ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 23-847

Improving elementary school students' reading comprehension through content-rich literacy curriculum: The effect of structured read-aloud supplements on measures of reading comprehension transfer

Douglas M. Mosher
Harvard University

James S. Kim
Harvard University

This study contributes to the science of teaching reading by illustrating how a ubiquitous classroom practice – read alouds – can be enhanced by fostering teacher language practices that support students' ability to read for understanding. This experimental study examines whether and to what extent providing structured teacher read aloud supplements in a social studies read aloud can allow students to leverage a familiar science schema and thereby positively impact reading comprehension outcomes. Treatment students received a single social studies read-aloud on the story of Apollo 11 with structured teacher read aloud supplements while control students received the same read-aloud story but without structured supplements. Effect sizes from hierarchical linear models indicated that students in the treatment condition significantly outperformed students in the control condition on four measures of domain-specific reading comprehension. Further exploratory analyses using structural equation modeling examined the extent that teacher language mediated the treatment effect. Results indicated that teachers going above and beyond the intervention script explained 67 percent of the treatment effect. Structured supplements for read alouds can help students see important connections between schemas, which ultimately aids in reading comprehension.

VERSION: September 2023

Suggested citation: Hartney, Mosher, Douglas M., and James S. Kim. (2023). Improving elementary school students' reading comprehension through content-rich literacy curriculum: The effect of structured read-aloud supplements on measures of reading comprehension transfer. (EdWorkingPaper: 23-847). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/w1h1-5g88

Improving elementary school students' reading comprehension through content-rich literacy curriculum: The effect of structured read-aloud supplements on measures of reading comprehension transfer

Douglas M. Mosher

James S. Kim

Harvard University

Abstract

This study contributes to the science of teaching reading by illustrating how a ubiquitous classroom practice – read alouds – can be enhanced by fostering teacher language practices that support students' ability to read for understanding. This experimental study examines whether and to what extent providing *structured teacher read aloud supplements* in a social studies read aloud can allow students to leverage a familiar science schema and thereby positively impact reading comprehension outcomes. Treatment students received a single social studies read-aloud on the story of Apollo 11 with structured teacher read aloud supplements while control students received the same read-aloud story but without structured supplements. Effect sizes from hierarchical linear models indicated that students in the treatment condition significantly outperformed students in the control condition on four measures of domain-specific reading comprehension. Further exploratory analyses using structural equation modeling examined the extent that teacher language mediated the treatment effect. Results indicated that teachers going above and beyond the intervention script explained 67 percent of the treatment effect. Structured supplements for read alouds can help students see important connections between schemas, which ultimately aids in reading comprehension.

Introduction

Attempting to improve reading comprehension outcomes is no small task, yet incremental research that bends the knowledge-seeking arc and provides teachers with practical and actionable practices is exactly what is needed to improve literacy outcomes for children. Indeed, students in the United States continue to struggle comprehending grade-level texts, as evidenced by the stagnant growth among fourth grade reading achievement for a decade prior to the COVID-19 pandemic (National Center for Education Statistics, 2019). With the Science of Reading in the forefront of public discourse, there is a timely need to highlight ways in which teachers can positively impact students' ability to comprehend texts. This study contributes to the science of teaching reading by illustrating how a ubiquitous classroom practice – read alouds – can be enhanced by fostering teacher language practices that support students' ability to read for understanding.

Thus, our aim was to examine the causal effects of embedding a number of *structured supplements* into a third grade read aloud lesson on various measures of reading comprehension through random assignment to treatment and control conditions. Structured supplements include concise definitions and meaningful examples of target vocabulary, discussion questions to engage students in using the target vocabulary to discuss larger topics related to the topic schema, a one-to-one schema mapping activity, and a review of key vocabulary words using a concept map. While there is ample research documenting the importance of the numerous components of the Science of Reading – developing phonological awareness, explicit phonics instruction, activating background knowledge, and building vocabulary knowledge – there is a need to better understand how teachers can use read aloud lessons to help students leverage, connect, and build the necessary knowledge to instantiate schemas efficiently in order to improve reading comprehension outcomes (Shanahan, 2020). Put differently, understanding how teachers can efficiently integrate background knowledge (i.e., schemas) while introducing new content to students is an area of research that needs further investigation.

Read Alouds Can Impact Vocabulary Knowledge and Comprehension Outcomes

There is compelling evidence to indicate that read alouds positively impact vocabulary acquisition among children ages 4-10 (Ard & Beverly, 2004; Biemiller & Boote, 2006; Elley, 1989; Robbins & Ehri, 1994; McKeown & Beck, 2014; Parsons & Bryant, 2016; Penno et al., 2002; Senechal & Cornell, 1993; Silverman, 2007), yet there is less evidence to show that they directly improve comprehension to the same extent (Coyne et al., 2010; Baker et al., 2020; McKeown & Beck, 2014). Specifically, in some studies, read alouds that included key vocabulary covered in the lesson had positive effects on listening comprehension measures (Coyne et al., 2010), while other studies did not detect meaningful effects on comprehension outcomes (Baker et al., 2020; McKeown & Beck, 2014). For studies measuring vocabulary acquisition, read aloud lessons that identified key vocabulary within texts and provided concise definitions of word meanings yielded positive results on vocabulary acquisition, especially for students with limited vocabulary knowledge (Coyne et al., 2004; Silverman, 2007). Students need rich exposures to word meanings and read alouds expose students to words both incidentally (Authors, 2021; 2022; Wright, 2022) and explicitly with word explanations, meaningful examples, structured discussions using vocabulary, and closer examination of the phonological and orthographic aspects of words (Silverman, 2007).

There is emerging evidence that the quantity and quality of teachers' language during a read aloud is an important pathway for improving their students' comprehension ability. Read alouds directly improve vocabulary acquisition, and although they are less effective in directly raising comprehension scores, there is reason to believe that they may indirectly improve comprehension outcomes. Engaging students in scaffolded and inferential discussions is an important element of read aloud lessons that has been shown to be positively associated with comprehension scores (van Kleeck, 2006; Collins, 2016). In order for students to comprehend texts and move beyond literal recall questions, teachers must create opportunities for rich, high-level discussions. A meta-analysis examining the effects of storybook interventions on vocabulary and comprehension outcomes showed that a range of effects from negative to positive (Swanson et al., 2011). Dialogic reading – read alouds including conversational turn-taking and open-ended questions (Ezell & Justice, 2005; Swanson et al., 2011) – were among the most common type of read aloud lesson and yielded an overall positive effect. Indeed, shifting away from students solely listening to a read aloud and answering occasional low-level questions to engaging students verbally and making the read aloud experience more interactive is essential to helping students acquire new knowledge. One way to facilitate discussions among students is to embed quality discussion prompts into the actual read aloud lesson. In fact, scripting lessons has been shown to lead to higher-quality teaching (Language and Reading Research Consortium et al., 2014). However, most read aloud interventions have involved multiple lessons over numerous days of instruction. There has yet to be a study examining the impact of a single read aloud lesson designed to build student vocabulary and domain knowledge, facilitate schema instantiation, and impact reading comprehension outcomes through the inclusion of targeted scripting for teachers.

Schemas are the Intellectual Structures that Help Student Organize Knowledge

Schemas are supportive intellectual structures that allow students to acquire, access, and retain relevant information (Anderson & Pearson, 1984). They are also essential to comprehending texts (Anderson, 2010). To comprehend a text, an individual needs to have "found a mental 'home' for the information in the text" or make the necessary adjustments to integrate any new information with that which is already established (Anderson & Pearson, 1984, p. 225). Thus, schemas are malleable, constantly evolving, and hierarchical (Kimball & Holyoak, 2000).

As students learn about different topics, they develop *topic schemas*. For example, as they learn about the various human body systems – muscular, skeletal, and nervous – students form a topic schema on how human body systems function to stay alive. If students continue to learn about, for example, different animal systems or the moon team system, they will develop corresponding topic schemas. Each of the three schemas just mentioned share a common overarching *general schema* that centers around the concept of 'systems'.

Indeed, schemas are an essential component to comprehending texts. In the construction-integration model, readers must build a robust situation model that merges a reader's literal understanding of a text with that reader's existing background knowledge (Kintsch, 1993; 2009). First, all learners must formulate propositions by accessing word meanings, utilizing working memory, and assessing the coherence of the various propositions (Kintsch & Kintsch, 2005). During the integration phase, however, readers must incorporate the literal propositions from the text with their existing knowledge and experiences – i.e., robust schemas – to form a mental model of the text and its meaning. Read alouds can function as a vehicle for teachers to scaffold and aid students in building and instantiating schemas so that as students become skilled readers, they can do so independently when engaging with various texts. Yet, minimal research has explored ways to help teachers improve this practice. If schemas are abstract, intellectual structures that are essential to reading comprehension, what are the observable and concrete components of schemas?

Networks of Domain-Specific Vocabulary Words are the Visible Parts of Schemas

Networks of domain-specific vocabulary knowledge represent the building blocks of schemas and are the visible components of a schema – "the exposed tip of the conceptual iceberg" (Anderson & Freebody, 1981, p. 82). Yet, there has to be a unifying thread connecting vocabulary words. Without one, words are simply disconnected. There is growing evidence that teaching domain-specific vocabulary in semantic networks has been shown to impact proximal measures of vocabulary knowledge (Kim et al., 2020; 2021; 2023a) and that networks of domain-specific vocabulary mediated the treatment effect documented in the 2022 Model of Reading Engagement (MORE) study (Authors, 2023b). Creating rich semantic representations allows students to store these word meanings in their long-term memory (Ericsson & Kintsch, 1995), thereby allowing them to retrieve word meanings much more efficiently and consequently aiding in understanding various texts (Fitzgerald et al., 2020). Furthermore, creating "an extended network of meanings for a given focal word represents instantiation of stronger and richer meaning" (p. Fitzgerald et al., 2020, p.857), and visualizing networks of vocabulary with concept maps can help students see the connections between words and concepts (Karpicke & Blunt, 2011; Novak, 1990). Thus, as students instantiate word meanings within a network of words and concepts, students are also developing a topic schema.

By building networks of domain-specific vocabulary knowledge, students are also acquiring the necessary domain and topic knowledge in disciplines such as social studies and science. These networks of word knowledge signify the various interconnections between concepts within a student's memory (Kendeou & O'Brien 2015). In the present study, we rely on an already established topic schema (human body system) to provide context for students to build a new topic schema (moon team system), both of which are nested within a general schema (systems). Figure 1 shows how the domain-specific vocabulary words serve as the kernels of domain and topic knowledge that comprise each schema. It also helps illustrate the conceptual similarities between both schemas via the overlapping vocabulary words.

Indeed, building networks of vocabulary knowledge is tightly coupled with acquiring domain and topic knowledge, both of which help to instantiate schemas and ultimately aid in linguistic and reading comprehension (Cabell & Hwang, 2020; Cervetti & Wright, 2020). Vocabulary knowledge and topic knowledge have been shown to be among the strongest predictors of reading comprehension (Cromley & Azevedo, 2007), and in a more recent study, vocabulary and topic knowledge were both found to explain the most variance in reading comprehension (Ahmed et al., 2016). Recent content literacy interventions that have focused on developing robust schemas through the acquisition of domain-specific vocabulary knowledge yielded positive treatment effects on vocabulary knowledge (Kim et al., 2020; 2021; 2023a) and reading comprehension measures (Kim et al., 2020; 2022a; 2023a). Having robust schemas are an essential element of comprehending nonfiction texts, and read alouds have the ability to make schemas more concrete by engaging students with rich texts designed to build knowledge.

Robust Schemas Allow for Students to Transfer Knowledge Across Topics and Domains

When students have developed robust schemas, they are able to "transfer knowledge from one situation to another by a process of *mapping*" (Gick & Holyoak, 1983, p. 2). More specifically, establishing a set of one-to-one correspondences that map from one concept to another – i.e., one topic schema mapping onto another schema – help students ultimately transfer and apply knowledge when they encounter an unfamiliar topic. This process allows students to leverage their existing knowledge organized in schemas as students encounter knew situations that vary along a continuum from familiar to unfamiliar (Kimball & Holyoak, 2000;Authors, 2023a). To move past lower-level comprehension tasks such as recalling facts and to progress towards a deeper understanding of knowledge and how it applies to other concepts, acquired knowledge "must be actively linked to semantic retrieval cues" to make acquiring knowledge from texts possible in unfamiliar situations (Kintsch, 2013, p. 812). Put differently, words that convey conceptual meaning that are organized in networks are essential to reading and comprehending informational texts. In sum, there is ample evidence to support how to help students comprehend texts, yet in practice, this is not always a reality.

Instruction Rarely Activates and Builds Student Knowledge

Numerous studies highlight the importance of activating and building students' background knowledge (McCarthy et al., 2018; O'Reilly et al., 2019; Kaefer, 2020), and there is evidence to indicate that activating student knowledge is not a common occurrence in some elementary school classrooms (Hattan et al., 2015). Instructional materials often fail to consistently prompt teachers to activate and build existing knowledge, thereby placing students who do not possess adequate knowledge at a disadvantage. Similarly, teachers who launch pre-reading activities that introduce new content rather than reviewing already established content to provide context for the subsequent read aloud also establish barriers for students who do not possess the necessary knowledge to access texts (Hattan et al., 2015; Kaefer, 2020). Indeed, the language teachers use in the classroom is a key factor for introducing and explaining new content. How can teachers effectively activate and build knowledge in the classroom?

Content Literacy Instruction Creates Opportunities to Build Knowledge

Research spanning the past decade has documented the positive effect of content literacy instruction on knowledge building and comprehension outcomes for elementary school children. In fact, a recent meta-analysis revealed that content-area instruction improved elementary school standardized comprehension by a quarter of a standard deviation (Hwang et al., 2021). Even more recent, a content-literacy intervention called the MORE intervention produced lasting effects on standardized reading comprehension outcomes (Kim et al., 2023a). Participating students received two 5-week units in first grade, one 5-week unit in second grade, and one 3-week unit in third grade. Results showed positive and significant effects on the standardized end-of-grade (EOG) assessment in third grade, and even in the absence of lessons in fourth grade, students who had received MORE the previous three years continued to outperform those who did not on a standardized reading assessment.

Yet, even as research suggests the positive effects of content-area literacy, domains such as social studies often are overlooked due to the emphasis placed on English Language Arts (ELA) and Math instruction (Huck, 2019). In some instances, teachers have perceived that the core priorities of schools and districts is to focus on English Language Arts and Math instruction, often at the expense of other subjects. Consequently, content-area domains such as social studies have been neglected even as recent research indicates the significant positive association between time spent on social studies content and rising literacy scores (Northern & Petrilli, 2020).

Content literacy provides opportunities for students to build their knowledge and vocabulary simultaneously, but there is still a need to better understand effective ways to acquire and integrate new knowledge with existing knowledge. In a recent mixed-methods study, interviewed teachers who had participated in the MORE content literacy intervention described how they provided concise explanations of domain-specific vocabulary word meanings, examples for each target word during instruction, and numerous ways in which target words were conceptually connected to other studied words (Mosher et al., 2023b). Furthermore, domain-specific vocabulary knowledge was found to be a significant mediator of the treatment effect. Indeed, teacher language is a crucial element of acquiring knowledge, and in intervention studies, teachers going above and beyond the lesson script to explain word meanings can make knowledge acquisition that much more impactful (Neugebauer et al., 2017). Thus, if building knowledge is an essential criterion for comprehending grade-level informational texts, then the quality of teacher language is a crucial element for facilitating learning in the classroom.

Teacher Language as a Potential Mechanism Through Which the MORE Intervention Operates

Previous studies of MORE have examined the extent to which domain-specific vocabulary knowledge mediates measures of reading comprehension transfer (Mosher et al., 2023b) as well as how domain-specific reading comprehension mediates argumentative writing outcomes for multilingual learners (Relyea et al., 2022b). Nevertheless, there has yet to be a study examining the extent that the quantity of multiple aspects of teacher language mediates treatment effects for content literacy interventions such as MORE. While there is evidence that student domain-specific vocabulary knowledge is a key mechanism of the MORE intervention, it seems plausible that the language teachers use when discussing vocabulary or content related to vocabulary would be an essential part of student domain-specific vocabulary acquisition and schema instantiation. Put differently, if a lesson plan for a read aloud provided concise vocabulary definitions and examples, could those embedded explanations of word meanings serve as a catalyst for more elaboration of word meanings and examples – i.e., language extensions (Neugebauer et al., 2017)? Similarly, would embedding discussion questions designed to promote use of target vocabulary words encourage teachers to use those target words with greater frequency and ask additional follow-up questions? Thus, the inclusion of embedded supports (structured supplements) around key language in a read aloud lesson could impact teachers so that they elaborate on lesson materials, which in turn could predict comprehension outcomes.

The Present Study

The primary goal of the study was to investigate certain malleable factors within the classroom that have the potential to help students acquire and transfer knowledge on a formative assessment of reading comprehension transfer. We sought to test the effectiveness of certain instructional strategies and the impact in helping students better read grade-level texts.

The research questions were as follows:

(1) Compared to a read aloud lesson without any structured supplements, to what extent does embedding structured supplements into a read aloud lesson improve: (a) third graders' recall of basic topics discussed in the text, (b) their domain-specific reading comprehension as measured by near- and mid-transfer assessments, and (c) their domain-general reading comprehension as measured by an End-of-Grade (EOG) reading comprehension assessment?

(2) To what extent does teacher language mediate effects of the read aloud lesson on student comprehension scores (recall, near, and mid transfer)?

To address these questions, we conducted a cluster randomized controlled trial where classrooms were randomly assigned to either the treatment (read aloud with structured supplements) or control conditions (read aloud without structured supplements). As shown in the theory of change in Figure 2, we hypothesized that students receiving structured read aloud supplements in the treatment condition would have higher comprehension scores than those in the control condition without structured supplements. That is, the structured supplements would provide ample opportunities for students to gain access to word meanings (and by proxy domain and topic knowledge), engage students in conversations using target vocabulary, and establish connections between two subjects often taught in isolation: science and social studies. Furthermore, we hypothesized that structured supplements would improve the quantity of teacher language – language not in the read aloud lesson plan – and that that greater quantity of teacher language would mediate the proposed treatment effect. The study was preregistered at the Registry of Efficacy and Effectiveness Studies and the lesson was taught in May of 2022.

Method

Design

The experimental design for this study blocked on schools with random assignment occurring at the classroom level. Randomization was achieved using Stata 17. Balance tests at the student level showed no significant differences between treatment and control groups in standardized measures of baseline reading and math or in baseline student demographic characteristics. Balance tests at the teacher level also showed no significant differences in years teaching or past experience teaching MORE.

Transparency Statement

We preregistered the study at the Registry of Efficacy and Effectiveness Studies prior to lesson implementation. The data and analysis will be made publicly available.

Participants

We recruited teachers from the larger 2022 MORE science intervention that had desirable compliance rates – i.e., teachers who implemented most MORE lessons (Figure 3). In total, 82 teachers agreed to participate, with a prospective total of 1,467 third-grade students from 20 elementary schools. In preparation for this initial efficacy trial, we piloted the read aloud lesson in multiple third-grade classrooms resulting in adjustments to the story and activities prior to implementation.

The final analytic sample included 965 students from 19 schools with an attrition rate of 34 percent. There were no significant differences on baseline on reading and math tests nor was there a significant differential attrition rate between treatment and control conditions. In short, no major threats to the internal validity of the study were observed.

Procedure

In the spring of 2022, all third-grade students within a large, urban district in the Southeastern United States participated in the MORE science intervention for three weeks where they learned about three important human body systems: the muscular, skeletal, and nervous systems. Students also learned key domain-specific vocabulary words that helped build their domain and topic knowledge about human body systems. As lessons progressed, students constructed a schema around how these different systems work together to help the body function. Thus, we hypothesized that if students have developed a robust topic schema from science (human body systems), they should be able to construct a new topic schema in the domain of social studies (moon team system) and establish a link between both schemas (Anderson & Pearson, 1984; Gick & Holyoak, 1983). Accessing an instantiated schema provides students with the necessary background knowledge to make connections to a new topic that is conceptually related. Yet, what scaffolds are needed to help make this link visible to students?

To test the impact of a read aloud lesson with structured supplements on comprehension outcomes, we wrote a nonfiction text on Apollo 11. The text focuses on how a team of individuals – Katherine Johnson, Chuck Lowry, Neil Armstrong, and John F. Kennedy – worked together as a system to overcome obstacles and accomplish their goal of sending astronauts to the moon and returning them safely to Earth. A recent intervention indicated that high school students who read stories about numerous struggles that scientists had to overcome improved student learning (Lin-Siegler et al., 2016). Thus, we wrote our own struggle story about the moon team's challenges and how they overcame them. Treatment teachers read the read aloud text with structured supplements. The various structured supplements are shown in Table 1 and included vocabulary instruction and discussion activities designed to help students build their vocabulary, domain, and topic knowledge while drawing on the schema developed during the MORE science intervention. Structured supplements also included schema mapping and concept mapping activities. Control teachers received the Apollo 11 read aloud story but without any of the structured supplements designed to identify and explain key vocabulary words, pose discussion questions, or help students link the science and social studies schemas. In essence, the read aloud text that control teachers received was a blank slate that gave full autonomy to those teachers to decide which, if any, words to introduce and what questions to ask. Students in both conditions completed identical assessments. Thus, all students were exposed to the social studies read aloud and random assignment to the treatment condition allows for an estimate of the effectiveness of structured supplements on student comprehension outcomes.

The read aloud text included three vocabulary words from the MORE science intervention students had already learned: *function*, *system*, and *diagnosis*. The text also introduced four new social studies words: *contribute*, *engineer*, *persist*, and *ingenious*. The treatment version of the text provides embedded explanations of these four social studies words along with relevant examples and specific questions prompting students to access the science schema and make connections between the science schema and the read aloud content. For example, after learning about the different members of the team that helped send astronauts to the moon, teachers asked students, "How is the moon team *system* similar to the different human body *systems*?" Students subsequently completed a schema-mapping activity designed to help students see the analogic link between science and social studies schemas, thereby engaging students in a 1:1 mapping of both topic schemas (Gick & Holyoak, 2000). In this activity, students are asked about the similarities between the moon team system and the human body system. Teachers then led students through a "madlibs" activity (Figure 4) where students had to complete the missing words scattered across two paragraphs using key vocabulary words from both the previously taught science unit (paragraph 1) and the read aloud lesson (paragraph 2). Both paragraphs had similar language and the main differences centered around which target words completed the paragraph. The activity aimed to make the 1:1 schema mapping explicit. Elsewhere in the lessons, students were posed with numerous questions to discuss with a partner. For example, "How was Chuck Lowry an *ingenious engineer*?" These types of prompts required students to use target vocabulary in their discussions via specific inclusion criteria (Table 1).

Measures

Pretests

Reading and Math pretest scores came from winter Measure of Academic Progress (MAP) scores (NWEA, 2011). The district uses the MAP assessment at three times points throughout the year to assess student achievement. The winter assessment date represents the mid-year assessment of student achievement and the most recent standardized measure of student reading and math ability prior to the start of the intervention lesson.

Posttests

Students took 4 assessments following the conclusion of the read aloud lesson. We developed three measures of domain specific reading comprehension and the fourth measures was the final state-wide EOG assessment, which was issued a month later.

Recall. To measure basic recall, students took a brief 5-item multiple choice assessment (with 4 options per item) measuring basic understanding of key details from the text. The measure had a Cronbach's reliability estimate of 0.63.

Near and Mid Transfer. Near and mid transfer domain specific reading comprehension measures each were comprised of 13 multiple choice items with four options per item. Recent research suggests measuring a continuum of transfer rather than simply assessing proximal and distal measures (Kim et al., 2022a). The near-transfer passage discussed the creation of the lunar module that enabled astronauts to explore the moon. While students had learned about the system that helped get astronauts to the moon, they had no prior knowledge of the lunar module. The mid-transfer passage focused on the making of the Empire State building, a topic much more removed from that of the read aloud content centered around the moon. Near- and mid-transfer assessments yielded a Cronbach's alpha reliability of 0.82 and 0.76 respectively. Results from a 2-paramenter logistic item response theory (IRT) model indicated location parameters ranging from -1 to 3 with almost all items providing substantial information.

Content Comprehension. Content comprehension represents both near- and mid-transfer measures combined.

End-of-Grade (EOG). Domain general reading comprehension was measured via the North Carolina third grade EOG assessment, which includes multiple-choice items (4 options) that are scaled using a 3-parameter IRT model.

Fidelity of Implementation

We gave each teacher two recorders to wear around their neck while delivering the read aloud lesson. Of the 80 teachers participating in the study, 75 teachers returned usable recorders. The unusable recorders were either broken, preventing us from accessing the recording, or the recorder was never switched on. Fidelity analyses involve the 75 teachers who returned usable recorders.

To assess fidelity of implementation, we used two facets of Dane and Schneider's (1998) facets of fidelity: adherence and dosage. We examined teacher adherence to the lesson by comparing the number of times a teacher said each target word within each condition. For example, the read aloud lesson with structured supplements used the word *contribute* 31 times while the lesson without structured supplements used *contribute* 10 times. Thus, when listening to the audio recordings, we would expect to hear treatment teachers say *contribute* at least 31 times and control teacher at least 10 times. We created dichotomous variables for each word to indicate if each of the seven target words were said the prescribed number of times. We also created a dichotomous variable for whether or not the number of scripted questions included in the lesson were asked. We then took row totals across the seven words and questions variables to create a composite with a maximum score of 8. Results from these comparisons are listed in Table 6 where percentage of adherence is also reported. There was no significant difference in adherence between treatment and control condition ($\beta = -.46, SE = .33, p = .16$).

To assess dosage, we report the length of audio measured in minutes for reach recording, as shown in Table 6. Unfortunately, two different teachers' recorders inadvertently shut off mid-lesson and their skewed adherence and dosage brings down the treatment average. Because the read aloud lesson with structured supplements had more text, we expected the lesson to be longer. Results indicate that on average, the treatment lesson was roughly 19 minutes longer ($t = 8.74, p < .001$).

Data Analyses

To assess the causal impact of structured read aloud supplements on five comprehension outcomes, we fit a series of 3-level hierarchical linear models (HLMs) with teacher random effects and school fixed effects (Raudenbush & Bryk, 2002). The model is as follows:

$$Y_{ijk} = \gamma_{00} + \alpha_k + \gamma_{01} TREATMENT_{jk} + \sum_{p=1}^{10} \gamma_{pjk}\, L1COV_{ijk} + \sum_{q=2}^{3} \gamma_{0qk}\, L2COV_{jk} + \varepsilon_{ijk} + u_{0jk}$$

$$\varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2) \quad u_{0jk} \sim N(0, \sigma_u^2)$$

where Y_{ijk} represents the 5 outcomes (recall, near transfer, mid transfer, content comprehension, and EOG) for student i in classroom j in school k. γ_{00} is the intercept for the school reference group and α_k represents each school's average deviation from the reference school intercept. γ_{01} is the adjusted causal effect of structured read aloud supplements, $\gamma_{pjk\,(p=1,\ldots,10)}$ is a vector of 10 student-level covariates including winter MAP pretests in reading and math as well as student demographic variables (race/ethnicity, gender, English Language Learner status, measure of neighborhood poverty), and $\gamma_{0qk\,(q=2,3)}$ is a vector of two teacher-level covariates (years of teaching experience, past experience with MORE) to improve the precision of the treatment effect estimates. ε_{ijk} represent the student-level residuals and u_{0jk} are the classroom-level random intercepts. Because we had five outcomes, we tested our results to the sensitivity of false discoveries using the Benjamini-Hochberg procedure with a false discovery rate set to 0.05 by outcome domain (Benjamini-Hochberg, 1995).

Some teachers administered the recall, near-, and mid-transfer assessments on different days, and as a result, we had varying numbers of students that completed each part of the assessment due to student absences. We used 2-parameter IRT-stored estimates for recall and near- and mid-transfer outcomes. For the EOG assessment, we used the district reported scale scores. Because of the discrepancy in student completion, we tested for a potential treatment effect on missing assessment data. That is, we wanted to determine if there was an imbalance in treatment and control missingness. HLMs revealed that there were no significant treatment/control differences on recall ($\beta = -.01, SE = .01, p = .23$) near ($\beta = -.05, SE = .05, p = .32$), or mid transfer ($\beta = -.03, SE = .06, p = .65$). Because we had missing pretest data, we conducted Little's missing completely-at-random (MCAR) test. Results revealed that missing values were MCAR ($\chi^2 = 1.74, df = 2, p = .42$). Subsequent HLMs use multiple imputation by simulating 20 data sets with potential values in place of the missing observations (StataCorp, 2021).

To address our second research question on the extent to which teacher language mediated the treatment effect, we used coded data from classroom recordings to specify a structural equation model (SEM) using *Mplus 7* (Kline, 2016; Muthén & Muthén, 2012). As our mediation model included latent variables for *Teacher Language* and *Reading Comprehension Transfer*, we specified two different confirmatory factor analysis (CFA) measurement models to verify each latent variable's properties. *Teacher Language* was comprised of nine indicator variables (Figure 5). Each target word (of which there were seven) was represented as an indicator variable that included the number of times the target words were uttered above and beyond the lesson script. We included an indicator variable for language extensions, which was the number of times a teacher provided additional explanations or examples of a word's meaning above and beyond the script. Language extensions were coded in Dedoose (2016) and 20 percent of the recorded audio transcripts were double coded with an inter-rater-reliability kappa estimate of $\kappa = .88$. The final indicator variable, teacher questions, included the number of times a teacher asked the class a question related to the text above and beyond the lesson script. The outcome factor, domain-specific reading comprehension, was comprised of three indicator variables: stored IRT estimates for recall, near transfer, and mid transfer questions (Figure 6). To account for the clustered nature of the data, we used the Maximum Likelihood Robust (MLR) estimator and Full Information Maximum Likelihood (FIML) to account for pretest missing data.

Results

Preliminary analyses

Descriptive statistics of the student analytic sample are shown in Table 2 and correlations for key variables for research question 1 are shown in Table 3. We conducted balance tests to determine if attrition resulted in differences between both baseline reading and math measures and found that there were no significant differences based on attrition. Descriptive statistics for the teacher analytic sample are shown in Table 3, and in Table 4 correlations are listed for both teacher-level covariates indicators comprising the latent variable *Teacher Language* for research question 2.

Both latent variables were assessed using CFA. All indicators were significantly correlated with the latent construct *Teacher Language* (Figure 5), with factor loadings for eight of the nine indicators above .70. One of the indicators had a weak correlation with the factor because the word (*diagnose*) only appeared in the text a few times (unlike the other six words), and consequently, there was minimal variation of teachers mentioning the word beyond the lesson script. Additionally, we allowed two different sets of indictors to covary as both sets of words were often mentioned within the same utterance as each other. For example, when the word *system* was mentioned, the word *function* was often used in the same phrase. The model fit for *Teacher Language* was deemed adequate (RMSEA = .020, CFI = .962, TLI = .945, SRMR = .049). For domain-specific reading comprehension (Figure 6), all three indicators were significantly correlated with the factor, with factor loadings above .60. There were no fit statistics for this factor given that the model was perfectly identified.

Research Question 1: Effect of Structured Read Aloud Supplements on Comprehension Outcomes

Table 7 displays the HLM results for recall, near transfer, mid transfer, content comprehension (near and mid transfer combined) and EOG. Model estimates are standardized and indicate that there were indeed significant treatment effects on recall ($\gamma_{01k} = .17, SE = .07, p < .05$), near transfer ($\gamma_{01k} = .17, SE = .06, p < .01$), mid transfer ($\gamma_{01k} = .18, SE = .07, p < .05$), and content comprehension ($\gamma_{01k} = .18, SE = .07, p < .05$) passages. Thus, there is evidence that structured read aloud supplements impacted domain-specific reading comprehension. Results indicated that the intervention did not improve student outcomes on the domain general EOG reading comprehension measure ($\gamma_{01k} = .01, SE = .02, p = .65$).

We tested for false discoveries according to the Benjamini-Hochberg procedure with a false discovery rate set to .05 and confirmed the significant treatment effects on recall, near-, and mid-transfer passages (see Appendix).

Research Question 2: Teacher Language as a Mediator for Comprehension Outcomes

SEM was used to determine if and to what extent teacher language mediated the treatment effect. Figure 7 shows the specified model. In our model, we included freely estimated paths between random assignment to structured read aloud supplements and students' domain-specific reading comprehension outcomes as well as the latent variable teacher language. The model fit to the data was adequate (RMSEA = .025, CFI = .914, TLI = .901, SRMR = .055). Results indicated that the language teachers use that goes above and beyond the intervention script explained 67% of the treatment effect of domain-specific reading comprehension with a significant total indirect effect ($b = .294, SE = .11, p < .01, \beta = .124$) and total effect ($b = .439, SE = .17, p < .05, \beta = .185$). Furthermore, random assignment to structured read aloud supplements significantly predicted an increase in the language teachers use exceeding the script ($b = 1.112, SE = .29, p < .001, \beta = .972$), which in turn significantly predicted domain-specific reading comprehension ($b = .264, SE = .09, p < .01, \beta = .127$).

Sensitivity Analyses

We conducted sensitivity analyses to confirm the robustness of our findings. For research question 1, we aggregated our data at the classroom level and fit a regression model with school fixed effects using the same covariates. Results were slightly higher, with effect sizes ranging from .21 to .25 (Table 8). We also used SEM with clustered standard errors at the classroom level to estimate the causal effect of structured read aloud supplements on comprehension outcomes. Consistent with the 2-level HLM, SEM revealed a significant treatment effect of .17, identical to the main findings (contact first author for details).

As a sensitivity check for determining the extent to which teacher language mediated the treatment effect, we employed the bootstrap method to estimate a population parameter to provide bias-corrected bootstrap 95% confidence intervals (CIs) for the indirect effect using

1,000 draws (MacKinnon et al., 2004; Preacher & Hayes, 2008). Results were nearly identical, although the standard errors were lower given that *Mplus* does not allow the BOOTSTRAP command and CLUSTER command to be run simultaneously (contact first author for details).

Discussion

The present study sought to examine the causal effects of structured supplements for teachers – embedded vocabulary explanations, targeted discussion questions, and a one-to-one schema mapping activity – in a read aloud lesson on five measures of reading comprehension for third-grade students: recall, near transfer, mid transfer, content comprehension (near/mid combined), and EOG. While all students had participated in three weeks of the MORE science intervention and had developed a robust schema centered on three human body systems, they had not learned about or been exposed to the Apollo 11 mission to the moon. We designed the Apollo 11 read aloud first to pose a question designed to help students access their already established science schema. The structured supplement embedded within the treatment read aloud asked students to share their responses to the questions and for the teacher to record those responses. This activity likely provided students the necessary time to retrieve relevant background knowledge (McCarthy et al., 2018; O'Reilly et al., 2019; Kaefer, 2020). With a familiar topic schema activated, teachers then read aloud the Apollo 11 story that focused on a group of people working together to send astronauts to the moon. Thus, we leveraged a familiar topic schema about the human body and mapped that schema onto the unfamiliar topic schema of the moon team (Anderson & Pearson, 1984; Gick & Holyoak, 1983). Embedded structured supplements in the read aloud lesson for treatment teachers positively impacted student outcomes with effect sizes ranging between .17 and .18 on four measures of reading comprehension: recall, near transfer, mid transfer, and content comprehension (near and mid transfer tests combined). What is the significance of these findings?

First, the language of the embedded structured supplements introduced each of the four social studies vocabulary words with explanations of word meanings and examples by making semantic connections to both the new social studies words and the previously taught science words. Because knowledge is often organized in clusters (Anderson & Freebody, 1981), helping teachers use explicit language to introduce new vocabulary and link those words to other key vocabulary likely helped students better acquire word meanings (Read, 2004). While students in the control condition were exposed to these words incidentally by listening to the read aloud, without the embedded structured supplements succinctly highlighting word meanings through explanations and examples linking words, students in the control condition were unable to efficiently acquire word meanings, which impacted their ability to comprehend the texts in the recall and transfer assessments.

Second, by providing targeted questions designed to reinforce word meanings and build a new social studies topic schema, teachers in the treatment condition created situations where students had to use target vocabulary in their discussions. Thus, students were pushed to apply their new knowledge, and through this application, begin to instantiate a new topic schema around the moon team system working together to send astronauts to the moon. Indeed, these domain-specific vocabulary words represented “the exposed tip of the conceptual iceberg” (Anderson & Freebody, 1981, p. 82). Our findings confirm that providing scripted questioning is not only associated with comprehension scores (van Kleeck, 2006), it can positively impact them.

Third, providing an explicit 1:1 schema mapping activity was essential to helping students see the connections between the human body system and the moon team system (Gick & Holyoak, 1983). In this activity, students were asked about the similarities between the moon team system and the human body system and tasked with using both science and social studies vocabulary to fill in the missing words in two paragraphs (Figure 4). This activity was key to helping students see the links between both topic schemas and likely helped students leverage these similarities when they read passages on the lunar module and the construction of the Empire state building. Our findings confirm that "expertise can sometimes be transferred to novel tasks within and beyond the initial domain" (Kimball & Holyoak, 2000), thereby confirming similar results from an earlier analysis of a different implementation of the MORE intervention (Authors, 2023c). The cumulative effect of structured supplements allowed students to leverage their networks of knowledge to construct a robust situation model that merged students' literal understanding of the texts with their existing background knowledge (science and social studies schemas). Consequently, students were able to leverage that knowledge when encountering the unfamiliar passages (Kintsch, 1993, 2009; Kintsch & Kintsch, 2005). One interesting finding is that effect sizes were consistent across recall, near-, and mid-transfer assessments, which suggests that students found the easiest task (recall) no easier than the more challenging near-transfer passage or the even more difficult mid-transfer passage. Past studies have shown a slight drop in effect size for mid-transfer compared to near-transfer results (Kim et al., 2022a). These findings suggest that the structured supplements may have reduced the difficulty of the mid-transfer passage. Put differently, a passage that qualified once as mid transfer became near transfer, likely due to the structured supplements.

In our theory of change (Figure 2), we identify teacher language as a potential mediator, and in our analyses, we found that teachers in the treatment conditions used target words in their language above and beyond language from the read aloud lesson significantly more than control teachers. Similarly, teachers in the treatment condition asked more questions and provided language extensions for word meanings above and beyond the script of the lesson. This suggests that providing structured supplements for teachers during a read aloud can serve as a catalyst for expanded opportunities for word learning and schema instantiation. Giving teachers targeted word explanations and questions may help direct their attention on key concepts, thereby allowing them to use their pedagogical skills to enhance structured supplements of a read aloud. Our results confirm that providing teachers with basic language embedded within a read aloud is a key mechanism in explaining the significant treatment effects and is consistent with Gick & Holyoak's (1983) study, which documented the role teacher language played in helping scaffold the mapping of one schema to another. In contrast to the control condition where teacher language focused on a variety of semantically unrelated words, the treatment teachers focused their activities on the domain-specific words, which provided retrieval cues that activated the schema in the reader's mind and constrained the semantic space of related words. Consistent with the lexical quality hypothesis (Perfetti, 2007), our findings suggest that teachers' language in the treatment afforded more meaningful contexts for students to learn the form and meaning of words in connected text.

We did not see an impact of structured read aloud supplements on EOG scores. However, it is unlikely that a single lesson would have such far reaching effects on such a distal measure of reading comprehension as a standardized EOG reading assessment.

Finally, it should be noted that students received only a single read aloud lesson, and that this lesson alone with embedded structured supplements for teachers caused improvements in recall and transfer passage outcomes. Including structured supplements in read aloud lessons is a promising practice that can aid teachers in identifying and targeting their instruction during read aloud lessons through accessing familiar schemas, building knowledge through domain-specific vocabulary instruction, and creating an opportunity to link two similar schemas.

Limitations and future directions

While the current findings are promising, there are a number of limitations. First, while results indicated that structured supplements impacted recall and transfer measures of comprehension, replications are needed to confirm that these results hold. Second, the recall and transfer assessments were designed to be given the day after the read aloud lessons were completed. Teachers were told to teach the lesson and have students complete all assessment activities within a 10-day window. Due to time constraints at schools, many teachers administered the assessments multiple days after the completion of the read aloud and sometimes even spread across multiple days. This may have impacted the magnitude of the treatment effect as it is unknown if waiting multiple days could have impacted student outcomes. Furthermore, when teachers administered the assessments across multiple days, there were often student absences that resulted in incomplete assessments. For example, some students may have taken the recall and near-transfer assessment but were absent for the mid-transfer assessment.

Future research should explore the impact of this type of lesson over the span of a unit. Does embedding structured supplements in multiple read aloud lessons during a unit of study positively impact comprehension outcomes to a greater extent than the effects detected in the present study? And how far do those effects travel along a continuum of transfer as documented in multiple MORE intervention studies (Kim et al., 2023a)? The present study did not include a far-transfer passage due to time constraints. A far transfer passage that was based on the schema taught during the lessons but included no directly taught words would also provide a more stringent test of whether the schema alone or in combination with the vocabulary networks supports text comprehension. Finally, there is a need for insight from teachers about their perceptions of structured read aloud supplements. While a brief survey to participating teachers yielded favorable responses, in-depth interviews might highlight ways in which structure supplements can be expanded on and improved.

In conclusion, the findings from this study provide causal evidence for including structured supplements for teachers in read aloud lessons. Structured supplements promote focused emphasis on domain-specific vocabulary instruction, questioning, and activities to promote schema instantiation. Moreover, the structured supplements used in our study are curriculum agnostic and were informed by previous research on read alouds, teacher language and discussion, and schema development. Thus, they could be easily incorporated into a range of classroom curricula and instruction practices, particularly content literacy programs that aim to improve background knowledge, vocabulary, and comprehension in the elementary grades (Cabell & Hwang, 2020). The present study provides evidence for the benefit of embedding structured supplements into read aloud lessons that focus on identifying links between concepts and domains. As the Science of Reading continues to gain traction within public discourse, it is important to provide teachers with tools, strategies, and usable knowledge that helps them integrate principles of the Science of Teaching Reading into their instruction. Indeed, "becoming a good reader requires books that explain how and why things function as they do. Becoming a good reader depends upon teachers who insist that students think about the interconnections among ideas as they read" (Anderson & Pearson, 1984, p. 286). Embedding structured supplements into a read aloud gives teachers the framework and support to focus on the interconnections between concepts that can have positive impacts on comprehension outcomes.

Acknowledgements

We thank the teachers and students for their participation in this study. We also thank Ethan Scherer, Josh Gilbert, Jackie Relyea, Mary Burkhauser, and Taylor Harrison for their insights and expertise.

Funding

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 funding agency.

Declaration of Interest/Disclosure statement

We have no competing financial interests of benefits that have arisen from the direct applications of our research.

Data availability statement.

The data set associated with this paper will be provided at a later date along with analysis code.

References

Ahmed, Y., Francis, D. J., York, M., Fletcher, J. M., Barnes, M., & Kulesz, P. (2016). Validation of the direct and inferential mediation (DIME) model of reading comprehension in grades 7 through 12. *Contemporary Educational Psychology*, *44–45*, 68–82. https://doi.org/10.1016/j.cedpsych.2016.02.002

Anderson, R. (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). Newark, DE: 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 Assn.

Anderson, R. C., & Pearson, P. D. (1984). A schema-theoretic view of basic processing in reading. In P. D. Pearson (Ed.), *Handbook of Reading Research* (pp. 255–291). New York: Longman.

Ard, L. M., & Beverly, B. L. (2004). Preschool word learning during joint book reading: Effect of adult questions and comments. *Communication Disorders Quarterly*, *26*(1), 17-28,56-58. http://dx.doi.org/10.1177/15257401040260010101

Baker, D. L., Santoro, L., Biancarosa, G., Baker, S. K., Fien, H., & Otterstedt, J. (2020). Effects of a read aloud intervention on first grade student vocabulary, listening comprehension, and language proficiency. *Reading and Writing: An Interdisciplinary Journal*, *33*(10), 2697–2724.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Biemiller, A., & Boote, C. (2006). An effective method for building meaning vocabulary in primary grades. *Journal of Educational Psychology*, *98*(1), 44–62. https://doi.org/10.1037/0022-0663.98.1.44

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

Cervetti, G. N., & Wright, T. S. (2020). The role of knowledge in understanding and learning from text. In *Handbook of reading research, volume v*. Routledge.

Collins, M. F. (2016). Supporting inferential thinking in preschoolers: Effects of discussion on children’s story comprehension. *Early Education and Development*, *27*(7), 932–956. https://doi.org/10.1080/10409289.2016.1170523

Coyne, M. D., Simmons, D. C., Kame’enui, E. J., & Stoolmiller, M. (2004). Teaching vocabulary during shared storybook readings: An examination of differential effects. *Exceptionality*, *12*(3), 145–162. https://doi.org/10.1207/s15327035ex1203_3

Coyne, Michael. D., McCoach, D. B., Loftus, S., Zipoli, R., Ruby, M., Crevecoeur, YvelC., & Kapp, S. (2010). Direct and extended vocabulary instruction in kindergarten: Investigating transfer effects. *Journal of Research on Educational Effectiveness*, *3*(2), 93–120. https://doi.org/10.1080/19345741003592410

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

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review*, *18*(1), 23–45. https://doi.org/10.1016/S0272-7358(97)00043-3

Dedoose Version 7.0.23, web application for managing, analyzing, and presenting qualitative and mixed method research data (2016). Los Angeles, CA: SocioCultural Research Consultants, LLC www.dedoose.com

Elley, W. B. (1989). Vocabulary acquisition from listening to stories. *Reading Research Quarterly*, *24*(2), 174. https://doi.org/10.2307/747863

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

Ezell, H., & Justice, L. (2005). *Shared storybook reading: Building young children's language & emergent literacy skills*. Baltimore: Paul H. Brookes.

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

Gilbert, J. B., Kim, J. S., & Miratrix, L. W. (2023). Modeling item-level heterogeneous treatment effects with the explanatory item response model: Leveraging large-scale online assessments to pinpoint the impact of educational interventions. *Journal of Educational and Behavioral Statistics*, 10769986231171710. https://doi.org/10.3102/10769986231171710

Hattan, C., Singer, L. M., Loughlin, S., & Alexander, P. A. (2015). Prior knowledge activation in design and in practice. *Literacy Research: Theory, Method, and Practice*, *64*(1), 478–497. https://doi.org/10.1177/2381336915617603

Huck, A. (2019). Elementary Social Studies Content Integration in CCLS: An Analysis of Content Integration. *Social Studies*, *110*(1), 1–16. https://doi.org/10.1080/00377996.2018.1524359

Hwang, H., Cabell, S. Q., & Joyner, R. E. (2021). Effects of integrated literacy and content-area instruction on vocabulary and comprehension in the elementary years: A meta-analysis. *Scientific Studies of Reading*, 1–27. https://doi.org/10.1080/10888438.2021.1954005

Kaefer, T. (2020). When did you learn it? How background knowledge impacts attention and comprehension in read-aloud activities. *Reading Research Quarterly*, *55*(S1), S173–S183. https://doi.org/10.1002/rrq.344

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*. 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*. 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*. 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). New York, NY: Oxford University Press.

Kintsch, W. (1993). Information accretion and reduction in text processing: Inferences. *Discourse Processes*, *16*(1–2), 193–202. https://doi.org/10.1080/01638539309544837

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

Kintsch, W. (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). Newark, DE: 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: Vol. Fourth edition*. The Guilford Press. http://ezp-prod1.hul.harvard.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=1078917&site=ehost-live&scope=site

Language and Reading Research Consortium, Pratt, A., & Logan, J. (2014). Improving language-focused comprehension Instruction in primary-grade classrooms: Impacts of the Let's Know! Experimental curriculum. *Educational Psychology Review*, *26*(3), 357–377. https://doi.org/10.1007/s10648-014-9275-1

Lin-Siegler, X., Ahn, J. N., Chen, J., Fang, F.-F. A., & Luna-Lucero, M. (2016). Even Einstein struggled: Effects of learning about great scientists' struggles on high school students' motivation to learn science. *Journal of Educational Psychology*, *108*(3), 314–328. https://doi.org/10.1037/edu0000092

MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*, *39*(1), 99–128. https://doi.org/10.1207/s15327906mbr3901_4

McCarthy, K. S., Guerrero, T. A., Kent, K. M., Allen, L. K., McNamara, D. S., Chao, S.-F., Steinberg, J., O'Reilly, T., & Sabatini, J. (2018). Comprehension in a scenario-based assessment: Domain and topic-specific background knowledge. *Discourse Processes*, *55*(5–6), 510–524. https://doi.org/10.1080/0163853X.2018.1460159

McKeown, M. G., & Beck, I. L. (2014). Effects of vocabulary instruction on measures of language processing: Comparing two approaches. *Early Childhood Research Quarterly*, *29*(4), 520–530. https://doi.org/10.1016/j.ecresq.2014.06.002

Mosher, D.M., Burkhauser, M.A. and Kim, J.S. (2023). Improving second-grade reading comprehension through a sustained content literacy intervention: A mixed-methods study examining the mediating role of domain-specific vocabulary. Under Review.

Muthén, L. K., & Muthén, B. O. (2012). Mplus User's Guide (7th ed.). Los Angeles, CA: Muthén & Muthén.

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

Northern, A. & Petrilli, M. (2000). Forward and executive summary. Tyner, A. & Kabourek. S. *Social Studies instruction and reading comprehension: Evidence from the early childhood longitudinal study.* 1-8. Thomas B. Fordham Institute.

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

Parsons, A. W., & Bryant, C. L. (2016). Deepening kindergarteners' science vocabulary: A design study. *The Journal of Educational Research*, *109*(4), 375–390. https://doi.org/10.1080/00220671.2014.968913

Penno, J. F., Wilkinson, I. A. G., & Moore, D. W. (2002). Vocabulary acquisition from teacher explanation and repeated listening to stories: Do they overcome the Matthew effect? *Journal of Educational Psychology*, *94*(1), 23–33. https://doi.org/10.1037/0022-0663.94.1.23

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

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Sage Publications.

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

Relyea, J. E., Kim, J. S., & Rich, P. (2022). Effectiveness of tier 1 content-integrated literacy intervention on early elementary english learners' reading comprehension and writing: Evidence from randomized controlled trial. In *EdWorkingPapers.com*. Annenberg Institute at Brown University. https://www.edworkingpapers.com/index.php/ai22-606

Robbins, C., & Ehri, L. C. (1994). Reading storybooks to kindergartners helps them learn new vocabulary words. *Journal of Educational Psychology*, *86*(1), 54–64. https://doi.org/10.1037/0022-0663.86.1.54

Senechal, M., & Cornell, E. H. (1993). Vocabulary acquisition through shared reading experiences. *Reading Research Quarterly*, *28*(4), 360. https://doi.org/10.2307/747933

Shanahan, T. (2020). What constitutes a science of reading instruction? *Reading Research Quarterly*, *55*(S1), S235–S247. https://doi.org/10.1002/rrq.349

Silverman, R. (2007). A comparison of three methods of vocabulary instruction during read-alouds in kindergarten. *The Elementary School Journal*, *108*(2), 97–113. https://doi.org/10.1086/525549

StataCorp. 2021. *Stata Statistical Software: Release 17*. College Station, TX: StataCorp LLC.

Swanson, E., Vaughn, S., Wanzek, J., Petscher, Y., Heckert, J., Cavanaugh, C., Kraft, G., & Tackett, K. (2011). A synthesis of read-aloud interventions on early reading outcomes among preschool through third graders at risk for reading difficulties. *Journal of Learning Disabilities*, *44*(3), 258–275. https://doi.org/10.1177/0022219410378444

U.S. Department of Education, Institute of Education Sciences, National center for Education Statistics, National Assessment of Educational Progress (NAEP), 2019 Reading Assessments & Achievement Levels.

van Kleeck, A., Vander Woude, J., & Hammett, L. (2006). Fostering literal and inferential language skills in head start preschoolers with language impairment using scripted book-sharing discussions. *American Journal of Speech-Language Pathology*, *15*(1), 85–95. https://doi.org/10.1044/1058-0360(2006/009)

Wright, T. S., Cervetti, G. N., Wise, C., & McClung, N. A. (2022). The impact of knowledge-building through conceptually-coherent read alouds on vocabulary and comprehension. *Reading Psychology*, *43*(1), 70–84. https://doi.org/10.1080/02702711.2021.2020187

Figure 1. Intervention General and Topic Schemas

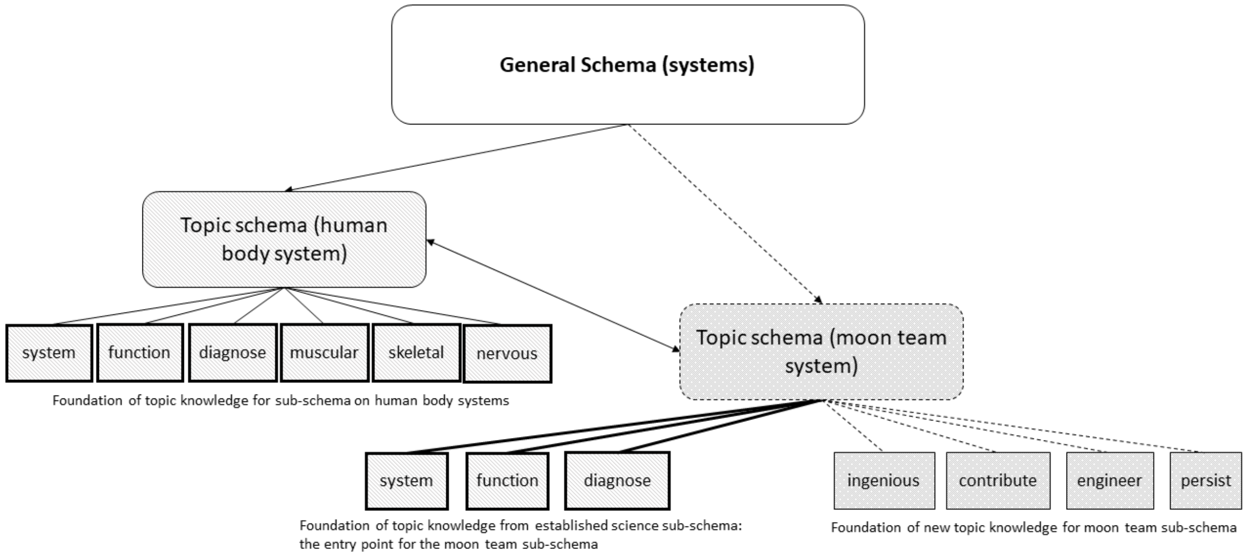


Figure 2. Study Theory of Change

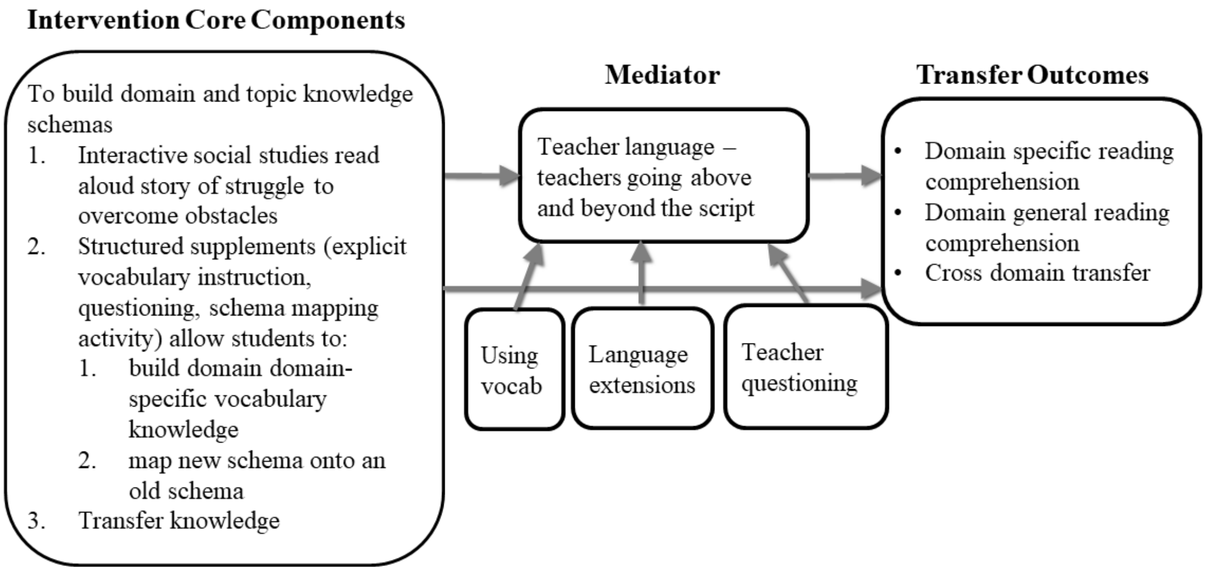


Figure 3. Consort Diagram for Sample Flow (Students)

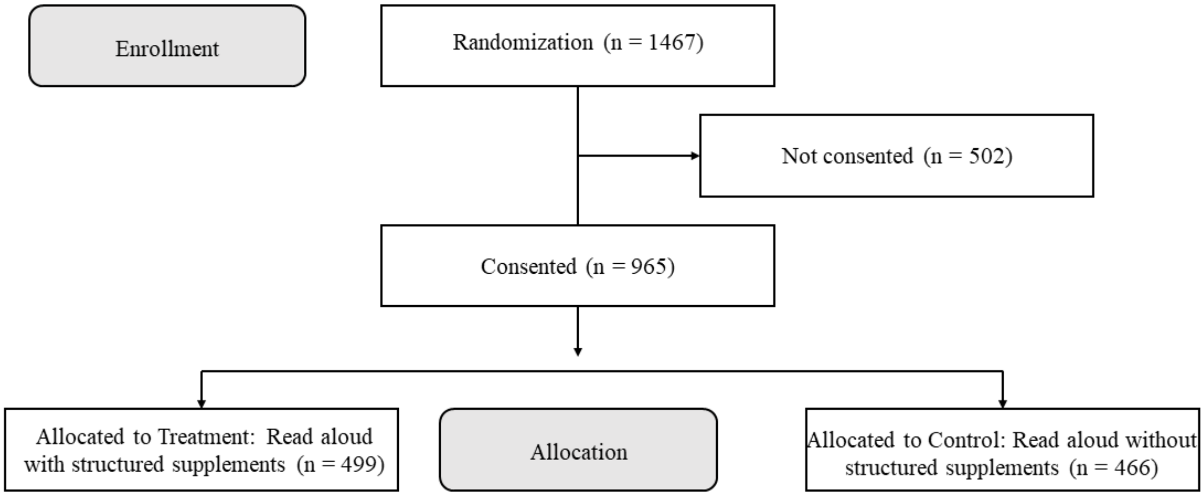


Figure 4: Schema Mapping Activity Using "Madlibs"

function muscles system nervous skeleton contribute

Human body

My human body is a system. The human body system has many parts. These parts function together to make the body work to keep us alive and healthy. One part of the human body system is your muscles. Your heart muscle has an important function. It moves on its own with each heartbeat. Another part of your body system is your skeleton. Your bones' function is to help support your body! Finally, the human body system needs the nervous system to tell your body what to do. All these different parts of your human body system contribute to helping your body function.

Moon team ***astronauts ingenious system function engineer contributed***

The moon team was a system. The moon team system had many people. These people functioned together to send astronauts to the moon and back. One part of the moon team system was an ingenious math expert like Katherine Johnson. She had an important function. She figured out the path to the moon and back. Another part of the moon team system was an ingenious engineer like Chuck Lowry. The engineer's function was to figure out how to slow down Apollo 11 when it fell to Earth! Finally, the moon team system needed astronauts like Neil Armstrong to fly to the moon. All these different people of the moon team system contributed to helping the moon team make it to the moon and back.

How is the moon team system similar to the human body system? (Pick the word that completes the sentence) (1 min)

persist engineer contribute ingenious

Both the human body system and the moon team system have many parts that contribute ***to helping it function.***

19

Figure 5: Teacher Language CFA

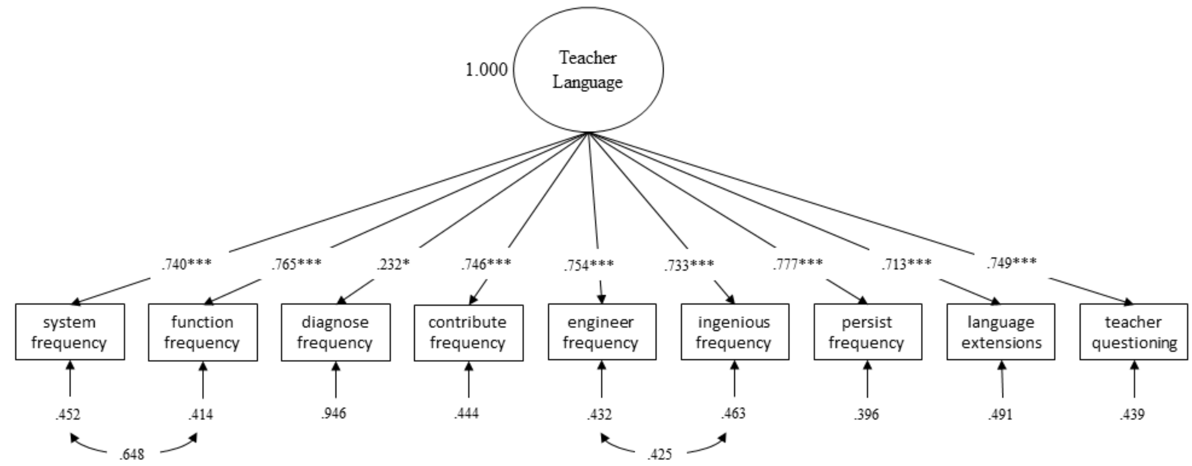


Figure 6: Domain-Specific Reading Comprehension CFA

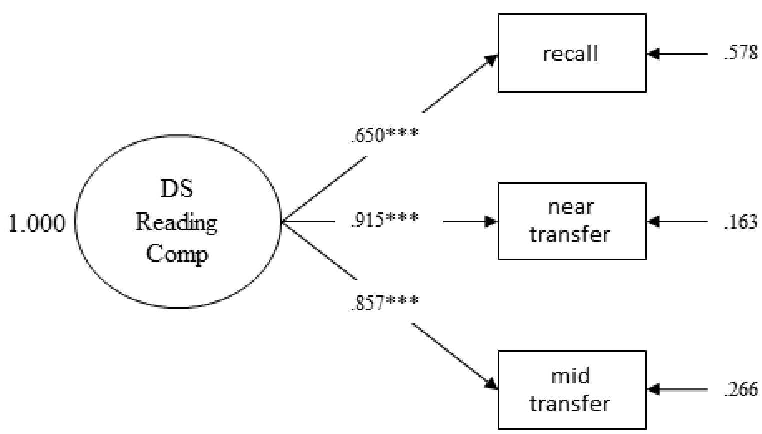


Figure 7: Mediation Model

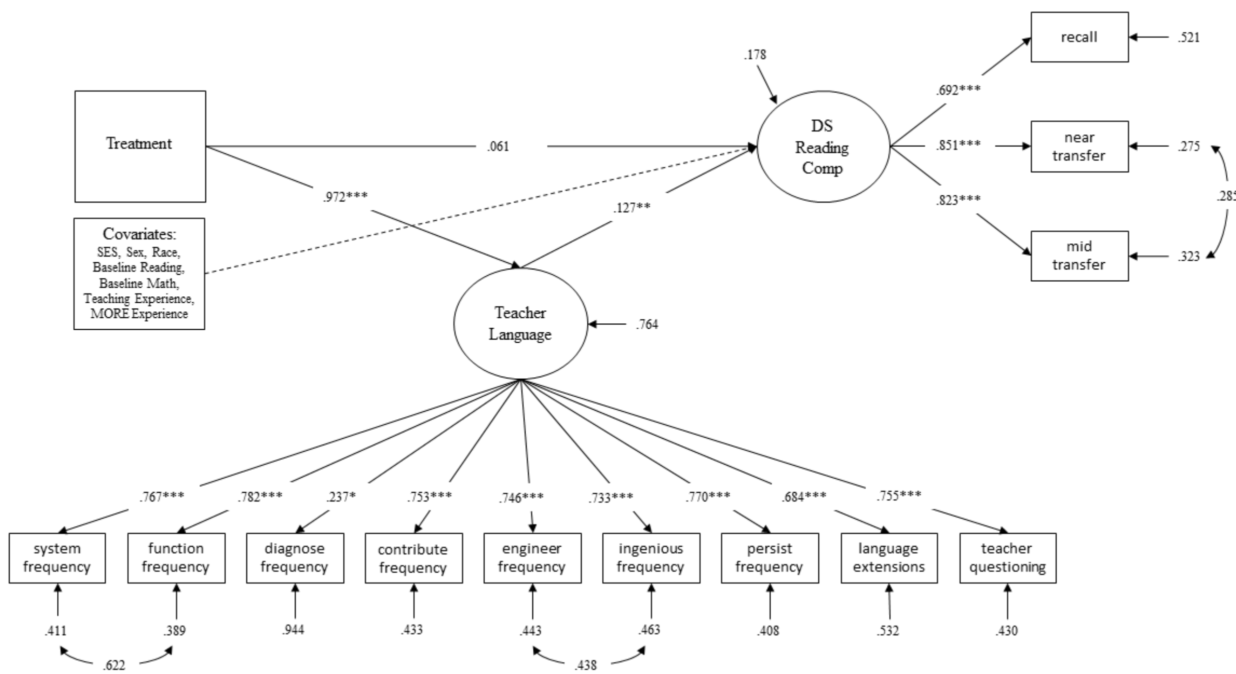


Table 1. Structured Supplements

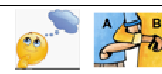
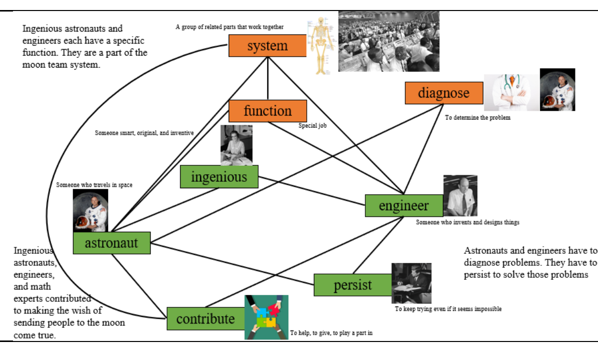
Structured Supplements	**Example**	
Activate existing schema	Class completes an anchor chart reviewing the three human body systems and their functions	***Human Body Systems*** **Part** | **Function** muscular system skeletal system nervous system
Introduce new domain-specific vocabulary with explanation of word's meaning and examples	*Let's say the word <u>ingenious</u> together.* *<u>Ingenious</u>.* *An <u>ingenious</u> person is someone who is original or inventive and smart. Before there were cars, people used to have to walk or ride horses everywhere. But <u>ingenious</u> <u>inventors</u> and <u>engineers</u> created cars and buses, making it so much easier for us to travel to different places – like from home to school. These people were <u>ingenious</u> because they were inventive and smart! They created something new.*	
Review topic/vocabulary knowledge	Partners answer questions with specific inclusion criteria for what students answers should contain (i.e., target words)	***How was Chuck Lowry an <u>ingenious</u> <u>engineer</u>? (1 min)***  **Must Have**: Answer includes the word <u>ingenious</u> or <u>engineer</u>. **Amazing**: Answer also includes <u>ingenious</u>, <u>engineer</u>, <u>contribute</u>, <u>function</u>, or <u>system</u>.
Map new topic schema onto different and already established schema	Partner discussion and "madlibs" schema mapping activity.	See figure 5
Concept mapping activity connecting words/concepts	Concept mapping activity	

Table 2. Descriptive Statistics for the Final Grade 3 Sample

Characteristics	N	Overall	Treatment	Control
Students	965	965	499	466
White	965	25%	24%	27%
Black	965	32%	34%	29%
Hispanic	965	31%	29%	33%
Asian	965	9%	9%	9%
Other	965	3%	4%	2%
Male	965	48%	48%	48%
Multilingual	965	25%	23%	26%
High SES	965	29%	27%	30%
Med SES	965	35%	35%	36%
Low SES	965	35%	37%	34%
MAP Reading, M (SD)	959	192.32 (18.34)	192.67 (18.33)	191.95 (18.36)
MAP Math, M (SD)	955	194.18 (14.87)	194.45 (14.92)	193.90 (14.84)
Recall, M (SD)	942	3.46 (1.46)	3.58 (1.43)	3.32 (1.47)
Near Transfer, M (SD)	934	8.28 (3.50)	8.63 (3.43)	7.89 (3.54)
Mid Transfer, M (SD)	910	6.87 (3.21)	7.14 (3.15)	6.57 (3.25)
Content Comprehension, M (SD)	965	14.49 (6.83)	15.30 (6.59)	13.63 (6.97)
EOG, M (SD)	964	538.29 (10.11)	538.46 (10.15)	538.12 (10.07)

Table 3. Correlation Matrix of Pre- and Post-Assessment Measures

Variable	1	2	3	4	5	6	7
1. MAP Reading Pre-test	-						
2. MAP Math Pre-test	0.80	-					
3. Recall	0.58	0.56	-				
4. Near Transfer	0.73	0.68	0.57	-			
5. Mid Transfer	0.69	0.64	0.55	0.76	-		
6. Content Comprehension	0.68	0.63	0.54	0.94	0.93	-	
7. EOG	0.86	0.76	0.58	0.73	0.69	0.69	-

All correlations statistically significant at the .05 level.

Table 4. Descriptive Statistics for Teacher-Level Variables

		Overall	Treatment	Control
Variable	N	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
Teachers	80	80	39	41
"System" Frequency	75	13.76 (14.18)	21.59 (15.60)	6.13 (6.66)
"Function" Frequency	75	4.75 (5.61)	7.38 (6.66)	2.18 (2.47)
"Diagnose" Frequency	75	0.27 (0.72)	0.27 (0.51)	0.26 (0.89)
"Contribute" Frequency	75	2.69 (4.41)	4.51 (5.41)	0.92 (1.98)
"Engineer" Frequency	75	1.79 (2.97)	2.73 (3.61)	0.87 (1.80)
"Ingenious" Frequency	75	4.57 (6.53)	7.38 (7.09)	1.84 (4.57)
"Persist" Frequency	75	3.65 (4.74)	5.11 (5.66)	2.24 (3.09)
Language Extensions	75	5.80 (6.49)	6.14 (6.38)	5.47 (6.67)
Number of teacher questions	75	65.65 (37.56)	80.05 (36.93)	51.63 (32.94)
Years Teaching	80	10.48 (8.88)	9.70 (9.37)	11.21 (8.44)
Past Experience with MORE	80	0.90 (0.94)	0.86 (0.92)	0.93 (0.96)

Table 5. Correlation Matrix for Teacher-Level Variables

Variable	1	2	3	4	5	6	7	8	9	10	11
1. System Frequency	-										
2. Function Frequency	0.83	-									
3. Diagnose Frequency	0.14	0.09	-								
4. Contribute Frequency	0.60	0.62	0.07	-							
5. Engineer Frequency	0.55	0.54	0.04	0.56	-						
6. Ingenious Frequency	0.53	0.55	0.06	0.60	0.71	-					
7. Persist Frequency	0.54	0.54	0.13	0.48	0.61	0.48	-				
8. Language Extensions	0.35	0.44	0.32	0.45	0.55	0.50	0.58	-			
9. Teacher Questioning	0.60	0.61	0.13	0.49	0.45	0.46	0.61	0.50	-		
10. Years Teaching	-0.22	-0.20	-0.16	-0.13	-0.11	-0.15	-0.08	-0.14	-0.09	-	
11. Experience with intervention	-0.08	0.04	-0.06	0.00	-0.20	-0.19	-0.16	-0.11	0.16	0.08	-

All correlations statistically significant at the .05 level.

Table 6. Fidelity of Implementation

	Overall	Treatment	Control	
	M (*SD*)	*M* (*SD*)	*M* (*SD*)	Difference
Lesson Adherence (0-7)	6.76 (1.63)	6.49 (1.89)	7.03 (1.28)	-.46 (.33)
Lesson Adherence (%)	84.5% (20.3)	81.1% (23.7)	87.8% (16.0)	5.8% (4.1)
Lesson Time	34.77 (14.42)	44.54 (12.42)	25.26 (8.80)	19.18^{***} (2.20)

Note. *M* = Mean; *SD* = Standard Deviation. The difference was estimated from a regression model that controls for fixed effects of randomization blocks (schools). Standard errors are in parenthesis.

$^{***}p < .001$.

Table 7. Results of Multiple Imputation Hierarchical Linear Models Predicting Treatment Effects of Structured Supplements on Measures of Reading Comprehension

Variable	Recall	Near Transfer	Mid Transfer	Content Comprehension	EOG
			Fixed effects		
Intercept, γ_{00}	-.15 (.2)	-.32† (.18)	-.5* (.21)	-.42* (.2)	0 (.07)
Treatment, γ_{01}	.17*(.07)	.17**(.06)	.18*(.07)	.18*(.07)	.01(.02)
			Variance Components		
Level 1, ε_{ijk}	.53	.34	.34	.28	.09
Level 2, u_{0jk}	.04	.04	.07	.06	0
N	942	934	910	934	964

Note. Point estimates derived from hierarchical models including the treatment indicator, school-fixed effects, student demographics, and reading and math pretest scores, with teacher-random intercepts.

† $p < .10$. ** $p < .05$. ** $p < .10$.

Table 8. Sensitivity Check: Results of Regression Models Predicting Treatment Effects of Structured Supplements on Measures of Reading Comprehension Using Aggregated Classroom-Level Data

Variable	Recall	Near Transfer	Mid Transfer	Content Comprehension	EOG
			Fixed Effects		
Intercept	1.57 (3.43)	-3.96 (2.55)	-5.11 (3.18)	-4.21 (2.67)	-1.31 (1.08)
Treatment	.23*(.1)	.25***(.07)	.21*(.09)	.24**(.08)	.04(.03)
R^2	.80	.90	.84	.89	.95
N	79	80	80	80	80

Note. Point estimates derived from OLS regression models using classroom-level aggregated data including treatment indicator, school-fixed effects, student demographics, and reading and math pretest scores.
** $p < .05$. ** $p < .01$. ** $p < .001$

Appendix. Benjamini-Hochberg Procedure to Test for False Discoveries

	Reported p-value (p_x)	X = p-value rank	M = # of p-values	New Critical p-value ($p_x{}'$ =0.05x/5)	Finding p-value < New Critical p-value?	Statistical Significance after BH Correction?	Largest p-value rank less than or equal to p_x'
Content							
Comprehension	0.008	1	5	0.01	TRUE	TRUE	0.04
Near transfer	0.00600	2	5	0.02	TRUE	TRUE	0.04
Recall	0.01200	3	5	0.03	TRUE	TRUE	0.04
Near transfer	0.01800	4	5	0.04	TRUE	TRUE	0.04
EOG	0.69100	5	5	0.05	FALSE	FALSE	0.04