 year of intervention implementation. Researchers have theorized that direct instruction of vocabulary over time might be enough to cause improvements in reading comprehension outcomes and incidental word learning (Wright & Cervetti, 2017), yet there is little known about the role growth of domain-specific vocabulary knowledge over multiple years plays in explaining outcomes on a continuum of domain-specific reading comprehension measures of transfer. In their systematic review of vocabulary research impacting text comprehension, Wright and Cervetti (2017) argue that vocabulary instruction should include a rich language environment (direct teaching of words, incidental exposure through interactive discussion, supporting word consciousness), yet they "found no studies that engage in this more comprehensive instruction" (p. 223). Consequently, comprehensive knowledge building via networks of vocabulary development is theorized to be essential to reading development and there is a need for empirical studies to examine this theory.

Teacher Practices That Build Depth of Vocabulary Knowledge

Concept Maps

One way teachers can build visual representations of semantic knowledge is to use a concept map. Particularly in the domain of science, using a concept map can help students visualize key concepts as well as the relationships between concepts, thereby enabling students to have a more comprehensive understanding of the overall subject matter (Karpicke & Blunt, 2011; Novak, 1990). Concept maps highlight the interconnectedness of words by linking words with lines and other connective shapes (Figure 2). Concept maps are also flexible in that words can be rearranged to feature certain semantic connections. Put differently, there is no specific way a concept map should be constructed other than it should highlight connections between words. Concept mapping has been shown to be associated with greater knowledge retention and transfer (Nesbit & Adesope, 2006), and in a recent study, first-grade students participating in the MORE intervention who received domain-specific vocabulary instruction using concept maps showed significant transfer effects on proximal measures of vocabulary knowledge for both words that were explicitly taught and words that were not taught but conceptually related to and appeared in the intervention content (Kim et al., 2020, 2021). Indeed, representing word, domain, and topic knowledge through concept maps that make visible the connectedness of words and concepts, allows learners to organize new knowledge in a way that fosters connections, for when it comes to comprehending texts, learners "must form connections between things that were previously disparate" (Kintsch, 1998, p. 93).

Incidental Exposure

In addition to seeing the interconnectedness of words, hearing target vocabulary incidentally (i.e., indirectly) is an essential element of increasing the breadth of exposure students have to a set of domain-specific vocabulary words. Incidental exposure refers to exposure students have to vocabulary words that are not focused on explicitly teaching word meanings. For example, using vocabulary words in conversation or during a read aloud exposes students to target vocabulary without specifically teaching word meanings, thereby giving students opportunities to hear words in varied contexts and settings. As stated in the lexical quality hypothesis, students need to be exposed to high-quality representations of word meanings in both varied contexts and with repeated exposures in order to expand their lexical network (Perfetti, 2007). Exposures to target vocabulary from varied contexts create memory traces that "may resonate or reactivate as fragments of decontextualized knowledge" (Bolger et al., 2008, p. 127). Thus, each encounter with a word creates memories, and if each word is semantically related to other words that share a conceptual link, then the greater the encounters students experience, the more likely they will internalize word meanings and concepts, thereby increasing domain knowledge of a given topic and aiding schema instantiation.

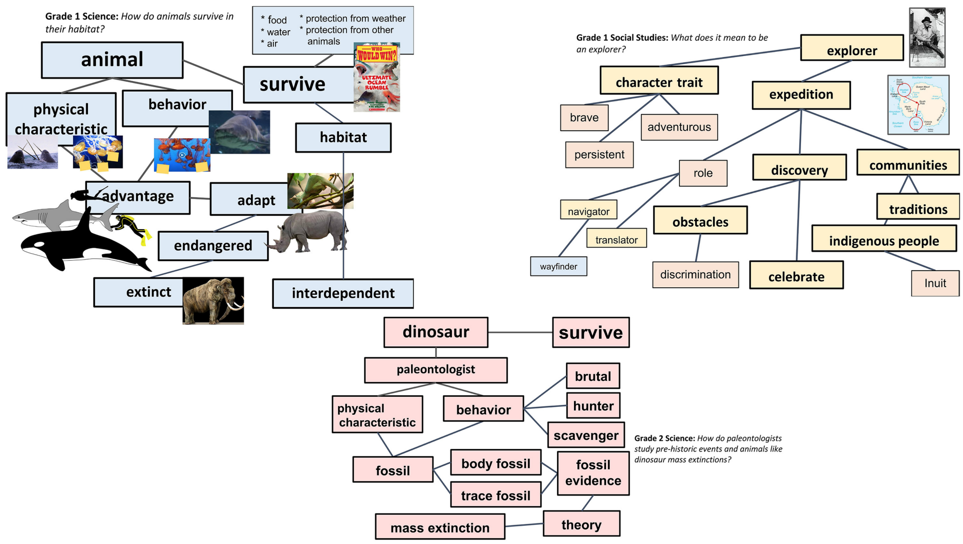
In one study, students who received frequent exposure to target vocabulary scored higher than students who received less frequent exposure, and substantially higher than students who received no exposure (McKeown et al., 1983). In a more recent content literacy intervention, recurrent exposures to target vocabulary resulted in improved domain-specific vocabulary knowledge as well as higher standardized assessment scores (Kim et al., 2020). Furthermore, analyses revealed a significant treatment effect on untaught words—that is, words that were not explicitly taught and that students were only exposed to incidentally through read aloud texts and lesson slides (Kim et al., 2020, 2021, 2022). The treatment effect on vocabulary knowledge of untaught words highlights the impact incidental exposure can have on student vocabulary acquisition, particularly when those words are conceptually related to words receiving explicit instruction.

Language Extensions

While students need to hear target vocabulary with frequency, they also need access to rich, explicit vocabulary instruction. In studies that have found that deviation from an intervention script actually improves student learning, evidence points to teachers' use of language extensions that exceed scripted language as a predictor of student vocabulary knowledge (Neugebauer et al., 2017; Sanetti & Kratochwill, 2009). While many literacy interventions seek to improve instructional practices, there is often variation in the degree that teachers either adhere to or deviate from intervention scripts to address student deficits or needs. Successful interventions are often

Figure 2

Grade 1 and Grade 2 Domain-Specific Vocabulary Concept Maps



Note. See the online article for the color version of this figure.

dependent on how teachers emphasize and favor specific aspects of an intervention over others (Snow, 2015). One particular study showed that teachers who elaborated on program definitions or provided additional definitions for words encountered in an aspect of the intervention resulted in students performing higher on the intervention assessment than students who did not experience language extensions by the teacher (Neugebauer et al., 2017). Even a single additional statement about a word relationship beyond what was scripted was associated with greater student achievement.

If teachers focus solely on a lesson script, they may miss noticing how students respond to instruction and subsequently overlook students' misconceptions. When teachers move past a script, however, they may be able to elaborate further on vocabulary meaning and highlight semantic links between multiple words. It seems plausible that additional statements that move beyond intervention scripts about the interconnectedness of domain-specific vocabulary might be positively associated with greater domain-specific vocabulary knowledge, which in turn may promote greater domain-specific reading comprehension in specific domains.

The Present Study

The intervention theory of change (Figure 1) visually describes the if-then statement that links the intervention core components to the mediator and to the transfer outcomes. The theory of change also explains why domain-specific vocabulary knowledge is likely to be both an immediate consequence of effective instruction and a mediator of transfer in domain-specific reading comprehension. If content literacy instruction helps children build domain and topic schemas (through thematic lessons, concept mapping, and interactive read alouds and wide reading of texts) and transfer those schemas (through argumentative writing, collaborative research, and word inquiry), then students will learn domain-specific vocabulary. And if teachers enact the practices and engage in practices that provide incidental exposure, then students will master the form and meaning of science and social studies words, thus facilitating word-to-text integration processes that support students' comprehension on domain-specific reading comprehension tests with high knowledge and vocabulary demands. In our previous studies, we find evidence that MORE has direct causal effects on domain-specific vocabulary knowledge in Grade 1 (ES = .30; ES = .50 Kim et al., 2020, 2021), and in Grade 2, effects on the untaught words acquired through incidental exposure were significantly larger than words that were explicitly taught (ES = .11, Kim et al., 2022). However, no previous study has attempted to examine the mediating effects of cumulative vocabulary learning from first to second grade on transfer measures of domain-specific comprehension. To test the intervention theory of change, the research questions for the current study were as follows:

> *Research Question 1 (RQ1):* To what extent does student knowledge of domain-specific vocabulary in Grades 1 and 2 mediate effects of the MORE intervention on domain-specific reading comprehension scores in Grade 2?

> *Research Question 2 (RQ2):* How do teachers describe (a) the ways students were exposed to vocabulary in their classrooms, and (b) through learning about their teaching practices, what inductively generated themes emerge that could potentially explain the quantitative results?

Method

We conducted a sequential quantitative–qualitative mixed-methods design to examine the mediating role of students' depth of vocabulary knowledge on a measure of domain-specific reading comprehension transfer and to use survey and interview data to inductively generate themes about potential explanations of the significant quantitative results. The data are drawn from an experimental longitudinal study examining the effects of a content literacy intervention on reading comprehension measures that included 30 schools from a large, urban district in North Carolina (Kim et al., 2022). Schools were stratified by achievement and demographic characteristics and then assigned to either treatment or control conditions. Table 1 displays the descriptive characteristics of students in the study. As the original study was longitudinal in nature, students participated in the intervention as first graders and second graders. For the final analytic sample, there were a total of 2,156 participating students: 1,176 in the treatment condition (55%) and 980 students in the control condition (45%).

Students were primarily Black (35%) and Hispanic (36%), with White (18%), Asian (8%), and Other (Multiracial, American Indian, Pacific Islander, 3%) races/ethnicities also represented. Forty-nine percentage of students were male. The school district divided socioeconomic status into three categories (low, medium, high) that accounted for parental education, home ownership, income, English language ability, and family composition. Students from low socioeconomic backgrounds comprised 40% of the sample, and 24% of students were classified by the district as having limited English proficiency.

The original study included 157 teachers across 30 schools and was longitudinal in nature, with multiple cohorts of teachers and students. At the start of Year 1 of the study (2019), 15 schools were randomly assigned to teach treatment lessons in first grade and another 15 schools were assigned to teach treatment lessons in second grade. In Year 2 of the study (2020), second-grade teachers who had been in the control condition the year prior became treatment teachers while second-grade teachers who had been in the treatment condition the previous year became control teachers. This design allowed students to remain in their assigned condition for the duration of the study. The present study examines only one cohort of students who either received the intervention 2 years in a row (in Grades 1 and 2) or not at all, as well as their (Year 2) second-grade teachers. The attrition rate for students matriculating from first to second grade was 27%, of which there was no statistically significant attrition rate based on the experimental condition (Kim et al., 2022). Due to the nature of the larger study design, most of the "control" teachers in the present study had taught the intervention the previous year, while most of the "treatment" teachers had previously been control teachers. Additionally, there were a handful of cases where because teachers switched grades and/or schools between Years 1 and 2, certain teachers taught the intervention 2 years in a row or not at all.

Table 1

Descriptive Characteristics of the Grade 1 and Grade 2 Sample (N = 2,156)

Characteristic	Overall	Treatment	Control
Teachers			
N	157	82	75
Students			
N	2,156	1,176	980
Male (%)	49	49	49
White (%)	18	16	22
Black (%)	35	36	35
Asian (%)	8	7	9
Hispanic (%)	36	39	31
Other race (%)	3	3	3
ELL (%)	24	25	21
Low SES (%)	40	48	32
Med SES (%)	38	33	44
High SES (%)	21	19	23
Baseline MAP reading, *M* (*SD*)	169.66 (15.65)	167.73 (15.61)	171.98 (15.38)

Note. ELL = english language learners; SES = socioeconomic status; MAP = measure of academic progress; other race = two or more races/Native American.

Transparency and Openness

All data from the original randomized controlled trial (Kim et al., 2022) are available at https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/LAWFFU. The present study was not preregistered, and while the data set is publicly available, the code is not. Please contact Douglas M. Mosher for questions. Quantitative data analyses used Stata 17 and Mplus 7; qualitative analyses used Dedoose.

The Intervention

Students within the treatment condition received the intervention during the science and social studies block while students within the control condition followed the prescribed district science and social studies curriculum. In the first year of the study, first-grade students within the treatment condition received a 10-lesson science unit on animal survival in the arctic and a 10-lesson social studies unit on arctic explorers. The following year, those same students, now in second grade, received a science unit on how paleontologists study dinosaurs. School closures due to COVID-19 prevented the second-grade social studies unit on inventors from being implemented.

The second-grade unit on how paleontologists study dinosaurs was comprised of five nine-lesson cycles for a total of 45 lessons. The first-grade curriculum from the previous year followed a similar sequence. Each cycle began with an activity that focused on building student interest in the topic followed by a read aloud lesson that helped students build knowledge about the given topic. The third lesson focused on key vocabulary, represented visually and conceptually by a concept map. Lesson 4 focused on another read aloud. The fifth lesson targeted notetaking to help students write some of the important conceptual knowledge presented in the read aloud and vocabulary lessons. Lesson 6 used an inquiry-based approach to examine words from the concept map in order to build morphological awareness. The final lessons of each cycle focused on having students answer a research question in collaborative research groups where students engaged with grade-level texts, collaborative discussion, and writing. Key vocabulary words were introduced in all lessons and added to the interactive class concept map.

There were 12 domain-specific vocabulary words for each unit. As first graders, participating students were exposed to 12 first-grade science words and 12 first-grade social studies words. As second graders, students were exposed to 12 second-grade science words. The target words formed the conceptual foundation for each unit, with each unit's content rooted in the meaning of each domain-specific vocabulary word. For all sets of 12 words, students received explicit instruction on seven "taught" words through a concept map. The remaining five "untaught" words were not taught explicitly, but they were conceptually related to the unit topic and students were exposed to the words indirectly. All words appeared in the text and in various lessons, including the "untaught" words. The seven second-grade science "taught" words for the intervention were extinct, fossil, brutal, evidence, theory, hunter, and paleontologist. The five second-grade "untaught" words were carnivore, hypothesis, organism, trait, and reptile. See Table 2 for the list of topic schemas and vocabulary words for each unit and Figure 2 for concept map examples.

Table 2
Intervention Topic-Specific Schemas and Corresponding Key Questions

Schema	Key questions and domain-specific vocabulary
Topic-specific schema in Grade 2 science	How do paleontologists study prehistoric events and animals, like dinosaur mass extinctions? • Taught vocabulary ○ extinct, fossil, brutal, evidence, theory, hunter, paleontologist • Untaught vocabulary ○ carnivore, hypothesis, organism, trait, reptile
Topic-specific schema in Grade 1 science	How do animals survive in their habitat? • Taught vocabulary ○ survive, species, behavior, advantage, adaptation, habitat, physical characteristic • Untaught vocabulary ○ potential, unique, camouflage, complex, diversity
Topic-specific schema in Grade 1 social studies	What does it mean to be an explorer? And why do we celebrate explorers? • Taught vocabulary: ○ expedition, discovery, obstacle, indigenous, explorer, community, persistent • Untaught vocabulary: ○ ancestor, navigation, settle, celebrate, route

While students in the control condition did not participate in the intervention unit on how paleontologists study dinosaurs in second grade, the district literacy curriculum also included a unit on dinosaurs. This contributed to an overlap of concepts and vocabulary between both treatment and control conditions. All students, including those in the treatment condition, participated in the district literacy unit on dinosaurs.

Teacher Subsample Data Collection

As a result of COVID-19 school closures, the research team was unable to collect audio recordings to measure fidelity of implementation. To gain greater insight into how teachers enacted the intervention curriculum, we selected a subsample of second-grade teachers who had indicated an interest in being interviewed on the end-of-intervention survey. Of the 89 teachers who were offered an hour-long interview, 48 teachers agreed to participate and were given a gift card in exchange for their time. Our subsample is not representative of all participating second-grade teachers, but rather those who agreed to an interview. Treatment and control teachers were evenly divided.

Prior to the interviews, we asked teachers to complete a preinterview questionnaire (Figure 3) that asked several questions about whether teachers taught the 12 second-grade target words, how students were exposed to those words (incidental exposure), the level of difficulty teaching the word, and questions regarding the district curriculum. All 48 teachers completed the questionnaire, but one teacher was unable to participate in the follow-up interview. Table 3 includes descriptive characteristics of the 48 teachers. Of the interview sample, 45 teachers identified as female (94%) and three as male (6%).

Teacher interviews were conducted by four members of the research team and lasted an hour. Each interview took place with one member of the research team and one participating teacher. The interview protocol that focused on vocabulary instruction was fairly scripted and drew from Weisner et al.'s (1997) ecocultural framework where we were interested in the culture and ecology of word learning in the classroom as well as to better understand the resources and barriers to promoting word learning. Prior to conducting the interviews, interviewers examined the teacher's preinterview questionnaire to identify which words were marked as explicitly taught and the types of indirect exposures students had to each target word. Then interviewers followed the protocol shown in Figure 4. Interviewers were allowed a degree of flexibility to probe teachers further, but the initial line of questioning was rooted in the aforementioned protocol. All teachers agreed to be audio/video recorded digitally on zoom, and zoom audio transcripts were prepared for qualitative analysis using Dedoose, Version 9.0.54.

Measures

Networks of Domain-Specific Vocabulary Knowledge

Students in first and second grades were given a semantic associations task at the conclusion of the intervention (Read, 2004). After selecting 12 domain-specific vocabulary words that formed the conceptual basis of the intervention lessons, the research team generated a semantic map for each word using an automated concept network

Figure 3
Excerpt From the PreInterview Survey

Think about your vocabulary instruction during the 2019-2020 school year as you answer the following questions.

	Did you explicitly teach the word?		How else were most of your students exposed to this word? (Check all that apply)			
	Yes	No	Saw or heard in a read aloud	Heard the word in a discussion	Used the word in a discussion or in the classroom	Never heard the word or used the word
extinct	○	○	☐	☐	☐	☐
fossil	○	○	☐	☐	☐	☐
brutal	○	○	☐	☐	☐	☐
evidence	○	○	☐	☐	☐	☐
theory	○	○	☐	☐	☐	☐
hunter	○	○	☐	☐	☐	☐
paleontologist	○	○	☐	☐	☐	☐
carnivore	○	○	☐	☐	☐	☐
hypothesis	○	○	☐	☐	☐	☐
organism	○	○	☐	☐	☐	☐
trait	○	○	☐	☐	☐	☐
reptile	○	○	☐	☐	☐	☐

Note. See the online article for the color version of this figure.

Table 3
Descriptive Characteristics of the Grade 2 Teacher Interview Sample

Characteristics	Overall	Treatment	Control
Teachers			
N	48	24	24
Female (%)	94	92	96

drawing on data from the top four best-selling science textbooks. Each semantic map consisted of numerous words represented as nodes with conceptually related words connected to each node. In this task, for each target word (taught and untaught), students were directed to circle the two words corresponding to the target word among four answer choices. The research team used the semantic network to identify two synonymous and conceptually related words for each target word. The semantic associations task included both taught and untaught words, and each item was scored 0–4, with a score of four being the highest possible score and indicating the student circled the two correct words and did not circle the remaining two words. Thus, the Grade 1 social studies and science semantic associations task had a combined maximum score of 96, and the Grade 2 science semantic associations task had a total score of 48. An answer key was used to score the assessment. Grade 1 science, Grade 1 social studies, and Grade 2 science assessments of domain-specific vocabulary knowledge had a Cronbach's α reliability estimate of .76, .80, and .92, respectively.

Domain-Specific Reading Comprehension of Near-, Mid-, Far-Transfer Passages

At the end of the second-grade unit, students took a 20-item domain-specific reading comprehension measure assessing near-, mid-, and far-transfer, with seven items per near-transfer passage and mid-transfer passage. The far-transfer passage had only six items since one item was dropped due to poor item function from item response theory analysis (Kim et al., 2022).

The design of the transfer passages was motivated by the two factors from Barnett and Ceci's (2002) taxonomy of transfer: content (what is being transferred) and context (where the learning is transferred to). Here, we refer to the domain-specific vocabulary words as "content" and the topics of each transfer passage as "context" given that the topic and context of each passage becomes less overtly related to the topic of the intervention when moving from near to far. The near-transfer passage discussed sea ammonites, included six target vocabulary words (content), and was very similar to the intervention topic of how paleontologists study dinosaurs (context). The mid-transfer passage explained how archeologists study the ruins of Pompeii and included four target vocabulary words (content). While conceptually related, the mid-transfer passage contained some distinct changes—substituting the concept of archeologists studying fossilized buildings in place of paleontologists studying dinosaur fossils, thus moving away from prehistoric extinctions (context). Finally, the far-transfer passage focused on genealogists and did not include any target vocabulary words (content). This passage was distantly related to the intervention topic (context), although it still centered on the general schema of scientists who study past events.

Drawing on Kintsch's construction-integration model of reading comprehension (Kintsch, 1988, 2013), students had to merge their literal understanding of the textbase with their prior topic and vocabulary knowledge (Kim et al., 2022). The comprehension items purposely assessed whether or not students could determine the main idea of the text, identify the meaning of words or phrases from the text, recognize correct explanations of scientific concepts, and merge knowledge with ideas, ultimately assessing the extent to which students could leverage the familiar and instantiated schema of the intervention (paleontologists studying dinosaur fossils) when reading and answering questions about unfamiliar and topics (fossils of sea ammonites, ruins of Pompeii, ancestors). Thus, the domain-specific reading comprehension measures of transfer assessed students' ability to comprehend complex texts requiring students to transfer the topic schemas of each transfer passage onto the overarching general schema consistent across the intervention and transfer passages—that of scientists studying past events.

The 20-item assessment had a Cronbach's α reliability estimate of .78. Each item had three answer choices with only one correct answer. In addition to the transfer passages and semantic associations task, all students took the standardized measure of academic progress (MAP) reading assessment at the start of the school year (Northwest Evaluation Association, 2011). Fall MAP scores served as a measure

Figure 4
Teacher Interview Protocol

- ***If explicitly taught:***
  - *Could you tell me more about how you taught the word _____?*
    - *I want to see it through your student's eyes. Walk me through the steps of how you would teach this word.*
- ***Level of difficulty teaching the word:***
  - *You said it was (very easy/easy/difficult/very difficult) to teach the word _____. What made it (very easy/easy/difficult/very difficult) to teach?*
- ***Incidental exposure:*** *You mentioned that students saw or heard the word during a read aloud or discussion.*
  - *How was the word used in discussion?*
  - *Did students need to be prompted?*
  - *Elaborate on what discussions looked like*
- ***If not explicitly taught:***
  - *You marked that you did not teach ______. What made you decide not to teach these words?*

of baseline equivalency. We were unable to obtain end-of-year MAP scores due to COVID-19 school closures.

Data Analysis

Analysis for RQ 1: Effects of the Grade 2 Intervention on Domain-Specific Reading Comprehension Scores Mediated by Domain-Specific Vocabulary Knowledge

To address our first research question, we specified an SEM (Kline, 2016) to examine the extent to which domain-specific vocabulary knowledge mediated the treatment effect on domain-specific reading comprehension scores using Mplus 7. Additionally, confirmatory factor analysis was used to verify the properties of the measurement models. To determine the unit of clustering due to the nested nature of the data, we fit a three-level hierarchical variance components model that included the stratification variable for randomization blocks. Results showed that between-school variance was only 4% and between-teacher variance was 13%. The same three-level model without stratification blocks showed between-school variance was 8%. As a result, we decided to proceed with subsequent analyses by clustering standard errors at the teacher level—the level with the most variance—while also including the stratification blocks to account for between-school variation. To do so, we used the maximum likelihood robust estimator in Mplus 7. We assessed adequate model fit for all models using cutoffs specified by Hu and Bentler (1999): root-mean-square error of approximation (RMSEA) $< .06$, comparative fit index (CFI) and Tucker–Lewis index (TLI) $> .90$, and standardized root-mean-squared residual (SRMR) $< .08$.

Missing Data

There were a total of 269 missing observations from the vocabulary assessment data, which amounted to 12% missing data for the vocabulary measures. All missing values were from the first-grade science and social studies vocabulary assessments. To address this, we conducted Little's missing completely-at-random (MCAR) test (Little, 1988). Results revealed that missing values were MCAR, $\chi^2(4) = 7.17$, $p = .13$. Subsequent SEMs used full-information maximum likelihood estimation to account for missingness. There were no cases of missing data for the subsample of Grade 2 teachers that completed the questionnaire about vocabulary practices.

Analysis for RQ 2: Quantitative Survey Analysis and Qualitative Thematic Analysis of Classroom Teacher Practices

We used ordinary least squares (OLS) regression to determine if there was a difference between treatment and control teachers' self-reported incidental exposure to target vocabulary. The model is as follows:

$$\text{Incidental exposure}_i = \beta_0 + \beta_1\text{treatment}_i + \beta_2\text{female}_i + \beta_3\text{taught}_i + \varepsilon_i, \quad (1)$$

where we controlled for teachers' self-reported scores for teaching target words in addition to controlling for gender. We treated survey results as a proxy for fidelity since the intervention study was disrupted by COVID-19 school closures, preventing the research team from conducting classroom fidelity observations. We hypothesized that teachers with higher levels of incidental exposure would be in the treatment condition given the nature of the intervention lessons.

The qualitative interviews were meant to surface new themes about classroom instruction that might explain the quantitative findings. Themes are represented as extended phrases or sentences identifying what the data is about (Saldaña, 2016). Using incidental exposure as a proxy for high and low fidelity allowed us initially to situate our qualitative analyses within the context of fidelity, although we ultimately looked at the continuum of low to high incidental exposure and explored the practices of those teachers. Before coding the interview transcripts, a list of codes was compiled from the literature on vocabulary instruction as potential codes that would surface in the transcripts, a procedure that appears in the literature (Chin & Phillips, 2004). Interview analyses were conducted using Dedoose software. Drawing from Braun and Clarke's (2006) thematic analysis, we initially adopted a theoretical thematic analysis approach centered on much of the theoretical and empirical research on vocabulary in elementary school grades. We were interested in hearing how teachers described vocabulary instruction and practices in their respective classrooms, and specific interview questions were scripted to gain a detailed understanding of exactly how teachers taught target vocabulary words. While we initially approached coding the interviews with a theoretical background and a set of existing codes, it became apparent early in the coding process that many of the initial codes were irrelevant and numerous themes were beginning to emerge that we had not originally anticipated. Thus, the codebook was heavily altered with over half the original codes being discarded as irrelevant to the interview data. Following Braun and Clarke's (2006) approach, we familiarized ourselves with the data, generated initial codes, searched for themes across codes, reviewed themes, and finally defined and named the themes, which emerged as a result of "repeated patterns of meaning" (p. 86). We also used elements of Ryan and Bernard's (2003) framework in conducting our thematic analysis by focusing in particular on repetitions of key topics, similarities and differences across units of data, and key words in context—that is identifying the most prevalent codes and the context in which those codes appear.

Results

Descriptive Statistics

All descriptive statistics and regression analysis were conducted using Stata 17. Pairwise correlations and descriptive statistics of study variables are listed in Table 4. Baseline reading measured at the start of Grade 1 correlated with all Grade 2 comprehension outcomes: near ($r = .42$, $p < .05$), mid ($r = .50$, $p < .05$), and far ($r = .50$, $p < .05$). Similarly Grade 2 vocabulary words that were not explicitly taught but students received exposure were similarly correlated with near-transfer ($r = .45$, $p < .05$), mid-transfer ($r = .40$, $p < .05$), and far-transfer ($r = .40$, $p < .05$) outcomes. Grade 1 and Grade 2 taught science vocabulary scores were correlated ($r = .28$, $p < .05$) and untaught science vocabulary was slightly less, but still significantly correlated ($r = .23$, $p < .05$).

Table 4
Descriptive Statistics for Analytic Sample (Mean and Standard Deviation) of Study Variables and Pairwise Correlation Matrix

	Treatment		Control		Correlation								
Variable	*M* (*SD*)	*n*	*M* (*SD*)	*n*	1	2	3	4	5	6	7	8	9
Baseline equivalency													
1. MAP pretest	167.73 (15.61)	1,176	171.98 (15.38)	980	—								
G1 sci vocabulary													
2. Taught words	19.70 (3.90)	1,039	19.05 (3.56)	861	**.48**	—							
3. Untaught words	13.85 (2.91)	1,039	13.52 (2.93)	861	**.39**	**.44**	—						
G1 SS vocabulary													
4. Taught words	21.72 (4.59)	1,042	18.78 (3.89)	845	**.47**	**.44**	**.40**	—					
5. Untaught words	13.17 (3.36)	1,042	12.65 (3.19)	845	**.42**	**.40**	**.34**	**.42**	—				
G2 sci vocabulary													
6. Taught words	19.77 (5.82)	1,176	20.64 (5.47)	980	**.52**	**.28**	**.22**	**.32**	**.29**	—			
7. Untaught words	13.50 (4.68)	1,176	13.59 (4.42)	980	**.49**	**.30**	**.23**	**.32**	**.27**	**.70**	—		
Comprehension													
8. Near transfer	3.66 (1.75)	1,176	3.49 (1.58)	980	**.42**	**.28**	**.22**	**.29**	**.24**	**.45**	**.45**	—	
9. Midtransfer	3.66 (1.93)	1,176	3.65 (1.81)	980	**.50**	**.32**	**.27**	**.34**	**.29**	**.40**	**.40**	**.40**	—
10. Far transfer	3.65 (1.83)	1,176	3.83 (1.80)	980	**.50**	**.29**	**.24**	**.29**	**.25**	**.42**	**.40**	**.36**	**.54**

Note. MAP = measure of academic progress; G1 = Grade 1; G2 = Grade 2; SS = social studies; sci = science. All bolded correlations are significant at the .05 level.

Balance tests reported in the original study (Kim et al., 2022) indicated an imbalance in reading favoring the control condition with a standardized mean difference of $-.23$ *SD*s ($SE = 0.09$, $z = -2.45$), which is below the threshold of 0.25 *SD*s established by What Works Clearinghouse Standards (2021). Consequently, to address this imbalance of baseline reading scores, we controlled for MAP in our mediation model.

Teacher characteristics from the interview sample can be found in Table 3. Descriptive statistics and pairwise correlations of study variables for the teacher interviews are displayed in Table 5. For the interview subsample of teachers, the self-reported totals of having explicitly taught target vocabulary were significantly correlated with self-reported incidental exposure ($r = .65$, $p < .05$).

Measurement Models for Domain-Specific Vocabulary Knowledge and Domain-Specific Reading Comprehension

All measurement models and subsequent mediation models were done using Mplus 7 (contact Douglas M. Mosher for details). Because our domain-specific reading comprehension measure was comprised of near-, mid-, and far-transfer passages and questions, domain-specific reading comprehension was represented as a latent variable with sum scores for near, mid, and far sections serving as continuous indicator variables (Figure 5). Fit statistics for the measurement model were just identified. Thus, there are no goodness of fit statistics to report. All three factor loadings were statistically significant and greater than .50.

We represented both Grade 1 and 2 domain-specific vocabulary knowledge as separate latent constructs with sum scores for taught and untaught words serving as the indicator variables (Figure 6). Fit statistics for the vocabulary measurement models were adequate: RMSEA = .032, CFI = .995, TFI = .991, and SRMR = .016. All factor loadings were statistically significant and greater than .60.

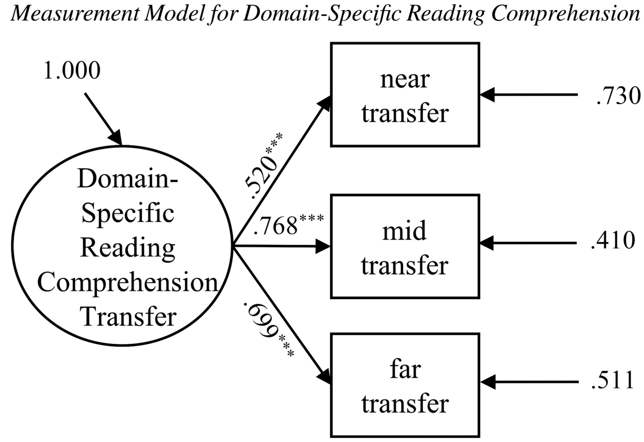
We fit a partial mediation model as well as a full mediation model. The fit for the data was adequate for both models: partial mediation (RMSEA = .036, CFI = .938, TLI = .924, SRMR = .026) and full mediation (RMSEA = .036, CFI = .938, TLI = .926, SRMR = .026). After calculating the Satorra-Bentler chi square difference test (Satorra & Bentler, 2010), we determined that there was no difference in

Table 5
Descriptive Statistics for Analytic Teacher Interview Sample (Mean and Standard Deviation) of Study Variables and Pairwise Correlation Matrix

	Treatment		Control		Correlation
Variable	*M* (*SD*)	*n*	*M* (*SD*)	*n*	1
1. Taught vocabulary	8.54 (2.34)	24	7.83 (3.06)	24	—
2. Incidental exposure	28.29 (5.43)	24	24.42 (5.79)	24	**.65**

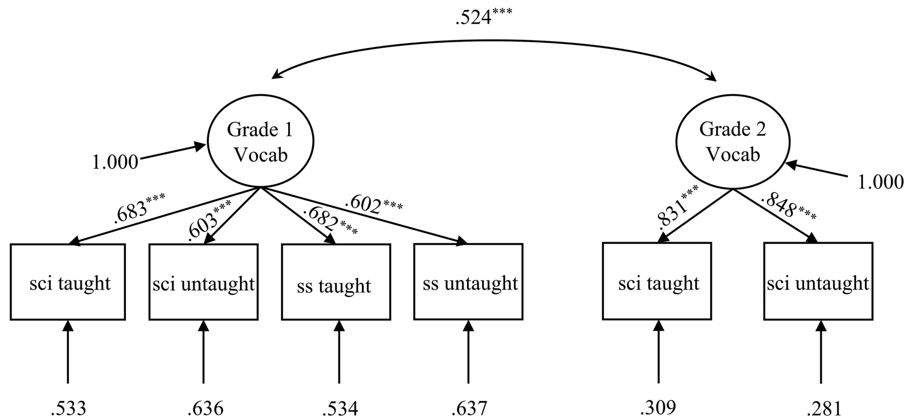
Note. Bolded correlation significant at the .05 level.

Figure 5
Measurement Model for Domain-Specific Reading Comprehension



Note. All coefficients are standardized. Significant coefficients are starred, and *p* values taken from unstandardized values.
*** $p < .001$.

Figure 6
Measurement Model for Grade 1 and Grade 2 Depth of Domain-Specific Vocabulary Knowledge



Note. All coefficients are standardized. Significant coefficients are starred, and *p* values taken from unstandardized values. sci = science; ss = social studies; vocab = vocabulary.
*** $p < .001$.

model fit, $\chi^2(1) = 1.31$, $p = .25$. Even though full mediation cannot not be ruled out, theoretically, it seems implausible that vocabulary would explain the entirety of the treatment effect when there are other unobserved factors. We proceed by interpreting the paths of a partial mediation model shown in Figure 7.

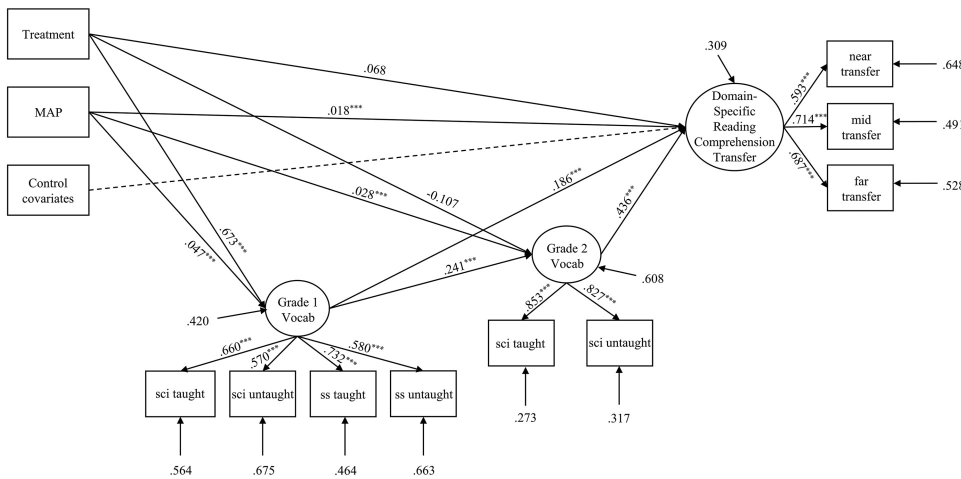
RQ 1: Effects of Grade 2 Intervention on Domain-Specific Reading Comprehension Scores Mediated by Domain-Specific Vocabulary Knowledge

To answer our first research question of whether domain-specific vocabulary knowledge mediated the treatment effect of the MORE intervention on students' domain-specific reading comprehension, we used SEM to fit a mediation model (Figure 7). The model includes standardized coefficients with statistical significance derived from unstandardized coefficients. In our model, we included freely estimated paths between random assignment to the intervention and students' domain-specific reading comprehension outcomes as well as Grade 1 and Grade 2 vocabulary knowledge. We also included paths between both vocabulary factors (i.e., Grade 1 predicting Grade 2) as well as paths between vocabulary factors and Grade 2 domain-specific reading comprehension. Collectively, these paths allowed us to test the hypothesis that vocabulary in Grades 1 and 2 served as a mediator in the effect of the MORE intervention on student domain-specific reading comprehension. Baseline reading, as measured by the MAP assessment, was also included as a key predictor of domain-specific reading comprehension as well as Grade 1 and Grade 2 domain-specific vocabulary knowledge. We included MAP as a key predictor for two reasons: baseline reading was moderately correlated with vocabulary knowledge and the imbalance of baseline equivalence that favored the control condition. Additionally, we controlled for student demographics, socioeconomic status, gender, English proficiency, and intervention randomization blocks when predicting domain-specific reading comprehension.

Our hypothesis that domain-specific vocabulary knowledge mediates the relationship between the intervention and domain-specific reading comprehension was supported by the data, with Grade 1 and Grade 2 vocabulary knowledge explaining 69% of the relation. Given that the MORE intervention emphasizes schema building and that students' depth of vocabulary knowledge in Grade 1 helped to lay the foundation for further knowledge acquisition in Grade 2, we see evidence that related vocabulary knowledge is a key component in explaining the treatment effect, as indicated by the significant direct path from treatment to Grade 1 vocabulary ($b = 1.040$, $SE = 0.098$, $p < .001$, $\beta = .673$) and Grade 1 vocabulary to domain-specific reading comprehension ($b = 0.217$, $SE = 0.051$, $p < .001$, $\beta = .186$). Unsurprisingly, there was no significant direct path from treatment to Grade 2 vocabulary ($b = -0.137$, $SE = 0.086$, $p = .111$, $\beta = -.107$). While students within the control condition did not participate in the intervention lessons on paleontologists and dinosaurs, both treatment and control students did learn about dinosaurs in the prescribed district literacy curriculum in Grade 2. There was no significant difference between treatment and control students' performance on the Grade 2 vocabulary assessment, although Grade 2 vocabulary knowledge did help to explain the relation between treatment and domain-specific reading comprehension ($b = 0.611$, $SE = 0.071$, $p < .001$, $\beta = .436$). The indirect effect of Grade 1 domain-specific vocabulary knowledge on domain-specific reading comprehension was statistically significant ($b = 0.226$, $SE = 0.061$, $p < .001$, $\beta = .126$) and explained 58% of the treatment effect. While the indirect effect of solely Grade 2 domain-specific vocabulary knowledge on domain-specific reading comprehension was not statistically significant ($b = -0.084$, $SE = 0.051$, $p = .101$, $\beta = -.047$), the effect of both Grade 1 and Grade 2 vocabulary knowledge was significant ($b = 0.127$, $SE = 0.033$, $p < .001$, $\beta = .071$) as was the total indirect effect ($b = 0.269$, $SE = 0.081$, $p < .01$, $\beta = .150$). With a null direct effect of the intervention on domain-specific reading comprehension scores ($b = 0.123$, $SE = 0.118$, $p = .295$, $\beta = .068$), our results suggest indirect-only mediation (Zhao et al., 2010). In sum, the cumulative effect of increased depth of domain-specific vocabulary knowledge from Grade 1 to Grade 2 mediated the overall treatment effect on domain-specific reading comprehension.

Figure 7

Structural Equation Model Examining the Mediating Role of Grade 1 and Grade 2 Depth of Domain-Specific Vocabulary Knowledge



Note. All coefficients are standardized. Significant coefficients are starred, and p values taken from unstandardized values. Control covariates include stratification blocks, gender, English language learner status, SES, and race/ethnicity. SES = socioeconomic status; MAP = measure of academic progress; sci = science; ss = social studies; vocab = vocabulary.
*** $p < .001$.

We conducted a number of sensitivity checks using single imputation, listwise deletion, and bootstrapping using 1,000 draws. Results were stable and there was no evidence of full mediation (contact Douglas M. Mosher for details).

RQ 2: Quantitative Survey Analysis and Qualitative Thematic Analysis of Classroom Teacher Practices

Quantitative Analyses

To answer our second research question, we first used OLS regression to analyze the teacher survey data to determine if there were reported differences of incidental exposure—that is the various exposures students had to target words as reported by participating teachers. Results (Table 6) indicated significant differences of self-reported incidental exposure between treatment and control conditions, with treatment teachers reporting a half standard deviation of greater exposure ($\beta_1 = .54$, $SE = 0.20$, $p < .01$).

Table 6

Results of Regression Model Predicting Main Effect of Content Literacy Intervention on Self-Reported Incidental Exposure to Vocabulary

Variable	Total incidental exposure
Treatment	0.544 (0.202)**
Female	1.082 (0.413)*
Taught words	0.610 (0.102)***
Intercept	−1.286 (0.421)**
N	48
F	18.29
R^2	.555

Note. Standard errors in parentheses. Taught words = number of target words teachers self-reported explicitly teaching on a scale from 0 to 12.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Qualitative Analyses

We conducted semistructured interviews with treatment and control teachers to investigate how teachers who self-reported high and low levels of incidental exposure described their vocabulary instruction. In the process of examining teachers with both high and low levels of incidental exposure, an overarching theme emerged from etic and emic codes: teachers engage in rich practices to make word meaning more accessible. From the superordinate theme emerged three dimensions: (a) expanded opportunities to engage in word use, (b) connecting words to topics, and (c) sustainability of intervention practices. In our analysis below, we include each teacher's self-reported incidental exposure score. All values are measured in standard deviation units and teacher identities are masked with pseudonyms. We also indicate whether quoted teachers were in the treatment or control condition for the present study, and when relevant, we also indicate assignment to treatment or control in the prior year.

Theme 1: Expanded Opportunities to Engage in Word Use. When asked to elaborate on their survey responses regarding the three types of scenarios students may have been exposed to target vocabulary (Figure 3), 83% of teachers discussed how students heard or saw the words in texts from both the intervention and district curriculum. One treatment teacher, Colleen ($M = -0.40$), mentioned that "all of the words on the list, those were definitely in a read aloud at least once," while another treatment teacher, Janiece ($M = 0.62$), said that the read aloud book did "a decent job explaining what the word theory meant for the kids to be able to kind of understand that it's an idea [...] That was a MORE book."

Eighty-nine percentage of teachers mentioned that students heard target vocabulary from discussions that occurred in their classrooms. Thus, the read alouds often spurred conversations about target vocabulary, many of which provided additional lexical representations to make word knowledge more accessible to students. Claire ($M = 0.96$) from the treatment condition made a point of highlighting, "We discussed it (vocabulary). Every time it came up in a book—'what does that word mean?'—again, you know, just getting it in their head." Similarly, Emily ($M = 1.64$) also from the treatment condition spoke of how when one target word came up, the students discussed the word with each other:

> I know there were texts within EL [district curriculum] and MORE that the word *theory* came up and then we would pause. I definitely had them use a picture or I definitely had them using conversation with each other: "Try and turn and talk and what is your *theory* on this?" I kind of gave them some examples.

In addition to numerous exposures to target words in read alouds, students encountered target vocabulary when reading texts in collaborative research groups and used the language in their discussions. Janiece ($M = 0.28$) describes one such scenario:

> When they had a minute to share or do the round table [collaborative group discussion protocol], you could tell that the kids would listen. The ones who were really excited would go first and then the children who weren't so interested in it would kind of repeat what they said and picked up off what their partner was saying. I feel like that helped them to gain some knowledge about those particular vocabulary words and those topics. I mean, they all knew about dinosaurs. Like I said, some of those vocabulary words like *extinct* they were able to pick up from their peers explaining it.

Incidental exposure not only occurred regularly but often prompted language extensions where teachers provided additional information about target vocabulary words for their students to better comprehend the text. Sixty-four percent of teachers discussed providing some sort of additional explanation of words, which spanned both teachers in treatment and control conditions. Treatment teacher Tanisha ($M = 0.79$) described how the initial mention of a word in a text resulted in a mini-vocabulary lesson that included her providing an explanation of the word, but also her students explaining the word, providing examples, and most importantly, using the word in their discussion:

> The very first time it [vocabulary] ever came up in a read aloud, I stopped and said, "Who can tell me what this or who already knows what this word means?" […] And they said it and I repeated it. […] I said, "Does anyone want to add on to that?" And you know they shared, dinosaurs are extinct or some of them were like "oh there's other things that are *extinct* like mammoths and stuff like that." They had a lot of background information, so they quickly caught on to what it meant.

Teachers who reported low and high incidental exposure both deviated frequently from the lesson script and provided explanations of vocabulary words beyond the curriculum. Morgan (control, $M = -1.59$), tried to provide greater visualizations of the word fossil so she "pulled up pictures of fossils and excavation sites to let them (students) see some fossils." Similarly, Janiece (treatment, $M = 0.28$) described how providing more visualizations served as the catalyst for even greater explanation of the target word:

> Sometimes I tend to go off on a tangent and I'll do a Google search to show them more of the fossils, but also explaining to them that a fossil is not just a footprint, you know in the mud or something like that. It's been there for years […] Talking about how it's being compounded by dirt and sand and things like that. So for us it was using the pictures in connection with the definition.

Other teachers relied on connecting words to the lived experiences of their students. Angie (treatment, $M = 0.45$), when attempting to explain evidence, tried to make her explanation relevant: "What evidence might there be if you say you are sick? What is the evidence of that? Oh, I am getting a tissue and I am blowing my nose or you know, I try to give a lot of different life examples to the kids to try and use evidence in a sentence." Similarly, Alice (control, $M = 1.64$), taught the word trait by stating that the word was "more like a state of being. When you see a person, how would you describe them? What are some of the things you notice about them? What are some of the things that attract you to that person and that type of thing."

In the absence of connecting target words to the lived experiences of students, teachers used familiar scenarios to situate the explanation of a word. For the word evidence, Janelle (treatment, $M = 0.62$) described the following:

> I have to relate to the kids. So, I remember saying if there was a bank robbery, how would the police know who did it? They need to look for *evidence*. And then I remember talking to them about the video cameras, the fingerprints, and stuff like that. And then I went on to say, well, *paleontologists* are investigators too because they have to look for *evidence*, but they just look for *evidence* of dinosaurs or ancient things.

While there were significant differences between treatment and control teachers' self-reporting of incidental exposure that favored the treatment condition shown from the survey analyses, the interview data suggests that while the quantity of incidental exposure varied, the quality was consistent across conditions.

Theme 2: Connecting Words to Topics. One of the core components of the MORE intervention is teaching words in networks. Our analyses revealed that many teachers with higher self-reported levels of incidental exposure described making connections between vocabulary words while teachers with lower self-reported levels of incidental exposure tended to describe teaching words in isolation. By teaching words in networks, teachers were not only building depth of domain-specific vocabulary knowledge, but building student topic knowledge that may have impacted schema instantiation. Aubrey ($M = -1.25$), who was in the treatment condition mentioned that the concept map helped students see "that all these words are connected to each other" during her vocabulary instruction throughout the MORE unit. Even though she had a below-average incidental exposure score, suggesting low fidelity of intervention practices, Aubrey explained how the concept map—an integral component of the MORE intervention—helped her teach words in networks. Similarly, Katy ($M = 0.79$), a control teacher in the present study who had previous experience teaching MORE in prior years and continued to use intervention practices even as a control teacher, described how the concept map impacted her way of teaching vocabulary:

> I used the concept map every day. (…) I had pictures up there of the read aloud and the *paleontologists* that they studied (…) It was my whole board. I took it to the next extreme I guess. I had pipe cleaners arrowing and everything. (…) The concept is that it's all interrelated and they all

> feed off of each other. There is no one right answer, it could go here, but it could also go here.

In direct contrast, Jackie ($M = -1.42$), another teacher in the control condition, spoke of how she visually represented vocabulary for her students to see while teaching the district curriculum: "I did posters, I had anchor charts with a picture and maybe the definition of the word." Unlike Aubrey and Katy, Jackie illustrates teaching vocabulary in isolation by using a picture and definition. While students may learn the word meanings this way, this method of instruction does not highlight conceptual connections between words.

In discussing how teachers introduced vocabulary words, many teachers similarly spoke of how introducing one word led to connecting it to other target words on the concept map. For example, Janelle ($M = 0.62$) in the treatment condition mentioned:

> Well, *evidence* goes hand in hand with *fossils*, because a *fossil* is *evidence* that the dinosaurs were here. So, I want to say when I introduced the word *fossil*, the word *evidence* kind of came in and then it also comes in with *theory* because you have to have *evidence* of these different *theories*.

Here, Janelle has indicated that in her explanation of one word, *evidence*, she is in effect connecting it to three other words. Instead of teaching words in networks, however, other control teachers like Jackie focused on one word when introducing vocabulary: "So theory is something that you think is going to happen. You know, I try to teach them, first of all, you give them the book knowledge, the definition, the Webster's definition." In this example, vocabulary is taught in isolation, preventing potential connections to other words and concepts.

Some teachers commented on how they tried to connect words from the previous year's first-grade concept map to the current second-grade concept map. According to Emily ($M = 1.64$) who had been a first-grade teacher in the treatment condition the prior year before moving to second grade as a treatment teacher for the present study, "From last year when we taught MORE, (. . .) endangered was on our concept map last year. And so we talked about endangered and extinct. So they were able to really pull from the last year's science." By highlighting the connection between these two words, Emily helped her students establish a link connecting the second-grade science schema to the first-grade science schema on animal survival in the arctic. Furthermore, teachers actively engaging in teaching words in networks touched on how making these connections between words provided a more robust schema of the concept of how paleontologists study dinosaurs:

> You know, I'm not super familiar with that but with my background knowledge, I'm just explaining that a *paleontologist* is a scientist who studies things from long ago and we typically think of it as someone who finds *fossils* and *dinosaurs*, but it could be other animals, it could be plants, shells, rocks, variety of things. And it could be things from like millions and millions of years ago, from thousands or hundreds of years ago. It's a pretty broad term or it can be, someone who is a paleontologist can study a broad range of topics. (Gaby; control, treatment the previous year, $M = -0.23$)

In contrast to teaching words in networks, Dave ($M = -0.57$) described a process of explicitly explaining the word without connecting it to other related words—in essence providing explicit instruction in isolation:

> So that one (*extinct*) would have been the one where we walked through, and we talked about specific examples. And we even talked about the scientific definition of what extinct is and then have kids put it into their own words. So that's how I taught that one.

Theme 3: Sustainability of Intervention Practices. One final theme that emerged during the coding process was that of control teachers in the present study explaining how they continued to use intervention practices after having participated as treatment teachers the year prior. Given the complex design of the larger study, many second-grade control teachers in the present study had served as treatment teachers the prior year. A number of these teachers detailed how they incorporated elements of the MORE intervention into non-intervention related lessons, particularly related to word learning.

> Talking about *paleontologists* making discoveries and *evidence*, *theory* kind of might have snuck in there just as a way to bridge the gap between *evidence* and ideas. Things like that. But it wasn't explicit. [...] I don't remember it [*theory*] being in the [district] curriculum but having the MORE background from the year before, I couldn't help pulling in some things just kind of naturally. But I don't remember it being in the curriculum. (Gaby, $M = -0.23$)

While there was some overlap of words from the MORE intervention and the district curriculum, there were a handful of intervention words that did not appear in the control condition. Nevertheless, Grade 2 control teachers with previous exposure to the intervention sometimes incorporated those intervention words into their discourse as Gaby demonstrated.

One of the prominent instructional practices from the intervention was the use of a concept map to help students see the connections between vocabulary words. Even though the concept map was not a part of the district curriculum, control teachers drew upon their past experiences with the MORE intervention and continued to utilize it while participating in the control condition.

> I actually used—because your study, the MORE study—it was very eye opening the year before for the concept map. So, I actually kept a lot of those words and then added whatever the EL [district] curriculum added and I just kind of used it the same way. We talked about the word, we kind of gave a definition. We put it on the wall, and I always made sure I put a picture because EL didn't have a picture, so I would print out extra pictures. (Sonya, $M = -1.76$)

While the concept map functioned as a tool to help teachers organize vocabulary for their students, a few Grade 2 control teachers spoke to the conceptual reason for why the concept map had changed the way they approached teaching vocabulary.

> All of this [vocabulary] would always be on the concept map and then sometimes, I'd let the kids decide where it goes. Because we have the word *paleontology* up there, you know, so then we've kind of figured out how they are all interconnected. So that was—that's something I learned in MORE last year that I did incorporate in EL because it was really relevant. I used the concept map every day. It was my whole board. (Katy, $M = 0.79$)

Here, Katy speaks to the idea of word networks and how the concept map functioned as a platform for helping her students see those connections. Similarly, Eve ($M = -0.74$) further emphasized how the concept map and her past experiences with the MORE intervention helped her see the value in making connections between words:

> Basically what we did was use what MORE taught us. […] EL was more like a word wall. It [vocabulary] was introduced almost as a slide with a definition and you would use that word all the time. It was more like a word wall than concept mapping—tying it all together—which is how I perceived concept mapping—to connect those dots.

This select group of teachers who were control group participants in Grade 2 and treatment group participants in Grade 1 seems to have internalized one of the core concepts of the intervention—making connections between words. Their inclusion of the concept map and understanding of its significance is evident in their descriptions of how they used the concept map and how they continued to implement intervention practices even when teaching a curriculum separate from MORE.

Discussion

Using mixed methods, this study aimed to examine a key mechanism of the MORE intervention (Kim et al., 2022) that led to positive treatment effects on second-grade domain-specific reading comprehension measures. We examined from a variety of different angles the extent to which (a) expansion of student domain-specific vocabulary knowledge over the course of 2 years promoted transfer on measures of domain-specific reading comprehension; and (b) the inductively generated themes that surfaced from teachers describing their vocabulary practices and how these themes could potentially explain the results from the mediation analysis.

Domain-Specific Vocabulary as a Mediator for Domain-Specific Reading Comprehension

SEM results indicated that both Grade 1 and Grade 2 vocabulary knowledge explained 69% of the relation between random assignment to the MORE intervention and students' domain-specific reading comprehension scores. Importantly, our finding is consistent with results emerging from the K-3 RfU projects, which also found that vocabulary was the most consistent mediator of Tier I curricular effects on reading comprehension (Connor et al., 2017; Language and Reading Research Consortium et al., 2019). What is the significance of these convergent findings?

The MORE intervention centers around building and expanding on schemas comprised of domain-specific vocabulary words that represent the tip of the conceptual iceberg (i.e., schema) (Anderson & Freebody, 1981). Each vocabulary word conveys relevant domain and topic knowledge (i.e., paleontologists, fossil) and represents the foundation of knowledge acquisition. Understanding these words helps students access the texts and key concepts discussed in each thematic unit. Thus, if domain-specific vocabulary words are at the core of the intervention and represent the visible components of the instantiated schemas, we would expect vocabulary knowledge to play a substantial role in explaining the treatment effect on the Grade 2 domain-specific reading comprehension transfer outcomes.

In order to access the near-, mid-, and far-transfer second-grade comprehension passages, readers must construct their situation model, requiring a synthesis of textual information and "relevant prior knowledge" (Kintsch & Kintsch, 2005, p. 73). That "prior knowledge" are the instantiated topic schemas from the second-grade intervention (how paleontologists study dinosaurs) and first-grade thematically related topic schemas (animal survival, arctic explorers) (Kimball & Holyoak, 2000). By anchoring the topics of each unit in domain-specific vocabulary, recognizing these words in the transfer passages helps students activate the topic-schema and enhance their situation model (Ericsson & Kintsch, 1995; Kintsch & Kintsch, 2005). As students master the form and meaning of words (Perfetti, 2007), they begin to know a word by the company it keeps (Firth, 1957). As networks of domain-specific vocabulary knowledge grow, students are able to instantiate topic schemas that they can then map onto a continuum of related topic schemas (i.e., transfer passages: paleontologists studying sea ammonites, archeologists studying the ancient city of Pompeii) (Gick & Holyoak, 1983; Kimball & Holyoak, 2000).

Each unit of the MORE intervention focuses on seven words that are explicitly taught and five semantically related words that appear in the lessons and books frequently but do not receive explicit instruction. In contrast, the second-grade district curriculum on dinosaurs and paleontologists explicitly teaches around 20 words. Instead of learning numerous challenging words from the texts, students in the treatment condition received targeted instruction on key domain-specific words that not only helped them acquire adequate domain knowledge, but also enabled them to develop a thematic schema. Furthermore, students in the treatment condition had the added benefit of expanding on existing general schemas and networks of words from first grade as evidenced by the fact that Grade 1 vocabulary explained 58% of the treatment effect. The Grade 2 schema builds on the Grade 1 science schema of animal survival by discussing dinosaur survival and their ultimate extinction in second grade. The Grade 2 schema also builds on the first-grade social studies schema of explorers in that paleontologists too are explorers that go on expeditions to discover dinosaur fossils (Figure 2). Indeed, domain-specific vocabulary words are the visible tip of the conceptual iceberg (Anderson & Freebody, 1981), and our findings suggests that developing and expanding topic and general schemas through the acquisition of networks of domain-specific vocabulary knowledge contribute to students' ability to read and understand texts "that vary in surface characteristics from previously encountered situations (i.e., intervention lessons)" (Kimball & Holyoak, 2000, p. 118). Even though only a few first-grade vocabulary words appeared in the Grade 2 transfer assessments, the conceptual connection of all Grade 1 words to the Grade 2 words and schema contributed to a more robust knowledge base among treatment students that could be leveraged when reading the different and unfamiliar transfer passages (Myers & O'Brien, 1998).

One remaining question lingers: If there was no significant treatment effect on vocabulary knowledge in second grade, why was there a significant treatment effect on the domain-specific reading comprehension transfer passages? We argue that even though students in both treatment and control conditions could recognize the two correct words accompanying each second-grade target word on the semantic associations assessment, the fact that second graders in the treatment condition outperformed control students on domain-specific reading comprehension suggests that treatment students were able to (a) access the topic schema (dinosaurs studying dinosaur fossils), which is comprised of the domain-specific vocabulary words representing the topic knowledge of the unit, and (b) transfer that schema to other, less familiar topic schemas (paleontologists studying sea ammonites, archeologists studying the ancient city of Pompeii, genealogists studying our ancestors). Put differently, the control students understood the word meanings, but the treatment students understood the word meanings and the larger concepts attached to them (i.e., general schema of scientists studying past events).

We sought to dig deeper into the vocabulary practices of teachers participating in the study by administering a survey and conducting qualitative interviews to address questions left unanswered by the quantitative analyses regarding the role of teacher vocabulary instruction as a lever of change.

Self-Reported Incidental Exposure From Survey Analysis

To understand the effects of content literacy instruction on domain-specific vocabulary instruction, we examined treatment-control differences in how teachers self-reported the types of exposures students received to each vocabulary word using a survey. We theorized that teachers' self-reporting of incidental exposure could serve as a proxy for fidelity where we would expect treatment teachers to report greater levels of incidental exposure given that the intervention lessons purposely embed target vocabulary in both the lesson slides and lessons texts. Results indicated that there were differences in the quantity of incidental exposure students received, although, as the interview data revealed, greater incidental exposure did not necessarily serve as a proxy for intervention practices or fidelity. This coarse measure simply revealed that treatment teachers provided greater exposure to the target vocabulary. Nevertheless, the difference in self-report of both taught and untaught words is consistent with the intervention findings where a treatment effect was detected on untaught words (Kim et al., 2022).

Teacher Interviews Reveal Insight Into Rich Practices to Make Word Meanings More Accessible

Conducting teacher interviews allowed us to expand on the quantitative survey results to surface a number of inductively generated themes highlighting teacher practices that build students' depth of vocabulary knowledge. Unscripted language extensions of key vocabulary may have played an important role in helping students fully grasp the meaning of target vocabulary (Neugebauer et al., 2017), thereby helping students develop high-quality lexical representations of word meanings in a variety of contexts (Perfetti, 2007). Teachers with both high and low levels of self-reported incidental exposure described deviating from the lesson scripts and making vocabulary accessible to students with contextual and relevant examples and helping students establish personal connections with the words. By fostering deep understanding rather than emphasizing surface-level definitions and quick memorization, teachers may have impacted students' ability to apply this knowledge on the transfer tasks (Bransford et al., 2000; Kendeou & O'Brien, 2016).

Many teachers with higher levels of self-reported incidental exposure described making connections between words and concepts. By making these distinct connections, teachers were providing high-quality lexical representations of words that allowed for varied representations of meaning, which may have allowed for the necessary meaning retrieval when reading the domain-specific reading comprehension passages (Perfetti, 2007). Furthermore, teaching words in networks may have also helped students instantiate the second-grade topic schema, with some teachers making connections to previous vocabulary words from the previous year (some teachers taught first grade the prior year), thereby connecting first-grade topic schemas to the second-grade science schema.

While some Grade 2 control teachers described teaching words in networks, many taught words in isolation. That is, they provided strong definitions and visuals of word meanings but did not help students see the conceptual links with other target words and concepts. Students who received vocabulary instruction emphasizing connections between words, and by proxy concepts, may have acquired the adequate depth of knowledge needed to understand these words in different contexts when reading the domain-specific comprehension passages. Making strong conceptual connections between words and concepts may have also enabled students to access the necessary schema to support meaning making when reading passages not directly tied to dinosaurs (Kimball & Holyoak, 2000). The qualitative analyses suggest that instruction that centers on teaching words in networks and connecting words to topics has the potential to help students develop robust schemas.

One dimension of scale is that of spread (Coburn, 2003), and in the context of intervention research, one question often lingers: Will the practices of an intervention endure once the program has ended? Interview data revealed that in fact the teacher moves discussed promote spread over time and across contexts. Because there were multiple cohorts of students, many control teachers participating in the present study had been randomly assigned to the treatment condition in prior years and thus had exposure to various intervention components. Analyses revealed that a subset of second-grade control teachers in the present study continued to use MORE intervention practices such as teaching words in networks via the concept map in the domains of science, English language arts, and social studies, thereby indicating that elements of the intervention continued to endure within classrooms and across subjects.

As shown in the theory of change (Figure 1), the combination of the findings from the quantitative mediation analysis and qualitative interview analyses suggest that what students learn may depend heavily on the rich word-learning practices teachers employ. While the quantitative findings confirm the role domain-specific vocabulary knowledge plays in aiding domain-specific reading comprehension, the qualitative themes suggest potential teacher practices that may enhance word learning and reading comprehension, although future research is needed to test the effectiveness of these practices.

Study Limitations and Future Research

Although the findings from this study provide initial support for our theory of change for a sustained content literacy intervention, future research should address the joint influence of both students' vocabulary learning and teachers' quality of vocabulary instruction on reading comprehension outcomes. MORE is a whole-school, multiyear intervention (American Institutes for Research, 2017–2018; Durlak et al., 2011, Quint et al., 2015; Vaughn et al., 2015) in which teachers in treatment and control classrooms implement different instructional frameworks and practices to improve students' reading comprehension. Thus, our theory of change (Figure 1) included both student-level mediators (domains-specific vocabulary knowledge) and teacher/classroom-level mediators (quality of vocabulary instruction). Due to COVID-19 school closures, however, we were unable to directly assess the quality of teachers' vocabulary instruction through classroom observations and audio recordings of lessons. Consistent with other classroom-based interventions that examine the mediating effects of vocabulary learning on students' reading comprehension (Clarke et al. 2010; Cromley & Azevedo, 2007; Language and

Reading Research Consortium et al., 2019, 2022), it is critical to address this limitation with multilevel SEMs (Preacher et al., 2010) that jointly model the mediating effects of both Level 1 student variables and Level 2 classroom-teacher variables. Doing so would paint a richer and perhaps more complete picture of the dynamic student and teacher variables that interact to ultimately improve students' reading comprehension.

Second, mediation analyses imply a causal relationship, yet because we cannot randomly assign the mediator, findings from our analyses are correlational and thus a limitation. In fact, the interview data suggests that within vocabulary instruction, there are many causal levers. To address this limitation, future research should explore randomly assigning teachers to the mechanism—for example, scripted supplements that support word learning—to treatment and control classrooms. Doing so would help isolate whether teacher language—in particular, teachers' ability to enact vocabulary and language extensions that support students' word knowledge—is the active ingredient driving both vocabulary acquisition and reading comprehension.

Third, there is a need for more mixed-methods studies in literacy research. While purely quantitative studies are critical to furthering the scientific understanding of vocabulary development and reading comprehension, they cannot generate new exploratory hypotheses and themes. The overarching theme of rich teacher practices for vocabulary instruction that emerged from the qualitative interviews not only expands on the quantitative findings but also provides insight into how teacher instruction can move beyond the intervention script to provide relevant and impactful lexical representations.

Implications for Practice

There are three main implications from the present study that are directly applicable to practice. First, domain-specific vocabulary instruction via networks is essential to schema instantiation, which allows students to organize and retrieve relevant domain and topic knowledge when encountering new concepts. Results of the mediation analysis suggest that careful selection of a targeted group of domain-specific words are essential to helping students build related schemas over consecutive years that foster transfer on domain-specific reading comprehension measures. Second, building and expanding schemas by way of acquiring networks of domain-specific vocabulary takes time, and we see the impact of this through the significant indirect effect of first-grade domain-specific vocabulary knowledge on measures of domain-specific reading comprehension transfer in second grade. Making connections between past units and expanding on schemas while building new schemas through the acquisition of domain-specific vocabulary is essential to helping students access, retrieve, and apply knowledge. Finally, while curriculums and intervention programs attempt to build in adequate explanations of words, qualitative interviews suggest that teachers engage in numerous practices that make vocabulary knowledge accessible to their students: extended opportunities to learn words and connecting words to topics.

Conclusion

There have only been a handful of studies that have examined vocabulary mediating treatment effects on reading comprehension outcomes (Connor et al., 2017; Cromley & Azevedo, 2007; Kim et al., 2021; Language and Reading Research Consortium et al., 2019), and even fewer studies to our knowledge that have interviewed teachers to learn more about classroom vocabulary practices. This study sheds light on a critical set of vocabulary-related mechanisms through which content literacy instruction appears to improve young children's reading comprehension. Domain-specific vocabulary instruction that focuses on teaching words in semantic networks is a critical element of content literacy instruction as are the numerous word-learning practices that teachers employ regularly.

References

Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2007). Lasting consequences of the summer learning gap. *American Sociological Review*, *72*(2), 167–180. https://doi.org/10.1177/000312240707200202

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

American Institutes for Research. (2017–2018). *Results from a three-year i3 impact evaluation of the Children's Literacy Initiative (CLI): Implementation and impact findings of an intensive professional development and coaching program*. Children's Literacy Initiative. https://cli.org/impact/i3-scale-up/

Anderson, R. C. (2013). Role of the reader's schema in comprehension, learning, and memory. In D. E. Alvermann, N. J. Unrau, & R. B. Ruddell (Eds.), *Theoretical models and processes of reading* (6th ed., pp. 476–488). International Reading Association.

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

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

Apthorp, H., Randel, B., Cherasaro, T., Clark, T., McKeown, M., & Beck, I. (2012). Effects of a supplemental vocabulary program on word knowledge and passage comprehension. *Journal of Research on Educational Effectiveness*, *5*(2), 160–188. https://doi.org/10.1080/19345747.2012.660240

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

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological Bulletin*, *128*(4), 612–637. https://doi.org/10.1037//0033-2909.128.4.612

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

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school* (Expanded ed.). National Academy Press.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Cabell, S. Q., & Hwang, H. (2020). Building content knowledge to boost comprehension in the primary grades. *Reading Research Quarterly*, *55*(S1), S99–S107. https://doi.org/10.1002/rrq.338

Chin, T., & Phillips, M. (2004). Social reproduction and child-rearing practices: Social class, children's agency, and the summer activity gap. *Sociology of Education*, *77*(3), 185–210. https://doi.org/10.1177/003804070407700301

Clarke, P. J., Snowling, M. J., Truelove, E., & Hulme, C. (2010). Ameliorating children's reading-comprehension difficulties: A randomized controlled trial. *Psychological Science*, *21*(8), 1106–1116. https://doi.org/10.1177/0956797610375449

Coburn, C. (2003). Rethinking scale: Moving beyond numbers to deep and lasting change. *Educational Researcher*, *32*(6), 3–12. https://doi.org/10.3102/0013189X032006003

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *8*(2), 240–247. https://doi.org/10.1016/S0022-5371(69)80069-1

Connor, C. M., Dombek, J., Crowe, E. C., Spencer, M., Tighe, E. L., Coffinger, S., Zargar, E., Wood, T., & 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. https://doi.org/10.1037/edu0000128

Connor, C. M., Phillips, B. M., Kim, Y. G., Lonigan, C. J., Kaschak, M. P., Crowe, E., Dombek, J., & Al Otaiba, S. (2018). Examining the efficacy of targeted component interventions on language and literacy for third and fourth graders who are at risk of comprehension difficulties. *Scientific Studies of Reading*, *22*(6), 462–484. https://doi.org/10.1080/10888438.2018.1481409

Cromley, J. G., & Azevedo, R. (2007). Testing and refining the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology*, *99*(2), 311–325. https://doi.org/10.1037/0022-0663.99.2.311

Doty, E., Kane, T. J., Patterson, T., & Staiger, D. O. (2022). *What do changes in state test scores imply for later life outcomes?* NBER Working Paper 30701. https://www.nber.org/papers/w30701

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, *82*(1), 405–432. https://doi.org/10.1111/j.1467-8624.2010.01564.x

Elleman, A. M., 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), 1–44. https://doi.org/10.1080/19345740802539200

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*(2), 211–245. https://doi.org/10.1037/0033-295X.102.2.211

Firth, R. (1957). *A synopsis of linguistic theory. Studies in linguistic analysis*. Oxford.

Fitzgerald, J., Elmore, J., Relyea, J. E., & Stenner, A. J. (2020). Domain-specific academic vocabulary network development in elementary grades core disciplinary textbooks. *Journal of Educational Psychology*, *112*(5), 855–879. https://doi.org/10.1037/edu0000386

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

Goldman, S. R., Varma, S., & Cotè, N. (1996). Extending capacity constrained construction integration: Toward "smarter" and flexible models of text comprehension. In B. K. Britton & A. C. Graesser (Eds.), *Models of text comprehension* (pp. 73–113). Erlbaum.

Hadley, E. B., Dickinson, D. K., Hirsh-Pasek, K., & Golinkoff, R. M. (2019). Building semantic networks: The impact of a vocabulary intervention on preschoolers' depth of word knowledge. *Reading Research Quarterly*, *54*(1), 41–61. https://doi.org/10.1002/rrq.225

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Language and Reading Research Consortium, Jiang, H., & Logan, J. (2019). Improving reading comprehension in the primary grades: Mediated effects of a language-focused classroom intervention. *Journal of Speech, Language, and Hearing Research*, *62*(8), 2812–2828. https://doi.org/10.1044/2019_JSLHR-L-19-0015

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

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

Kim, J. S., Burkhauser, M. A., Mesite, L. M., Asher, C. A., Relyea, J. E., Fitzgerald, J., & Elmore, J. (2020). Improving reading comprehension, science domain knowledge, and reading engagement through a first-grade content literacy intervention. *Journal of Educational Psychology*, *113*(1), 3–26. https://doi.org/10.1037/edu0000465

Kim, J. S., Burkhauser, M. A., Relyea, J. E., Gilbert, J. B., Scherer, E., Fitzgerald, J., Mosher, D., & McIntyre, J. (2022). A longitudinal randomized trial of a sustained content literacy intervention from first to second grade: Transfer effects on students' reading comprehension. *Journal of Educational Psychology*, *115*(1), 73–98. https://doi.org/10.1037/edu0000751

Kim, J. S., Relyea, J. E., Burkhauser, M. A., Scherer, E., & Rich, P. (2021). Improving elementary grade students' science and social studies vocabulary knowledge depth, reading comprehension, and argumentative writing: A conceptual replication. *Educational Psychology Review*, *33*(4), 1935–1964. https://doi.org/10.1007/s10648-021-09609-6

Kimball, D. R., & Holyoak, K. J. (2000). Transfer and expertise. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 109–122). Oxford University Press.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, *95*(2), 163–182. https://doi.org/10.1037/0033-295X.95.2.163

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

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

Kintsch, W. (2013). The construction-integration model of text comprehension and its implications for instruction. In D. E. Alvermann, N. J. Unrau, & R. B. Ruddell (Eds.), *Theoretical models and processes of reading* (6th ed., pp. 807–839). International Reading Association.

Kintsch, W., & Kintsch, E. (2005). Comprehension. In S. G. Paris & S. A. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 71–92). Routledge.

Kline, R. B. (2016). *Principles and practice of structural equation modeling: Fourth edition* (4th ed.). The Guilford Press. https://ezp-prod1.hul.harvard.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=1078917&site=ehost-live&scope=site

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, *83*(404), 1198–1202. https://doi.org/10.1080/01621459.1988.10478722

Language and Reading Research Consortium, Lo, M.-T., & Xu, M. (2022). Impacts of the let's know! Curriculum on the language and comprehension-related skills of prekindergarten and kindergarten children. *Journal of Educational Psychology*, *114*(6), 1205–1224. https://doi.org/10.1037/edu0000744

Masten, A. S., & Cicchetti, D. (2010). Developmental cascades. *Development and Psychopathology*, *22*(3), 491–495. https://doi.org/10.1017/S0954579410000222

McKeown, M. G., Beck, I. L., Omanson, R. C., & Perfetti, C. A. (1983). The effects of long-term vocabulary instruction on reading comprehension: A replication. *Journal of Reading Behavior*, *15*(1), 3–18. https://doi.org/10.1080/10862968309547474

Myers, J. L., & O'Brien, E. J. (1998). Accessing the discourse representation during reading. *Discourse Processes*, *26*(2–3), 131–157. https://doi.org/10.1080/01638539809545042

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

Nesbit, J. C., & Adesope, O. O. (2006). Learning with concept and knowledge maps: A meta-analysis. *Review of Educational Research*, *76*(3), 413–448. https://doi.org/10.3102/00346543076003413

Neugebauer, S., Coyne, M., McCoach, B., & Ware, S. (2017). Teaching beyond the intervention: The contribution of teacher language extensions to vocabulary learning in urban kindergarten classrooms. *Reading and Writing*, *30*(3), 543–567. https://doi.org/10.1007/s11145-016-9689-x

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*(3), 249–272. https://doi.org/10.1598/RRQ.46.3.3

Newmann, F. M., Smith, B., Allensworth, E., & Bryk, A. S. (2001). Instructional program coherence: What it is and why it should guide school improvement policy. *Educational Evaluation and Policy Analysis*, *23*(4), 297–321. https://doi.org/10.3102/01623737023004297

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

Novak, J. D. (1990). Concept mapping: A useful tool for science education. *Journal of Research in Science Teaching*, *27*(10), 937–949. https://doi.org/10.1002/tea.3660271003

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

Pearson, P. D., Palinscar, A., Biancarosa, G., & Berman, A. (Eds.). (2020). *Reaping the rewards of the reading for understanding initiative*. National Academy of Education. https://doi.org/10.31094/2020/2

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, *11*(4), 357–383. https://doi.org/10.1080/10888430701530730

Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading*, *18*(1), 22–37. https://doi.org/10.1080/10888438.2013.827687

Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods*, *15*(3), 209–233. https://doi.org/10.1037/a0020141

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

Read, J. (2004). Plumbing the depths: How should the construct of vocabulary knowledge be defined. In P. Bogaards & B. Laufer-Dvorkin (Eds.), *Vocabulary in a second language* (pp. 209–227). John Benjamins Publishing Company.

Ryan, G. W., & Bernard, H. R. (2003). Techniques to identify themes. *Field Methods*, *15*(1), 85–109. https://doi.org/10.1177/1525822X02239569

Saldaña, J. (2016). *The coding manual for qualitative researchers*. Sage.

Sanetti, L. M. H., & Kratochwill, T. R. (2009). Toward developing a science of treatment integrity: Introduction to the special series. *School Psychology Review; Bethesda*, *38*(4), 445–459.

Satorra, A., & Bentler, P. M. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika*, *75*(2), 243–248. https://doi.org/10.1007/s11336-009-9135-y

Snow, C. E. (2015). 2014 Wallace Foundation Distinguished Lecture: Rigor and realism: Doing educational science in the real world. *Educational Researcher*, *44*(9), 460–466. https://doi.org/10.3102/0013189X15619166

Stahl, S. A., & Fairbanks, M. M. (1986). The effects of vocabulary instruction: A model-based meta-analysis. *Review of Educational Research*, *56*(1), 72–110. https://doi.org/10.2307/1170287

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2019 & 2022 Reading Assessments & Achievement Levels. https://www.nationsreportcard.gov/ndecore/xplore/NDE

Vaughn, S., Roberts, G., Swanson, E. A., Wanzek, J., Fall, A.-M., & Stillman-Spisak, S. J. (2015). Improving middle-school students' knowledge and comprehension in social studies: A replication. *Educational Psychology Review*, *27*(1), 31–50. https://doi.org/10.1007/s10648-014-9274-2

Weisner, T. S., Coots, J. J., Bernheimer, L. P., & Arzubiaga, A. (1997). *The ecocultural family interview manual*. UCLA Center for Culture and Health.

What Works Clearinghouse. (2021). *WWC Reporting guide for study authors*. Institute of Education Sciences. https://ies.ed.gov/ncee/wwc/Docs/ReferenceResources/WWC_Author_Guide_Jul2021.pdf

Wright, T. S., & Cervetti, G. N. (2017). A systematic review of the research on vocabulary instruction that impacts text comprehension. *Reading Research Quarterly*, *52*(2), 203–226. https://doi.org/10.1002/rrq.163

Zhao, X., Lynch, J. G., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of Consumer Research*, *37*(2), 197–206. https://doi.org/10.1086/651257

Received January 12, 2023
Revision received December 20, 2023
Accepted February 20, 2024 ■