ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 22-600

Long-Term Effects of a Sustained Content Literacy Intervention on Third Graders' Reading Comprehension Outcomes

James S. Kim
Harvard University

Patrick Rich
Harvard University

Ethan Scherer
Harvard University

To address the challenge of improving third grade reading comprehension, we developed and evaluated the long-term effects of a sustained content literacy intervention called the Model of Reading Engagement (MORE), which emphasizes building domain and topic knowledge schemas from Grade 1 to Grade 3. The MORE intervention emphasizes thematic lessons that provide an intellectual framework for helping students connect new learning to a general schema (e.g., how scientists study past events, how systems function properly). Over three years, the treatment group students participated in (a) spring Grade 1 thematic content literacy lessons in science and social studies, (b) fall to spring Grade 2 thematic content literacy lessons in science, (c) remote Grade 3 thematic content literacy lessons in science, and (d) wide reading of thematically related informational texts in the summer months following Grade 1 and Grade 2. During the third grade school year (SY 2020-21), the COVID-19 pandemic required remote schooling to be in place from fall to spring and the Grade 3 MORE was provided to both treatment and control students. Accordingly, we examine long-term effects on third graders' outcomes comparing a treatment group that received the Grade 1, Grade 2, and Grade 3 MORE treatment to a control condition that received the Grade 3 MORE treatment. Intent-to-treat estimates show that the students randomly assigned to the treatment condition outperformed control students in reading comprehension (ES = 0.11) and mathematics (ES = 0.14) on third grade state standardized assessments. Subgroup analyses also revealed positive impacts for student living in low- to moderate-socioeconomic status neighborhoods on both reading comprehension (ES = .13) and mathematics (ES = .20). Findings indicate that a sustained content literacy intervention may be a scalable approach for accelerating and equalizing third-graders' reading comprehension and math outcomes.

VERSION: July 2022

Suggested citation: Kim, James S., Patrick Rich, and Ethan Scherer. (2022). Long-Term Effects of a Sustained Content Literacy Intervention on Third Graders' Reading Comprehension Outcomes. (EdWorkingPaper: 22-600). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/g6m2-ns12

Long-Term Effects of a Sustained Content Literacy Intervention on Third Graders' Reading Comprehension Outcomes

James S. Kim, Patrick Rich, Ethan Scherer

Harvard Graduate School of Education

June 2022

Abstract

To address the challenge of improving third grade reading comprehension, we developed and evaluated the long-term effects of a sustained content literacy intervention called the Model of Reading Engagement (MORE), which emphasizes building domain and topic knowledge schemas from Grade 1 to Grade 3. The MORE intervention emphasizes thematic lessons that provide an intellectual framework for helping students connect new learning to a general schema (e.g., how scientists study past events, how systems function properly). Over three years, the treatment group students participated in (a) spring Grade 1 thematic content literacy lessons in science and social studies, (b) fall to spring Grade 2 thematic content literacy lessons in science, (c) remote Grade 3 thematic content literacy lessons in science, and (d) wide reading of thematically related informational texts in the summer months following Grade 1 and Grade 2. During the third grade school year (SY 2020-21), the COVID-19 pandemic required remote schooling to be in place from fall to spring and the Grade 3 MORE was provided to both treatment and control students. Accordingly, we examine long-term effects on third graders' outcomes comparing a treatment group that received the Grade 1, Grade 2, and Grade 3 MORE treatment to a control condition that received the Grade 3 MORE treatment. Intent-to-treat estimates show that the students randomly assigned to the treatment condition outperformed control students in reading comprehension (ES = 0.11) and mathematics (ES = 0.14) on third grade state standardized assessments. Subgroup analyses also revealed positive impacts for student living in low- to moderate-socioeconomic status neighborhoods on both reading comprehension (ES = .13) and mathematics (ES = .20). Findings indicate that a sustained content literacy intervention may be a scalable approach for accelerating and equalizing third-graders' reading comprehension and math outcomes.

Introduction

The ability to read and evaluate grade-level text proficiently by third grade represents a key milestone in a student's educational history. Children who can read proficiently by third grade are more likely than their non-proficient peers to graduate from high school and be prepared for college and a career (Hernandez, 2011). To be a proficient reader, however, a student must have strong word reading skills and have the requisite domain and topic knowledge to read complex nonfiction texts with high knowledge demands (Anderson & Pearson, 1984; Cervetti & Wright, 2020; Reardon et al., 2012). Data from the National Assessment of Educational Progress Long-Term Trend ("Trend NAEP") further suggests that US 9-year-olds have made no gains in reading comprehension over the decade and that gaps between the lowest-performing and highest-performing students have widened over time (National Center for Education Statistics, 2019). Even more troubling is the evidence indicating that gaps between low- and high-performing students are larger on reading comprehension than basic word-reading skills (D'Agostino & Rodgers, 2017; Fryer & Levitt, 2006). Furthermore, contextual factors like poverty are amplifying the negative impacts of the pandemic on third graders' reading performance in high-poverty schools (Goldhaber et al., 2022; Lewis et al., 2021). Why are so many elementary grade students struggling to read proficiently by third grade?

The limited opportunities for children to participate in sustained content literacy interventions that emphasize the integration of content knowledge building into literacy instruction may play a pivotal role. Indeed, several descriptive studies have shown that the overemphasis on English language arts and math instruction limits students' opportunities to acquire science and social studies content knowledge, particularly for students of color and students in high-poverty schools (Arold & Shakeel, 2021; Blank, 2012). In response to these systemic challenges, researchers and policymakers have emphasized the need for multi-year content literacy interventions that sustain and align content and practices across consecutive grades and across school and home contexts (Alexander et al., 2007; Bailey et al., 2017; Bronfenbrenner & Morris, 2006; Nagy, 2005; Newmann et al., 2001; Pressley et al., 2007). Indeed, early literacy interventions that begin in first grade and continue to third grade may be critical to building not only students' word knowledge but also domain and topic knowledge.

The explanatory theory guiding this study was the following: multi-year, sustained content literacy interventions are needed to help students transfer their newly acquired domain and topic knowledge to passages about related topics. A key feature of students' domain and topic knowledge is the mental representation of knowledge in the form of schemas and schema instantiation facilitates readers' comprehension and transfer to new topics (Alexander, 2003; Anderson, 1984; Anderson & Pearson, 1984). Importantly, it takes sustained effort for several years to help students develop schemas that can be accessed when reading about new, related topics in science and social studies. However, sustained content literacy interventions that continue for more than one year are rare in the elementary grades and evidence of transfer remains elusive. To address this research gap, we developed, implemented, and evaluated the Model of Reading Engagement (MORE), a multi-year content literacy intervention that sustains and aligns content and practices from first to third grade.

Theoretical Background

Sustaining and Aligning Content Through Systematic Schema-Building Lessons

There is general agreement among researchers and practitioners that sustained content literacy interventions share the common goal of improving the quantity and quality of students' domain and topic knowledge through science and social studies content as a lever for improving reading comprehension (Cromley & Azvedo, 2007; Kendeou & O'Brien, 2016; Kintsch, 1988, 2009). Consistent with this aim, we define MORE as sustained content literacy intervention in which thematic lessons provide recurrent exposure to schema-building activities so that knowledge is accessible to children as they read increasingly complex informational texts (Cervetti & Wright, 2020; Kim, Relyea et al. 2021; Smith et al., 2021).

To prepare students to understand and evaluate complex informational texts in later grades, we designed a spiral MORE intervention that helps teachers stay on topic across grades to systematically build novice learners' schemas. Schemas are intellectual structures that help students acquire, organize, and apply knowledge while learning from school texts across content areas (Alexander, 2003). Schemas are mental structures that help novice learners build domain expertise and organize topics within a given domain like science. To be useful, schemas can be neither too specific nor too decontextualized, but "linked to abstract, generalizable features of situations" (Kintsch, 2009, p. 231). For example, in the first grade MORE lessons in science, children learned about the science topic: how do animals survive in their habitat? They started with concrete examples of living animals such as polar bears and what they look like, how they behave, where they live, and how they survive and adapt. Then, to further sustain content across school and home contexts in the summer months, students read thematically related books on topics studied during the school at home in the summer months. To continue to build on the schema for animal survival, students in second grade studied how scientists—paleontologists—study dinosaur mass extinctions by using fossil evidence to develop theories about dinosaurs' physical characteristics, behaviors, traits, and how they survived, adapted, and became extinct. Finally, third grade MORE science lessons focused on the schema for "system" as students studied how sub-systems (e.g., skeletal, muscular, nervous system) work and move together to keep the human body functioning properly.

Over time, children can form a general and topic specific schema for survival or system. Furthermore, they can acquire knowledge about key concepts (e.g., physical and behavioral characteristics, habitat, survival, adaptation, extinction, evidence, theory, system). The topic-specific schemas in MORE shared structured features that were designed to help young children situate and connect newly acquired knowledge to the general schema that unified content across first and second grade (i.e., how scientists study past events) and extended content through third grade (i.e., how systems function properly). Thus, the MORE intervention includes a spiral curriculum in which content is presented in a way that enables the expansion and deepening of schemas over time.

Sustaining and Aligning Practices Through Systematic Schema-Building Lessons

The theory of change for MORE in Figure 1 displays how the core intervention components work to promote transfer. We hypothesize that transfer measures of third-graders' domain specific and domain general general reading comprehension will improve if students can build and transfer their domain and topic-specific schemas as indicated by their vocabulary knowledge. We conceptualize transfer (Barnett & Ceci, 2002) as involving two factors: the content (i.e., what is transferred) and the context (i.e., when and where is learning transferred to and from). In essence, the MORE theory of change indicates that interleaving and sustaining practices across grades will first improve students' domain specific vocabulary knowledge and ultimately promote transfer in reading (i.e., domain specific and domain general reading comprehension) and cross domain transfer (i.e., mathematics) on third-grade outcomes.

Key practices to build domain and topic knowledge include the following: (a) the use of thematic lessons to foster situational interest, (b) concept maps to reinforce schema development, (c) interactive read alouds of thematically related informational texts during in-school lessons and as well as summer support for reading thematically related informational text in print and through a digital app. Key practices to transfer domain and topic knowledge included (a) argumentative writing, (b) collaborative research and (c) structured word inquiry. Details on the MORE practices are reported in previous studies of first and second grade implementations (Kim, Relyea et al., 2021, Kim, Burkhauser et al., 2022).

Present Study and Impacts of COVID-19 on Longitudinal Research

This study builds on previous research on the implementation and effectiveness of the MORE intervention. Previous research provides evidence of short-term impacts on one and two-year implementations of MORE. In an early efficacy study of first grade MORE science lessons (Kim, Burkhauser et al., 2021), students in the treatment group outperformed control students on vocabulary knowledge (ES = .30), argumentative writing (ES = .24), and reading comprehension (ES = .11). Following the first-grade MORE study, we designed a test of the full MORE intervention (first, second, third grade). Originally, this study was designed to compare a treatment group of students who received MORE from first to third grade versus a control group that received typical instruction. However, third grade implementation occurred during the COVID-19 pandemic year (school year 2020-21) when schools were closed for most of the year and remote instruction took place. To accommodate these changes and continue implementation, both treatment and control schools were given access to third-grade MORE. Accordingly, this study examines whether the timing of early implementation of MORE in first and second grade has long-term impacts on third graders' outcomes in contrast to a later implementation of third grade MORE only in the control condition.

Potential Mechanisms Driving Long-Term Impacts

Although previous studies demonstrate consistent and positive impacts on short-term outcomes (Kim, Burkhauser, et al., 2021, 2022), the question of the long-term impact on third grade outcomes has not been studied. There are several potential mechanisms through which MORE could potentially promote long-term impacts through third grade. First, it may be critical for intervention researchers to target trifecta skills "that are malleable, fundamental, and would not have developed in the absence of intervention” (Bailey et al., 2020, pp. 66–67). In many ways, the MORE intervention targets both malleable and fundamental domain specific vocabulary knowledge and comprehension outcomes that are typically difficult to improve and measure in the early elementary grades. In particular, the MORE lessons highlight building students’ abstract knowledge structures or “schemas” that lay the groundwork for further disciplinary learning and literacy development. In other words, schemas beget further development and extension of knowledge, and can thus be considered fundamental “springboards” that can alter the course of development. These schemas are malleable for text comprehension, particularly during the early elementary grades. It is also possible that the schemas developed by MORE may not be developed through normal business-as-usual instruction (Kim et al., 2022; Kim, Burkhauser et al., 2021; Kim, Relyea et al., 2021). In addition to the skill-building components, building sustaining environments is critical to sustained intervention impacts on early skill development (Bailey et al., 2017). The MORE spiral curriculum that students receive across successive years may help to ensure that the schemas were deepened and broadened over time.

Second, it may be critical to target unconstrained competencies which are often under-emphasized in the early elementary grades. Unconstrained competencies like domain specific vocabulary and schema development in science and social studies become increasingly important in school (Paris, 2015; Snow & Matthews, 2016). However, in the absence of coordinated efforts to promote schema development, it is difficult to sustain and align content and practices across multiple grades and between school and home contexts. To overcome these institutional challenges, we implemented this study in the context of a long-term research-practice partnership that targeted improvement in unconstrained competencies which are crucial for all school learning and later educational success.

Third, it is critical to examine how far intervention effects travel. To do so, we measured impact on both domain specific and domain general measures of reading comprehension. Furthermore, we assessed cross domain transfer from reading to mathematics. It is theoretically plausible to expect cross domain transfer if the schemas and vocabulary networks that children acquire via the MORE intervention are useful for reading and evaluating academic texts across a variety of school subjects. In other words, vocabulary knowledge is critical to understanding written text that students encounter in science, social studies, and mathematics textbooks (Fitzgerald et al., 2019). In addition, unconstrained competencies like vocabulary knowledge, nonverbal reasoning, working memory, and attention drive both reading and mathematics development (Cirino et al., 2018; Geary, 2010) and may be facilitated through the intervention core program components in MORE (e.g., see Figure 1).

Research Aims and Questions

The main contribution of this study, then, is to address the questions of whether and to what extent the impacts of MORE (a) persist into third grade and (b) promote cross-domain transfer from reading to mathematics, and (c) generalize to all subgroups of students and schools. In particular, this study addresses the following research questions (RQs):

1. Compared to students who only participated in the third grade MORE intervention, what is the intent-to-treat (ITT) and treatment-on-the-treated (TOT) impact the MORE first to third grade sustained content literacy intervention in science on third graders' domain specific vocabulary knowledge and comprehension and domain general reading comprehension? Are there cross-domain transfer effects on students' end-of-grade 3 mathematics outcomes?
2. To what extent to student characteristics (i.e., initial reading comprehension scores, racial and ethnic backgrounds, and English learner status) and school characteristics (i.e., neighborhood socioeconomic status) moderate the ITT impacts on third grade outcomes?

Methods

Analytic Sample

The students included in this analysis are those who were first grade in 2018-19 and third grade in 2020-21. Thirty schools from a large district in the southeast United States were included in the study: 15 schools were randomized to treatment and 15 to control. The students in the treatment schools had the opportunity to receive the MORE science and social studies curriculum in 2018-19 (30 hours), MORE science in 2019-20 (30 hours), while the control students received the counterfactual of regular science and social studies instruction in both years. In 2020-21, due to the disruptions caused by COVID-19, MORE science (10 hours) was offered to both treatment and control students during remote schooling. Thus, this study compares two groups of students: (a) MORE students who were randomly assigned to receive MORE from first to third grade, compared to (b) control students in a counterfactual condition in which students were randomly assigned to receive MORE instruction in only third grade. Figure 2 visually displays the timeline for the implementation activities and the assessment of long-term impacts on third graders' reading comprehension and math outcomes. In the present study, we report impacts on North Carolina end-of-grade reading and mathematics tests.

Baseline Equivalence Between Conditions

Table 1 shows the balance checks between treatment and control groups at baseline. School-level randomization was assigned based upon publicly available historical third grade EOG means, school size, and prior experience with MORE. For pretest measures in reading and mathematics at baseline of first grade, the treatment and control students were not balanced, with treatment group students performing significantly lower relative to control group students. Furthermore, treatment group students were more likely to live in low socioeconomic status neighborhoods (51%) than control group students (35%). The two groups were balanced on most other demographic categories. To address these imbalances, our analyses included pretest and demographic variables as covariates. In addition, because by chance the treatment group performed lower on baseline assessment, we view our results to be a lower bound on the effects.

Fidelity of Implementation: Program Differentiation in the Amount of Instructional Time by Subjects

To assess fidelity of implementation, we focused on program differentiation between MORE and control conditions in the amount of instruction time spent on English language arts (ELA)/reading, science, social studies, and mathematics subjects. We conducted a survey of all teachers in 2019, 2020, and 2021 to obtain the information about the amount of instruction time by subjects (the question item about mathematics instruction time was only included on the 2020 and 2021 surveys). Table 2 summarizes the average amount of weekly exposure to ELA/reading, science, social studies, and mathematics by treatment condition. For ELA/reading and mathematics, treatment and control teachers spent the similar amount of instruction time (no statistical difference between treatment and control, $p > .05$).

Treatment group teachers devoted significantly more instruction time to science and social studies than control group teachers ($p < .05$). In sum, treatment group students had more exposure to science and social studies content than their counterparts over the course of the three-year study.

Measures

Domain Specific (Science) Vocabulary Knowledge. We developed a 36-item measure to assess student’s science vocabulary knowledge depth immediately after the third grade implementation. Each semantic association task assesses students’ definitional knowledge of taught science words and their ability to identify relations between the target word and other known words. We replicated a semantic association task (Kim et al., 2020) for our study to assess third graders' ability to identify semantically related words and their knowledge of how words are networked to each other. For example, in Grade 3, the task included seven domain-specific words taught in the third grade MORE science lessons (i.e., taught words): *skeletal, muscular, nervous, diagnosis, structure, system, function.* The task also included five associated words that were not directly taught in the MORE lessons (i.e., untaught words): *signal, repair, organ, fracture, sensory*. The prompt asked students to "circle all of the words that go with the word" (e.g., signal) and presented four options (e.g., metal, messenger, transmit, similar). We also included 12 first grade and 12 second grade MORE science words taught in previous grades. Each item is scored 0 to 4, where students also get credit for not circling unrelated words. Cronbach alpha reliabilities were .85 for taught words and .77 for untaught words in our earlier efficacy study involving Grade 1 students (Kim et al., 2020).

Domain Specific (Science) Reading Comprehension. For the science reading comprehension measure in third grade, we developed a 29-item multiple choice test to assess students' ability to read three passages: a near, mid, and far transfer passage. Near and mid passages included taught words (i.e., word associations) in context whereas the far passage did not include taught words. The passages focused on the muscular system of primates (monkeys), the skeletal system of birds, and non-living things (skyscrapers). Cronbach's alpha reliabilities were above .80.

Domain General Reading Comprehension and Mathematics Achievement. Both domain-general reading comprehension and mathematics ability in third grade were measured using statewide EOG standardized assessments. The assessments were given during the last days of school and had well-documented psychometric characteristics (North Carolina Department of Public Instruction, 2020, North Carolina Department of Public Instruction, 2022).

NWEA Formative Assessments in Reading (Pretest). The Measure of Academic Progress (MAP) Reading (Northwest Evaluation Association, 2019) was used to assess students' reading comprehension (test-retest reliabilities = .79 to .86) and mathematics ability (test-retest reliabilities = .79 to .86). MAP pretests from first-grade winter were used as covariates.

Student Demographic Characteristics. We included a set of dichotomous variables for student demographic characteristics, based on administrative data in the data analyses. These data included, gender, race/ethnicity, neighborhood socioeconomic status (SES), individual education plan (IEP) status, eligibility for receipt of English-as-a-second-language instruction. For the SES variable, we used the categorization provided by the school district, which divided the student's census tracts into three groups (i.e., low, medium, high SES) based on the district's census tract characteristics.

Data analytic plan for the intent-to-treat (ITT) analysis

To conduct our ITT analyses, we specified multiple linear regression model as follows:

$$Y_{ij} = \alpha + \beta MORE_j + \Gamma X_{ij} + \phi_b + \epsilon_{ij}$$

in which Y_{ij} is the outcome for student i in school j, MORE is an indicator for being randomly assigned to MORE, ϕ is a vector of randomization blocks, X is a vector of student-level covariates, and ϵ is the error term.

Data analytic plan for treatment-on-the-treated analysis

To extend these analyses we also estimate the treatment-on-the-treated (TOT) effects, using two stage least squares (2SLS) instrumental variables estimation to examine the impact of MORE for students who participated in the full first and second grade intervention. Prior efficacy work has indicated that additional years of MORE could lead to larger impacts (Kim, Relyea et al., 2021; Kim, Burkhauser, et al., 2022). We therefore leverage random assignment as our instrument to isolate the exogenous variation in the number of years participating in more.

The first stage model is

$$YearsMORE_{ij} = \alpha + \beta(MORE)_j + \Gamma X_{ij} + \phi_b + \epsilon_{ij}$$

Where $YearsMORE$ is the potential mediator, $MORE$, the excluded instrument, is the treatment assignment indictor, $X$is a vector of student-level covariates, ϕ is a set of randomization block fixed effects, and ϵ is an error term.

The second stage model is:

$$Y_{ij} = \alpha + \beta(Year\widehat{sM}ORE)_{ij} + \Gamma X_{ij} + \phi_b + \epsilon_{ij}$$

We included block fixed effects in both the first and second stage models so the variation identifying the effect of the years with MORE on end-of-year reading scores was treatment-control comparison within blocks. We clustered standard errors at the school-level.

Results

Research Question 1: Intent-to-Treat (ITT) and Treatment-on-the-Treated (TOT) Impacts on Third Graders' Domain Specific Vocabulary Knowledge and Reading Comprehension and Domain Reading Comprehension

Table 3 reports both intent-to-treat and treatment-on-the-treated impacts at end-of-treatment and at two-month follow-up in third grade. The first two columns of Table 3 presents the impacts on domain-specific vocabulary knowledge and comprehension. For these outcomes, the sample size was attenuated by missing data from students who did not take the vocabulary knowledge and reading comprehension assessment during the pandemic school year (SY 2020-21). On these outcomes, students assigned to treatment outperformed control students on the domain-specific science vocabulary knowledge measure (ES = .13) and reading comprehension measure (ES = .14). Moving to the right panel of Table 3, columns 3 and 4 there were also positive and statistically significant impacts on domain general reading comprehension (ES = .11) and mathematics (ES = .14) outcomes. Figure 3 shows the impact on end-of-grade reading comprehension based on the covariate-adjusted average scale scores for the treatment and control groups, and Figure 4 presents similar results for math outcomes. Using typical improvement in reading from the state beginning-of-grade assessment to the end-of-grade assessment for all students in our partner site, these effects are equivalent to approximately 9.5 weeks of learning.

Research Question 2: MORE Impacts Moderated by Student and School Characteristics

Table 4 presents results from moderation analyses involving interactions between MORE treatment and student demographic characteristics (i.e., pretest first-grade reading, Black or Hispanic student, English learner, and low or middle SES neighborhoods). The results for the domain-general reading comprehension and mathematics in third grade reveal no evidence of moderation by student characteristics, suggesting that MORE was equally effective across a wide range of subgroups. Further exploratory analyses also revealed positive and statistically significant treatment impacts for students in low socioeconomic status neighborhoods for both reading comprehension (ES = .13) and math (ES = .20).

Discussion

This study analyzes the long-term impacts of a Tier I (general education, core classroom) sustained content literacy intervention by following one cohort of students from first to third grade. The students randomized to the treatment condition had the opportunity to receive exposure to MORE from first through third grade, whereas control students received MORE in third grade only. Results show that treatment group students enjoyed larger gains on domain specific measures of science vocabulary and comprehension, which were echoed on far transfer measures of domain general reading comprehension.

Overall, the magnitude of the end-of-treatment effects suggest the importance of previous exposure to the schemas that were taught in the MORE first to third grade lessons. Importantly, both treatment and control group students were given 10 lesson hours of instruction about the human body system and were thus exposed to all the semantically related words (i.e., directly taught words included system, structure, function, muscular, skeletal, nervous, diagnosis) during third grade. Despite having exposure to the same content, treatment group students demonstrated higher performance on domain specific vocabulary knowledge and reading comprehension than control students. Given that treatment students participated in the full first- to third-grade intervention, these results imply that early intervention is critical to building the schematic structures that lay the foundation for later learning in third grade and beyond.

Indeed, the most important finding from this study is the impact on end-of-third grade reading and mathematics outcomes. How might we explain both the far transfer on domain general reading comprehension and cross domain transfer to mathematics? In considering the characteristics of interventions that have long-term effects, Bailey and colleagues (2020) proposed that such interventions should target skills that are malleable, fundamental, and would not have occurred without the intervention. In terms of malleability, linguistic comprehension and the development of robust schemas are unconstrained outcomes that are sensitive to intervention efforts but cannot be improved easily through short-term interventions (Paris, 2005; Snow & Matthews, 2016). Through the study of state and common core standards, as well as leading textbooks, we identified words and concepts that are arguably fundamental to students' ability to read across subjects, particularly in science and social studies (Kim, Burkahauser et al., 2021; Kim & Relyea et al., 2021). In addition, the long-term effect for the students who received MORE in first and second grade may stem from the fact that the schemas introduced by MORE facilitated the acquisition of new knowledge.

Certainly, results from the teacher surveys clearly indicate that treatment group students had more opportunities to learn science and social studies content. The MORE teachers reported spending an average of 33 minutes/week on science and 32 minutes/week on social studies more than control teachers. Over the course of the two years when there was a treatment-control contrast, this equates to a total of 9.8 additional hours in science and 9.7 additional hours in social studies than the control students. Thus, while MORE students had greater access to these topics their exposure to ELA and math was similar to the control group.

Biosocial developmental contextualism posits that whether childhood interventions succeed depends on four factors: timing, flexibility, intensity, and environmental supports (Ramey & Ramey, 1998). The MORE intervention incorporated these four elements in designing and implementing the curriculum. First, MORE targeted a critical developmental time period, first to third grades, as students are rapidly developing critical literacy skills and acquiring foundational content knowledge (Lipsey et al., 2012). Second, for flexibility, MORE was adapted and designed to fit into the classroom and district context, such as by adapting the intervention to better align with a district's new literacy curriculum. MORE also provided teachers with discretion over when to teach MORE lessons. Third, MORE is an intensive, multi-year intervention that tightly couples learning across grades and across school and home contexts. Finally, MORE was designed to enhance environmental supports and broadly impact the literacy ecosystem to facilitate alignment throughout the district, with outreach and engagement ranging from senior district leadership to classroom teachers. Importantly, students received MORE over several years, ensuring that students' learning environments were supported during a multi-year timeframe.

MORE was theorized to have the most immediate impact on students' literacy abilities but there was also evidence of impact on math. Additional research is needed to understand cross domain transfer effects, but research suggests that similar cognitive process are involved in reading and math, such as vocabulary, nonverbal reasoning, working memory, and attention (Cirino et al., 2018; Geary, 2010; Shannon et al., 2021). In other words, the MORE concepts taught from first to third grade may be useful for children as they learn domain specific and domain general vocabulary and read and evaluate math problems that include words from Grade 1 MORE (e.g., unique, complex), Grade 2 MORE (evidence, theory, hypothesis), and Grade 3 MORE (structure, function, diagnosis). Furthermore, the practices that were sustained across grades require students to plan, organize, and execute reading and writing skills while acquiring science and social studies content knowledge. Future research should focus more specifically on the vocabulary and language demands and cognitive processes that are involved in both reading and math achievement.

Finally, MORE is a relatively low in cost when compared other interventions with similar effect sizes. We estimate a back-of-the envelope cost of the intervention of $437.09 per student (in 2021 dollars) including personnel, books, professional services & contracts (e.g., cost to create the app and corresponding database), computer & technology (e.g., iPads), travel (e.g., site visits, a convening for participants), and other miscellaneous expenses (e.g., equipment, printing, shipping). In addition, as we have made improvements over the years (i.e., the MORE app is web-based and can work on any computer now), we expect implementation costs will be lower, about $219.94 per student, but need to ensure these changes do not alter the effectiveness. In comparison, a cost analysis of Project READS, a summer book program that provided free books and paper activities related to the book for elementary school students, found that program expenses were relatively low (between $250-$480 per student) for the observed ITT effects of 0.04 standard deviations (Kim et al., 2016). Similarly, Reading Partners, a program that uses volunteers to provide one-on-one tutoring to struggling readers in elementary school, found effects of 0.10 standard deviations for costs between $480-$1,270 per student (Jacob et al., 2015). Thus, MORE could be a cost-effective literacy solution and future work will work quantify its elements more precisely.

Limitations and Next Steps

In its exploration of mechanisms, this study is limited in that it is based on teacher self-reported data, and not a direct measure of students' classroom experience. In addition, we only have data on instructional topics for one year. Thus, we are limited in the ability to paint a comprehensive picture of the treatment students' school experience, and how it may differ from the control group.

This study is also limited in that it occurred in one large school district. The schools in our study are comprised of a racially, ethnically, and socioeconomically diverse sample of students, which is provisionally suggestive of the intervention's potential external validity outside the district. However, to gain further insights on its external validity, it will be important to study MORE in other districts and settings, to see whether treatment effects remain on a larger sample.

Finally, we did not explore examine whether gains on domain specific vocabulary knowledge mediate the impact of MORE on transfer outcomes measured at the end of third grade. Thus, it is critical to consider whether and to what extent improvements in vocabulary knowledge are key mechanisms through which MORE improves students' domain specific and domain general reading comprehension as well as mathematics outcomes. Despite these limitations, findings from this study indicate that a sustained content literacy intervention may be an evidence-based and scalable approach for improving the academic achievement outcomes of all learners.

References

Alexander, P. A. (2003). The development of expertise: The journey from acclimation to proficiency. *Educational researcher*, *32*(8), 10-14.

Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2007). Lasting consequences of the summer learning gap. *American sociological review*, *72*(2), 167-180.

Anderson, R. C. (1984). Some reflections on the acquisition of knowledge. *Educational researcher*, *13*(9), 5-10.

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

Arold, B. W., & Shakeel, M. (2021). The Unintended Effects of the Common Core State Standards on Non-Targeted Subjects. *The Unintended Effects of the Common Core State Standards on Non-Targeted Subjects (June 4, 2021)*.

Bailey, D. H., Duncan, G. J., Cunha, F., Foorman, B. R., & Yeager, D. S. (2020). Persistence and fade-out of educational-intervention effects: Mechanisms and potential solutions. *Psychological Science in the Public Interest, 21*(2), 55–97. https://doi.org/10.1177/1529100620915848

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.

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

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

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

Catts, H. W., Hogan, T. P., & Adlof, S. M. (2005). Developmental changes in reading and reading disabilities. In Catts, H. W. and Kamhi, A. G. (Eds.), *The connections between language and reading disabilities* (pp. 38-51). Psychology Press.

Cervetti, G. N., & Wright, T. S. (2020). The role of knowledge in understanding and learning from text. In *Handbook of reading research, volume V* (pp. 237-260). Routledge.

Cirino, P. T., Child, A. E., & Macdonald, K. T. (2018). Longitudinal predictors of the overlap between reading and math skills. *Contemporary Educational Psychology, 54*, 99–111.

Connor, C. M., Dombek, J., Crowe, E. C., Spencer, M., Tighe, E. L., Coffinger, S., ... & Petscher, Y. (2017). Acquiring science and social studies knowledge in kindergarten through fourth grade: Conceptualization, design, implementation, and efficacy testing of content-area literacy instruction (CALI). Journal of Educational Psychology, 109(3), 301-320

Cromley, J. G., & Azevedo, R. (2007). Testing and refining the direct and inferential mediation model of reading comprehension. *Journal of educational psychology*, *99*(2), 311.

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

Fitzgerald, J., Elmore, J., Relyea, J. E., & Stenner, A. J. (2019, August 22). Domain-Specific Academic Vocabulary Network Development in Elementary Grades Core Disciplinary Textbooks. *Journal of Educational Psychology*. Advance online publication. http://dx.doi.org/10.1037/edu0000386

Fryer Jr, R. G., & Levitt, S. D. (2006). The black-white test score gap through third grade. *American law and economics review*, *8*(2), 249-281.

García, J. R., & Cain, K. (2014). Decoding and reading comprehension: A meta-analysis to identify which reader and assessment characteristics influence the strength of the relationship in English. *Review of educational research*, *84*(1), 74-111.

Geary, D. C. (2010). Mathematical disabilities: Reflections on cognitive, neuropsychological, and genetic components. *Learning and Individual Differences, 20,* 130–133.

Goldhaber, D., Kane, T., McEachin, A., Morton E., Patterson, T., Staiger, D., (2022) The Consequences of Remote and Hybrid Instruction During the Pandemic. Research Report. Cambridge, MA: Center for Education Policy Research, Harvard University

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

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

Hernandez, D. J. (2011). Double jeopardy: How third grade reading skills and poverty influence high school graduation. Annie E. Casey Foundation.

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

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

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: 1935-1964. https://doi.org/10.1007/s10648-021-09609-6

Kim, J. S., Burkhauser, M. A., Relyea, J. E., Gilbert, J. B., Scherer, E., Fitzgerald, J., Mosher, D. & McIntyre, J. (2022). Longitudinal Randomized Trial of a Sustained Content Literacy Intervention from First to Second Grade: Transfer Effects on Students' Reading Comprehension. *Journal of Educational Psychology*. Advanced Online. https://DOI: 10.1037/edu0000751

Kintsch, W. (1988). The role of knowledge in discourse comprehension: a construction-integration model. *Psychological review*, *95*(2), 163.

Kintsch, W. (2009). Learning and constructivism. In *Constructivist instruction* (pp. 235-253). Routledge.

Lewis, K., Kuhfeld, M., Ruzek, E., & McEachin, A. (2021). Learning during COVID-19: Reading and math achievement in the 2021-21 school year. NWEA Research.

Lipsey, M.W., Puzio, K., Yun, C., Hebert, M.A., Steinka-Fry, K., Cole, M.W., Roberts, M., Anthony, K.S., Busick, M.D. (2012). Translating the Statistical Representation of the Effects of Education Interventions into More Readily Interpretable Forms. (NCSER 2013-3000). Washington, DC: National Center for Special Education Research, Institute of Education Sciences, U.S. Department of Education. This report is available on the IES website at http://ies.ed.gov/ncser/

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

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.

North Carolina Department of Public Instruction (2022). The North Carolina Department of Public Instruction Grades 3-8 Reading End of Grade (EOG) High School English II End of Course (EOC) Edition 5 Technical Report 2020-21. Accessed at https://www.dpi.nc.gov/media/14305/open.

North Carolina Department of Public Instruction (2020). The North Carolina Department of Public Instruction Mathematics 3-8 End of Grade (EOG) NC Math 1 and NC Math 3 End of Course (EOC) Technical Report 2018-2019. Accessed at https://www.dpi.nc.gov/media/10219/open

NWEA. (2019). MAP® Growth™ technical report. Portland, OR: Author.

Paris, S. G. (2005). Reinterpreting the development of reading skills. Reading Research Quarterly, 40(2), 184–202. https://doi.org/10.1598/RRQ.40.2.3

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

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

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

Ramey & Ramey, 1998. Early Intervention and Early Experience. *American Psychologist, 53*, 2, 109-120.

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

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

Shannon, K. A., Scerif, G., & Raver, C. C. (2021). Using a multidimensional model of attention to predict low-income preschoolers' early academic skills across time. Developmental Science, 24(2), e13025. https://doi.org/10.1111/desc.13025

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

Snow, C. E., Matthews, T. J. (2016). Reading and language in the early grades. The Future of Children, 26(2), 57–74. https://doi.org/10.1353/foc.2016.0012

Vellutino, F. R., Tunmer, W. E., Jaccard, J. J., & Chen, R. (2007). Components of reading ability: Multivariate evidence for a convergent skills model of reading development. *Scientific Studies of Reading*, *11*(1), 3-32.

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

Table 1

Balance Checks for Analytic Sample of Students Remaining in the Long-Term Impact Analysis

	Treatment		Control		
	N	Mean (SD)	N	Mean (SD)	Difference
White	1147	.14 (.35)	891	.19 (.39)	-.064 (.054)
Black	1147	.38 (.49)	891	.35 (.48)	.027 (.059)
Hispanic	1147	.39 (.49)	891	.34 (.47)	.054 (.051)
Asian	1147	.07 (.25)	891	.08 (.28)	-.016 (.025)
Other	1147	.03 (.17)	891	.03 (.17)	-.001 (.008)
Male	1147	.52 (.5)	891	.51 (.5)	.002 (.013)
English learner	1147	.27 (.44)	891	.24 (.43)	.034 (.04)
IEP	1147	.07 (.26)	891	.10 (.29)	-.022 (.011)*
Low SES	1147	.51 (.5)	891	.35 (.48)	.184 (.086)*
Mid SES	1147	.31 (.46)	891	.43 (.5)	-.13 (.109)
High SES	1147	.17 (.38)	891	.21 (.41)	-.054 (.098)
Baseline MAP Reading	1123	167.02 (15.67)	874	171.66 (15.96)	-5.01 (1.579)**
Baseline MAP Math	1122	168.35 (16.67)	869	172.88 (17.26)	-4.865 (1.643)**

Note. Differences derived from regression model with treatment indicator, fixed effects for randomization block, and standard errors clustered at the school level. IEP = child has an individualized education plan. MAP = NWEA Measure of Academic Progress at baseline (Grade 1). SES = neighborhood socioeconomic status

*p < .05, **p < .01, ***p < .001

Table 2
Weekly Exposure to Courses (Minutes/week) from 2019 – 2021, by Treatment Status

Subject	Treatment *Mean (SD)*	Control *Mean (SD)*	t	df	ES	Raw Diff.
English language arts (ELA)/reading	552.39 (210.62)	573.39 (253.81)	-1.66†	295	-.09	-21
Science	145.16 (120.34)	112.48 (112.45)	2.58**	294	.28	33
Social Studies	138.93 (116.93)	106.67 (109.91)	2.57*	294	.28	32
Math	441.78 (156.1)	439.05 (165.79)	.25	153	.02	3

Notes: df = Degrees of freedom. These have been adjusted to account for the clustering of teachers within schools.
†p < .10, *p < .05, **p < .01, ***p < .001

Table 3

Intent-to-Treat and Treated-on-the-Treated Impact on Third Graders' EOG Reading Comprehension

	End of Treatment (G3) Domain Specific Reading		Two Month Follow-up Domain General Reading and Math	
	Content Comprehension	Vocabulary	EOG Reading	EOG Math
ITT	.14 (.066)*	.125 (.056)*	.108 (.036)**	.144 (.047)**
2SLS	.145 (.065)*	.13 (.056)*	.118 (.039)**	.157 (.049)**
N Students	1277	1277	2049	1985
2SLS 1st Stage F	21436	21436	348.07	329.07

Note. ITT and 2SLS models include controls for demographics, strata, cubic baseline reading & math scores. Errors clustered at school-level, the unit of randomization. Years of MORE is instrumented by initial randomization assignment. EOG = End of Grade.

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

Table 4

Student Moderators of ITT Impacts on Third Graders' End of Grade Reading Comprehension and Math Outcomes

	EOG Reading - 3rd grade				EOG Math - 3rd grade			
	Pretest Reading	Black or Hispanic	EL	Low or Mid SES	Pretest Reading	Black or Hispanic	EL	Low or Mid SES
MORE	.111 (.037)**	.189 (.07)**	.097 (.043)*	.208 (.111)+	.145 (.047)**	.14 (.066)*	.125 (.048)**	.213 (.081)**
Interaction	.005 (.027)	-.109 (.074)	.059 (.057)	-.123 (.116)	0 (.029)	.007 (.07)	.083 (.05)+	-.087 (.101)
N	2049	2049	2049	2049	1985	1985	1985	1985

Note: Models include controls for demographics, randomization blocks, and pretest scores in reading (linear term only) and math (cubic polynomial). Standard errors clustered at school-level, the unit of randomization in parenthesis. EL = English learner. SES = neighborhood socioeconomic status.

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

Figure 1

Theory of Change Underlying a Sustained Content Literacy Intervention

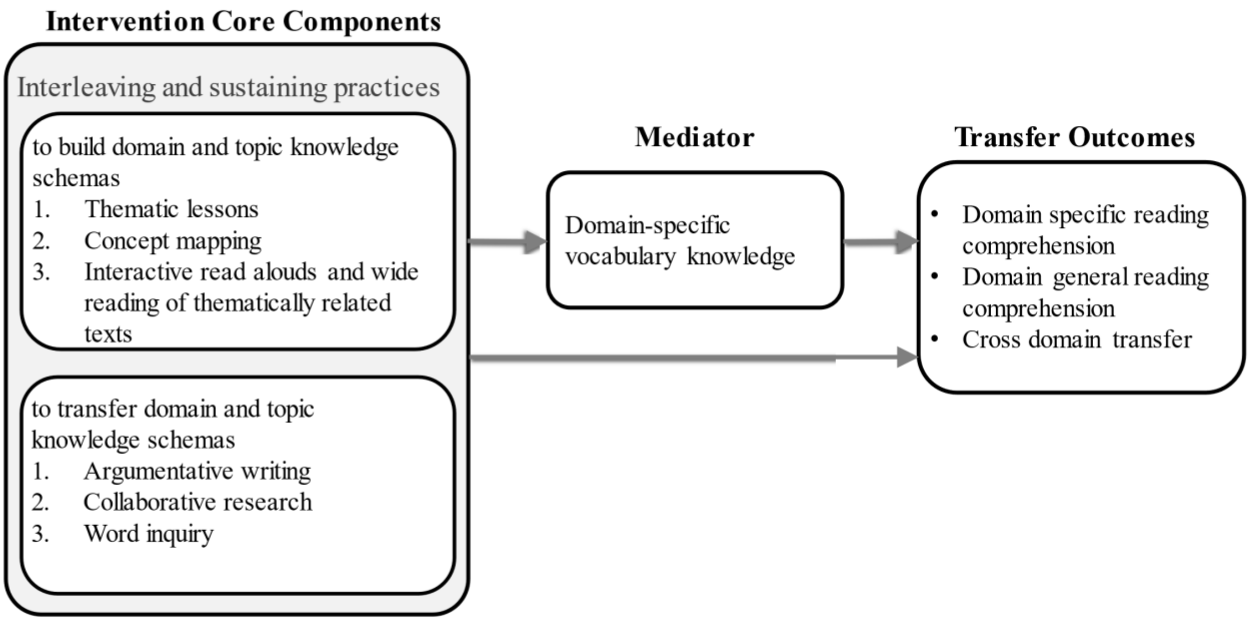

Figure 2

Visual Displaying Study Design for Implementation and Intervention Activities and Evaluation of Long-Term Impact on End-of-Treatment and Two-Month Follow-Up Grade 3 Outcomes

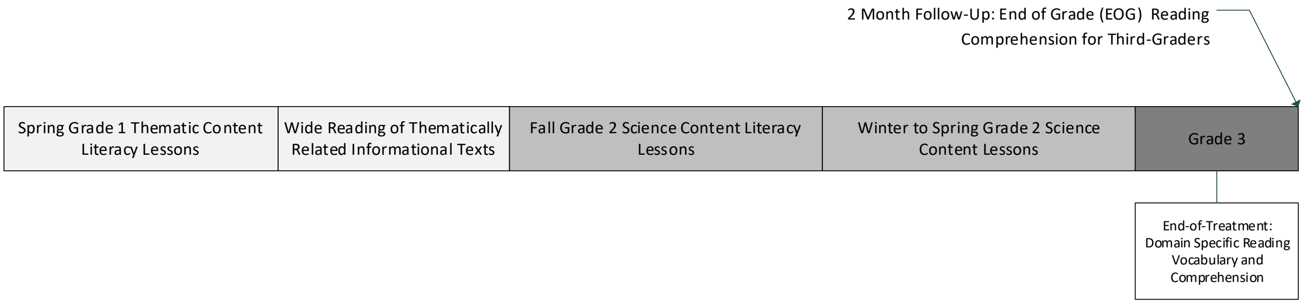

Figure 3

ITT Impacts on Third Graders' EOG Reading Comprehension Scaled Scores by Treatment Status

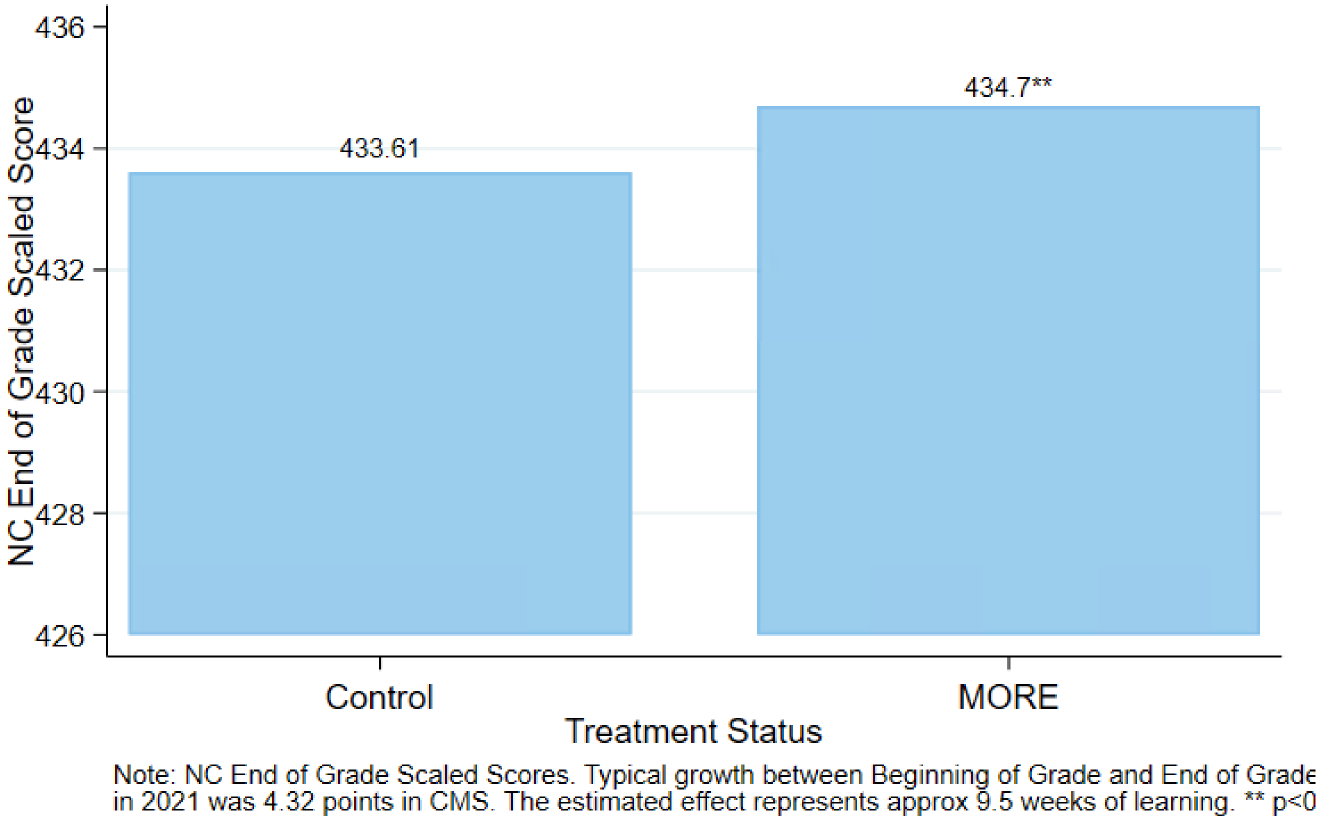

Figure 4

ITT Impacts on Third Graders' EOG Math Scaled Scores by Treatment Status