NOTE: THIS DRAFT WAS SHARED ON 4/18/24 AND IS CURRENTLY IN THE PEER-REVIEW PROCESS AND IS THEREFORE NOT THE FINAL VERSION. PLEASE REACH OUT TO THE FIRST AUTHOR FOR MORE INFORMATION.

Impact of a Content-Rich Literacy Curriculum on Kindergarteners' Vocabulary, Listening Comprehension, and Content Knowledge

Sonia Q. Cabell
https://orcid.org/0000-0002-2586-5735
Florida State University, Florida Center for Reading Research
James S. Kim
https://orcid.org/0000-0002-6415-5496
Harvard University, Graduate School of Education
Thomas G. White
https://orcid.org/0000-0002-0569-473X
University of Virginia, School of Education and Human Development
Charles J. Gale
Harvard University, Graduate School of Education
Ashley A. Edwards
https://orcid.org/0000-0002-5268-4235
Florida State University, Florida Center for Reading Research
HyeJin Hwang
https://orcid.org/0000-0002-1049-7941
University of Minnesota, Department of Educational Psychology at College of Education and Human Development
Yaacov Petscher
https://orcid.org/000-0001-8858-3498
Florida State University, Florida Center for Reading Research
Rhonda M. Raines
https://orcid.org/000-0003-0788-1501
Florida State University, Florida Center for Reading Research

Author Note

We would like to thank the district leaders, teachers, parents, and children who made this work possible. We would also like to thank Marcy Wyatt and Ansley Kramer, who worked tirelessly to manage the project activities and data. The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A170635 (PI: Cabell), and by the Office of Special Education Programs, U.S. Department of Education, through Grant #H325D190037, awarded to Florida State University. The opinions expressed are those of the authors and do not represent views of the Institute, Office, or the U.S. Department of Education. We have no known conflicts of interest to disclose. Study 2 was preregistered prior to data analyses in the Registry of Efficacy and Effectiveness Studies (ID # 1812.1v1). In addition, raw data from these studies are available in the LDBase repository.
Correspondence concerning this article should be addressed to Sonia Cabell, Florida State University, Florida Center for Reading Research, 2010 Levy Ave Suite 100, Tallahassee, FL 32310. Email: scabell@fsu.edu

Abstract

This study examined the impact of a widely used content-rich literacy curriculum on kindergarteners' vocabulary, listening comprehension, and content knowledge. In combined findings from two randomized controlled trials, the second being a replication of the first, 47 schools in large urban U.S. districts were randomly assigned to implement the Core Knowledge Language Arts: Knowledge Strand kindergarten curriculum (CKLA: Knowledge) or to a waitlist control condition. CKLA: Knowledge focuses instruction on language comprehension through interactive read-alouds that systematically build content knowledge. Teachers received two days of professional development workshops, along with light-touch support from facilitators during implementation. Participants included 1,194 kindergarten students, who were administered individual pre- and posttest measures of proximal and standardized oral language and content knowledge (i.e., science, social studies). After approximately one semester of curricular implementation, CKLA: Knowledge demonstrated a positive and significant impact on vocabulary, both for words taught in the curriculum and for general expressive vocabulary. In addition, there was a main impact on children's proximal science and social studies knowledge. Significant interactions were found for proximal measures of vocabulary and knowledge, such that children who began the year with relatively higher receptive vocabulary scores derived a greater benefit of learning the words and content knowledge taught in the curriculum. The present work is unique in that it tests the effects of a content-rich literacy curriculum, integrating oral language and content-area instruction, in driving far transfer effects on standardized measures of oral language and science and social studies knowledge in primary grade children.

Impact and Implications Statement

This study tested the impact of a widely used content-rich literacy curriculum, Core Knowledge Language Arts: Knowledge Strand, on kindergarteners' vocabulary, listening comprehension, and content knowledge. After one semester of instruction using the curriculum, students demonstrated significantly higher vocabulary and content knowledge scores than students who did not have instruction using the curriculum, including an effect on a far transfer measure of generalized vocabulary. Findings provide further evidence that approaches that integrate content instruction with oral language instruction (i.e., content-rich literacy approaches) are powerful for student learning.

Keywords: content-rich literacy; content literacy; vocabulary; kindergarten; language comprehension

Impact of a Content-Rich Literacy Curriculum on Kindergarteners' Vocabulary, Listening Comprehension, and Content Knowledge

Over the past two decades, public elementary school education in the United States has focused attention on improving the reading achievement of students in the primary grades (K-2; No Child Left Behind, 2002; National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010). Despite the time and effort spent on reading instruction, reading comprehension scores among fourth graders, including those in underrepresented groups and those living in poverty, have remained consistently low in national assessments for years, and the COVID-19 pandemic exacerbated inequality in reading performance (National Center Education Statistics, 2022). Specifically, between 2020 and 2022, the National Assessment of Education Progress (NAEP) results indicated that not only did reading scores among 9-year-olds experience the largest decline in decades, students performing at the 10th percentile had five times the amount of score decrease as those at the 90th percentile.

Given the alarming NAEP trends, it is important to test malleable instructional levers for improving children's reading comprehension. Substantial theoretical and empirical research documents robust links between oral language and knowledge development during the early grades and later reading comprehension (e.g., Grissmer et al., 2023; Language and Reading Research Consortium [LARRC] et al., 2019; Cromley & Azelvedo, 2007). Language and knowledge take time to develop (Neuman, 2006), and experts are increasingly recommending attention to both areas early in students' academic careers (e.g., Cervetti & Hiebert, 2019). Thus, a critical question is whether a stronger, early emphasis on oral language development and content knowledge acquisition can be implemented in a Tier 1 (i.e., universal instruction in general education) whole class setting, benefitting all children, including students with weaker

oral language skills at school entry, students acquiring English language proficiency, and students from high poverty backgrounds.

There has been increasing interest in U.S. elementary schools to offer content-rich literacy instruction for all learners as a novel approach for improving comprehension. Content-rich literacy instruction deliberately infuses content area instruction into English Language Arts instruction to simultaneously boost children's literacy skills (particularly oral language) and content knowledge needed for later success in reading comprehension (Cabell & Hwang, 2020). *Core Knowledge Language Arts: Knowledge Strand* (CKLA: Knowledge; Core Knowledge Foundation & Amplify Education, 2017) is one such widely used program that systematically builds children's oral language and content knowledge in the domains of science and social studies through interactive read-alouds. This article presents the combined findings of two randomized controlled trials (RCTs)—the second being a conceptual replication of the first in a different district context—that examined the effects of CKLA: Knowledge on kindergarteners' vocabulary, listening comprehension, and content knowledge across two large urban districts in the United States. To our knowledge, the present work is unique in that it tests the effects of a content-rich literacy curriculum, that includes both science and social studies knowledge, in driving far transfer effects on standardized measures of oral language and knowledge in primary grade children.

Improving Language Comprehension by Integrating Oral Language and Content Knowledge Instruction

Targeting Oral Language Skills

Language comprehension is the "ability to extract and construct literal and inferred meaning from linguistic discourse represented in speech" (Hoover & Tunmer, 2022). Thus, a key

component of language comprehension includes developing oral language (e.g., receptive and expressive vocabulary, listening comprehension) during the early childhood years, which, coupled with efficient word reading skills, is critical for proficiency in later reading comprehension (e.g., Hjetland et al., 2020; Kendeou et al., 2009; Language and Reading Research Consortium [LARRC], 2015; Storch & Whitehurst, 2002). Yet, research efforts to improve young children's oral language skills have demonstrated mixed results. Although there are consistent positive effects on children's proximal language skills, such as learning the words taught in a given intervention, relatively few studies report statistically significant impacts on generalized, standardized measures of oral language (e.g., Coyne et al., 2022; Dickinson et al., 2011; Marulis & Neuman, 2010; Silverman et al., 2020).

As part of the 120M Reading for Understanding (RfU) initiative, the LARRC research team conducted a cross-sectional pre-kindergarten and kindergarten randomized trial on the Tier 1, whole class *Let's Know* supplemental curriculum, which targeted both vocabulary and listening comprehension skills. Although *Let's Know* did incorporate content area topics, the primary focus of instruction was on improving oral language skills (rather than building content knowledge). They reported an immediate and sustained effect on words taught in the curriculum, along with other curriculum-aligned comprehension measures. Yet there were no statistically significant impacts on standardized measures of listening comprehension (including the Test of Narrative Language; LARRC et al., 2022). Moreover, improvement on proximal vocabulary did not mediate effects on these outcomes. These findings are in keeping with a recent meta-analysis that found that targeted vocabulary instruction doesn't typically transfer to generalized vocabulary knowledge (Cervetti et al., 2023). The LARRC team encouraged future research which "might also test whether a broader range of language and cognitive skills should be

targeted to achieve robust transfer effects" (p. 1221). Thus, the most up-to-date research underscores the need for rethinking the theory of change underlying Tier 1 whole class language-focused interventions.

Integrating Oral Language and Content Knowledge

Language comprehension includes not only oral language skills but also background knowledge (Scarborough, 2001). One novel, though largely untested, approach to developing children's language comprehension is to integrate instruction focused on building oral language and content knowledge. We define *content-rich literacy instruction* as "instruction in which literacy activities (reading and/or writing) serve as a tool to cultivate content knowledge (science and/ or social studies) while, at the same time, content teaching serves as a lever to facilitate [growth in] literacy skills (vocabulary and/or comprehension)" within the English Language Arts instructional setting (Cabell & Hwang, 2023; Hwang et al., 2022, p. 224). Inarguably, *content knowledge* in the domains of science and social studies (i.e., knowledge about the natural and social world) makes an important contribution to comprehension (Cabell & Hwang, 2020; Cervetti & Wright, 2020; Grissmer et al., 2010; Hwang & Duke, 2020). Indeed, comprehension is heavily dependent on the knowledge that a reader or listener brings to a text (Anderson & Pearson, 1984; Willingham, 2006). When approaching a text, the reader or listener constructs a situation model through integrating prior knowledge with the textbase, the propositions found in the text itself (Kintsch, 1988, 2005; Zwaan & Radvansky, 1998). The situation model helps the reader or listener to fill in gaps not directly stated in the text, make inferences, and connect ideas across sentences to create an "enriched level of comprehension beyond the literal meaning of a text" (Stafura & Perfetti, 2017, p.9). It follows that if one lacks content knowledge on the topic

of the text, the ability to create a coherent mental model, and thus more fully understand and learn from the text, is hindered.

Integrating oral language and content knowledge instruction has the potential to yield valuable benefits for children's language comprehension. Content-rich literacy instruction in the primary grades invites teachers to systematically build science and/or social studies knowledge through sets of readings and discussions (e.g., interactive read-alouds) focused on specific topics, while repeatedly exposing children to semantically related words and ideas. This type of instruction that integrates language and knowledge has potential affordances that may positively impact children's language comprehension. The notion of preferential attachment further suggests that as networks grow in size and density, they become stickier, making it easier for children to remember new words and to situate them in a semantic network (Steyvers & Tenenbaum, 2005). Such networks can be leveraged as children build mental representations of texts as they listen to texts, discuss, and write about science and social studies topics.

Change Model for CKLA: Knowledge

The change model shown in Figure 1 shows how theoretically informed intervention components of CKLA: Knowledge drive the underlying processes that improve kindergarten outcomes in vocabulary, listening comprehension, and content knowledge. The content-rich literacy instruction that is the focus of the present study has two main features that serve as underlying processes of the hypothesized intervention effect. These processes are also reflected in the small pool of K-2 content-rich literacy interventions that have demonstrated evidence for efficacy on proximal and/or standardized vocabulary or comprehension outcomes (Connor et al., 2017; Kim, Burkhauser, et al., 2021; Kim, Relyea, et al., 2021; Neuman & Kaefer, 2018; Vitale & Romance, 2012).

First, primary grade content-rich literacy instruction involves systematically building children's content knowledge through interacting with texts. It employs teachers reading aloud texts, discussion, vocabulary, and writing with the goal of improving not only vocabulary and comprehension, but also science and social studies knowledge (Hwang et al., 2021). Specifically, instruction is logically organized around science and/or social studies content topics, with knowledge building serving as a chief aim of instruction. Importantly, knowledge organized within and across grades in a connected fashion can help to develop children's schemata about specific topics, which can lead to not only better understanding the given topic but also transferring knowledge to other related topics (Kim & Burkhauser, 2022; Kim, Burkhauser, et al., 2023). Conceptually coherent text sets are used for teachers to interactively read aloud (or in some cases for student reading), with multiple texts designed to deepen content knowledge over successive readings, with vocabulary words and concepts that repeat across texts (Cervetti et al., 2016; Wright et al., 2022). A range of genres can be used including informational, mixed genre, and narrative texts (Neuman & Kaefer, 2018). Content-rich literacy instruction also features discussion before, during, and after reading of a text that supports knowledge building while actively promoting students' participation through the use of both literal and inferential questioning with opportunities for children to discuss their ideas, practices that have been found to be important for students' oral language development in other research (e.g., Brabham & Lynch-Brown, 2002; Dickinson & Smith, 1994; Feitelson et al., 1993; Morrow, 1992; Santoro et al., 2007; van Kleeck et al., 2003). Importantly, these interactive read alouds encourage topic-focused teacher-child conversational interactions that could improve both teacher talk and child language use in the classroom (Cabell et al., 2013; Kook & Greenfield, 2021; Wright & Gotwals,

2017). Students are also expected to engage in writing in response to texts in ways that support knowledge building.

Relatedly, the second underlying process of content-rich literacy instruction implicated in our theory of change (Figure 1) is strengthening semantic networks (or knowledge networks), which can facilitate listening or reading comprehension by increasing depth of word knowledge and rapid retrieval of word meanings (Beck, Perfetti, & McKeown, 1982; Ouellette, 2006). Building content knowledge through the use of coherent text sets can facilitate incidental word learning through the generation of accurate inferences about unknown words (Cervetti et al., 2016; Kaefer et al., 2015; Pulido, 2003, 2007). For example, a person who is knowledgeable about dinosaurs may understand the meaning of related words, such as *omnivore*, *carnivore*, and *herbivore*. In addition, research has demonstrated that explicit teaching of vocabulary words, including the relationships among words, fosters children's vocabulary development (e.g., Neuman & Kaefer, 2018; Neuman et al., 2011, 2021).

Content-rich literacy instruction contrasts with traditional primary grade literacy instruction, which (in addition to phonics instruction) is generally organized and driven by instruction focused on vocabulary and listening comprehension through reading aloud to students, with limited attention to the ways words and ideas are connected through content knowledge (Paige et al., 2021). When science and social studies topics are incorporated in traditional literacy instruction, even using material deemed high quality (e.g., EdReports, 2023), content is not a driving feature and, therefore, texts and vocabulary aren't necessarily selected to deepen content knowledge; discussion and writing do not consistently reinforce content knowledge; and knowledge is activated rather than being systematically developed.

Evidence for Content-Rich Literacy Instruction in the Primary Grades

Extant (quasi) experimental research suggests the promise of content-rich literacy instruction in elementary grade settings (Hwang et al., 2022; 2023), though the work in the primary grades (K-2) is quite limited (Cabell & Hwang, 2020). In a meta-analysis of integrated literacy and content instructional K-5 approaches more broadly (n = 35 studies), Hwang et al. (2022) reported statistically significant positive effects on proximal vocabulary outcomes (ES = 0.86) and on both proximal and standardized comprehension outcomes (ES=0.54 and 0.25, respectively). A supplementary analysis also demonstrated an effect on proximal, researcher-developed content knowledge outcomes (ES= 0.94). When focusing on the subset of content-rich literacy approaches in K-5 (i.e., taught in the English Language Arts setting), Hwang et al. (2023) identified only five studies that met the What Works Clearinghouse (WWC) standards for causal validity (What Works Clearinghouse, 2017). The result of this more focused meta-analysis indicated an impact on proximal vocabulary measures (*g*=0.29) and overall comprehension measures (*g*= 0.24) [though statistically non-significant for proximal or standardized measures when measured separately likely due to lack of statistical power]. Across both meta-analyses, effects on listening comprehension could not be isolated from effects on reading comprehension due to lack of detail in individual study reporting. There was also a lack of impact on standardized vocabulary and content knowledge measures, with the caveat that there are only a handful of studies that used distal measures of these outcomes.

Although these findings suggest a promising effect on vocabulary, comprehension, and content knowledge, there are a number of limitations in the extant literature. First, there are very few studies examining content-rich literacy instructional approaches in K-2 settings that meet a high standard of rigor (Connor et al., 2017; Gray et al., 2022; Kim, Burkhauser, et al., 2021; Kim, Relyea, et al., 2021; Neuman & Kaefer, 2018; Neuman et al., 2021; Vitale & Romance,

2012). Even fewer of these are Tier 1 whole-class approaches replacing [rather than supplementing] literacy instruction (Gray et al., 2022; Kim, Burkhauser, et al., 2021; Kim, Relyea, et al., 2021), with only a single study featuring a commercially available, widely used curriculum (Gray et al., 2022). Only one study tested an intervention longer than 10 weeks (Gray et al., 2022), and only two included the integration of instruction in oral language with both science and social studies domains (Connor et al., 2017; Kim, Relyea, et al., 2021). No studies to our knowledge have examined the impact of content-rich literacy instruction in K-2 settings on both proximal and standardized outcomes of oral language and content knowledge.

The present study is a unique contribution to the extant literature in that it examined the effects of a content-rich literacy intervention in both science and social studies domains in a Tier 1 whole-class setting as a lever for improving proximal and standardized kindergarten oral language and content knowledge outcomes. The most closely related study is a cross-sectional RCT of an RfU-funded K-4 intervention, in which Connor et al. (2017) tested the supplemental Content-Area Literacy Instruction (CALI) program that integrated science and social studies knowledge-building instruction with literacy instruction for 12 weeks in daily small group lessons led by teachers hired by the research team. Although the intervention had a statistically significant, positive effect on researcher-designed measures of proximal science and social studies knowledge in K-2 and on standardized effects for older students (4th grade vocabulary and listening comprehension), there were no significant effects noted on standardized measures of vocabulary or listening comprehension in the K-2 grade band. Although, in the CALI study, researchers were focused on impacting content knowledge without reducing the impact on reading comprehension, the lack of significant effects on standardized language outcomes underscores the need for additional work in this area. In our view, content-rich literacy

approaches not only add content teaching without detracting from literacy learning (Williams et al., 2009) but can serve to enhance both content knowledge and literacy outcomes (Hwang et al., 2022). In addition, the CALI study did not systematically build knowledge through developing underlying schemata. Schema development is important in that it is more likely to lead to broad general comprehension gains over time, as children need broad exposure to science and social studies content to understand the broad range of topics they will encounter in the upper grades and beyond (Kim, Gilbert, et al., 2023).

Potential Moderators of the Intervention Effect

Whole class content-rich literacy curricula like CKLA: Knowledge are designed to support the learning of all children. However, there is substantial research indicating the potential moderating role of student characteristics. First, research examining child-by-instruction interactions suggests that initial skills may moderate effects (Cronbach & Snow, 1977; Connor et al., 2012). Findings are mixed from studies that explore the moderating effects of initial language skills on language outcomes, with some studies reporting that children with weaker initial language skills benefit most from oral language interventions (e.g., Justice, Meier, & Walpole, 2005), whereas other studies report that children with stronger initial language skills benefit most from oral language interventions (Cabell et al., 2011; Coyne et al., 2009). There is much less information about child-by-instruction interactions for content learning. In a longitudinal correlational study of second and third grade students, Connor et al. (2012) found that the relation between science activities and gains in science knowledge were moderated by initial science knowledge, such that students entering with stronger knowledge appeared to benefit from the activities whereas those entering with the weakest knowledge did not. Thus, in the present study, we probed both main effects and interaction effects.

In addition, other child characteristics, including English Language Learner (ELL) and poverty status, may moderate the association between the treatment and children's outcomes (e.g., Vadasy & Sanders, 2012). Research on academic language interventions in the upper elementary and middle grades consistently suggest stronger gains for ELLs versus monolingual English speakers (Truckenmiller et al., 2019). Neuman and Kaefer (2018) likewise indicated that preschoolers and kindergartners who were ELL students benefited more from a content-rich literacy intervention than native English speakers. There is a need to continue to examine moderation by language proficiency in kindergarten. It is also important to probe whether the effect of the intervention varies by poverty status. Proponents of content-rich literacy interventions argue that building knowledge can level the playing field for children living in poverty, who may have fewer content-related experiences upon which to draw (e.g., Hirsch, 2003).

Study Purpose

In the present studies, we tested the efficacy of the widely used CKLA: Knowledge Strand in kindergarten classrooms with two RCTs that took place in two urban U.S. districts. Because these studies were identical in procedure and intervention, we report the combined results of these RCTs to increase precision and statistical power to detect effects (Lortie-Forgues & Inglis, 2019). Separate Study 1 and Study 2 results are reported in online supplemental materials (see Appendices A and B, respectively).

Our confirmatory question was: What is the intent-to-treat effect of CKLA: Knowledge on kindergarteners' oral language skills (i.e., vocabulary and listening comprehension) and content knowledge (i.e., science and social studies)? Our exploratory research question was: Is

the impact moderated by students' initial skills at kindergarten entry, by English Learner status, or by economic disadvantage (data available for Study 2 only)?

We hypothesized that there would be statistically significant positive effects on kindergarteners' language and knowledge, particularly for proximal outcomes in keeping with the extant literature (e.g., Hwang et al., 2022; Silverman et al., 2020). We made no firm hypotheses about the exploratory research question, given the mixed findings in this area.

Method

Transparency and Openness

This study began data collection in 2017, in the midst of the recent movement toward open science in education (e.g., IES Standards for Excellence in Education Principles introduced in 2019). While we did not preregister the study design nor hypotheses, the analysis plan for Study 2 was preregistered prior to data analyses in the Registry of Efficacy and Effectiveness Studies (ID # 1812.1v1). This manuscript is the primary reporting for Research Questions 1-3 specified in the registry. Due to the disruption in schools caused by the COVID-19 pandemic, we were unable to collect longitudinal data as we had planned in first and second grades to address the remaining research questions in the registry, including growth over time. Raw data from larger project are available in the LDBase repository (LINK MASKED FOR BLIND REVIEW). Analysis code for this study can be obtained by emailing the first author. We report how we determined our sample size, all data exclusions, all manipulations, and all outcome measures in the study (Simmons et al., 2012).

Research Design

In this cluster randomized controlled trial (RCT), designed to meet What Works Clearinghouse Standards 4.0 (2017) without reservation, 47 schools within two large, urban

districts in the Mid-Atlantic and South were randomly assigned per study, block randomized by school size, to one of two conditions: (1) CKLA: Knowledge (n = 23; n = 11 in Study 1, n = 12 in Study 2), or (2) business-as-usual control (n = 24; n = 12 in Study 1, n = 12 in Study 2). Random assignment was conducted by the second and fourth authors after the collection of baseline student assessment data (November 2017 for Study 1, November 2018 for Study 2), and a deliberate firewall between assignment/analyses and implementation was maintained such that these authors were not directly involved in the recruitment of sites or RCT implementation. Likewise, there was a firewall between the research team and: (1) the developer (Core Knowledge Foundation) and (2) the publisher (Amplify Education) of CKLA: Knowledge. Specifically, the developer Core Knowledge Foundation provided recruitment presentations, the professional development workshops for teachers, training for research facilitators who provided light coaching support to teachers, consultation on units to include in the reduced kindergarten implementation along with a pacing schedule to ensure that teachers were provided with implementation guidance in keeping with the design of the curriculum. Materials were purchased from the publisher Amplify Education, but otherwise the publisher was not involved in the study.

Schools were recommended for participation by the districts due to a priori factors, such as schools coming from common geographic locations, schools that were performing poorly, and the willingness of the school leadership to adopt the CKLA: Knowledge curriculum. Initially, principals of 50 schools signed letters of agreement with the research team, indicating a commitment to adopting CKLA: Knowledge in all kindergarten classrooms, with the understanding that schools that were randomly assigned to the control condition would receive the curricular materials and training workshops the following school year. One school in Study 1

and two schools in Study 2 declined participation prior to random assignment and thus were not considered part of the study samples.

To conduct the random assignment in each study, schools were divided into two groups (strata) based on school size operationalized by the number of kindergarten classrooms and were subsequently randomly assigned to condition within each group (stratum). In Study 1, small schools included 1 or 2 teachers, and large schools included 3, 4, or 5 teachers. In study 2, small schools included 2 or 3 teachers, and large schools included 4, 5, or 6 teachers. Balance between condition was immediately assessed after random assignment via the school-average Peabody Picture Vocabulary Test-IV (PPVT-IV) pretest standard score (Dunn and Dunn, 2007), as well as the number of students and classrooms. There were no statistically significant differences between conditions in either study. The 35 schools in the treatment condition began implementation of the curriculum in December 2017 or 2018 (i.e., soft start implementing lessons from *The Five Senses* unit), with full implementation starting in January 2018 or 2019 (for Study 1, Study 2, respectively) through the end of the academic year. Posttest data collection occurred in the spring of the academic year.

Kindergarten teachers within each of the 47 schools provided voluntary consent for teacher-level data collection (i.e., demographic surveys, reading logs, videotaping, facilitation meetings) and at the beginning of the school year, sent home parent permission forms (in a variety of languages [English, Arabic, Burmese, Chinese, French, Kurdish, Somali, Spanish] depending on the home language of the parent). Parent permission allowed for individual student testing and access to district demographic data. Multiple attempts were made to obtain parent permission, including a verbal consent phone protocol approved by the university IRB and the district research offices. Parent permission was obtained for 56.4% students enrolled in

kindergarten classrooms across schools (range per classroom = 23%-90%). Of students for whom we had parent permission, approximately 30 students per school were drawn evenly across classrooms (e.g., target of 10 students per class in a school with 3 Kindergarten classes). If a given classroom had a greater number of consents than our target number, we randomly selected consented students for testing. Assessed students were considered participants in this study, though it is important to note that all kindergarten teachers in a given school and students enrolled in their classrooms received the same access to the CKLA: Knowledge curriculum, regardless of study participation status.

Participants

Participants were 1,194 students (Treatment = 232 in Study 1, 333 in Study 2; Control = 312 in Study 1, 317 in Study 2) enrolled in 134 classrooms (Treatment = 24 in Study 1, 41 in Study 2; Control = 29 in Study 1, 41 in Study 2) across 47 schools (Treatment = 11 in Study 1, 12 in Study 2; Control = 12 in Study 1, 12 in Study 2). Of the 1,217 students randomly selected for testing, 23 were unable to be pretested due to a change of grade, transfer, or move from the district prior to initial data collection. Since pretesting occurred prior to random assignment of schools, these students were not considered part of the intent-to-treat (ITT) study sample. However, it is important to note that students who were not pretested due to absence remained in the ITT sample and were tested in the spring. CONSORT charts for the two studies are in Appendix C of the supplemental materials. Table 1 displays the demographic information of the combined sample, including gender, race/ethnicity, language status, and disability status. According to publicly available data, 82% of students in Study 1 were eligible for free or reduced lunch. About 59% of students in Study 2 were considered economically disadvantaged by the school district. Table 2 reports sample sizes, means, and standard deviations of pretest

measures including age at pretest. Standard scores on pretest assessments indicated that students performed in the average range across measures, with consistently low average scores on the Test of Narrative Language (see Appendix D). There were no statistically significant differences between treatment and control groups across measures or demographic variables; however, there were a few significant differences between Study 1 and Study 2 participants. In Study 2, there was a greater percentage of White, Hispanic, and multiracial participants and a smaller percentage of Black participants. Study 1's sample was predominantly Black, which is representative of the district in Study 1. Study 1 participants performed better at baseline on the standardized science measure at pretest, but there were no statistically significant differences in vocabulary and listening comprehension pretest measures. There were no statistically significant differences in demographic variables between conditions in either study or in the full sample across studies. For Study 1, there was a significantly larger proportion of children with an IEP in the treatment group as compared to the control group. For Study 2, there were more ELLs in the treatment group than in the control group.

Of the 123 teachers who completed an end-of-year survey (out of 134 total), 26.8% reported that a Bachelor's degree was their highest level of education, with 50.4% reporting holding a Master's degree and 7.3% reporting a degree beyond the Master's degree. Teachers reported having 6.44 years of teaching experience (SD = 3.44; 0-21 years) and 94.3% reported being certified to teach Kindergarten. Teachers were overwhelmingly female (96.7%), with race primarily reported as either White/Caucasian (66.7%) or Black/African American (27.6%). Only 3.3% reported their ethnicity as Hispanic or Latino.

Participation Incentives. To prevent attrition from the study, a series of incentives were provided to schools, teachers, and children. Schools were provided with curricular materials and

training at no cost; control group schools were offered professional development (PD) days and curricular materials a year after the treatment schools implemented CKLA: Knowledge (i.e., waitlist control). Teachers were provided with $25 gift cards for offering the opportunity for voluntary child participation to parents in their classrooms. Children were given a book at both fall and spring testing windows. The only difference between studies was in teacher compensation due to district regulations. In Study 1, treatment teachers received monetary compensation for any PD training days that occurred outside of the district established PD dates, and teachers in the control group were provided with $200 in gift cards for classroom supplies. In Study 2, control group teachers received up to $200 for research task completion (i.e., completing surveys, completing reading logs, meeting with project staff).

Attrition. All schools remained in the sample for the duration of the study. There were some teachers in each condition who were replaced during the year and a few classroom composition changes (see CONSORT charts in Appendix C of supplemental materials for detailed information). Of the 1,194 children, three who initially began in the treatment group moved to the control group, and two from the control group moved to the treatment group. For the ITT analyses conducted in this manuscript, these students' scores were analyzed according to their initial condition assignment of classroom and school. Overall student attrition, when considering students with at least one pretest measure and no posttest measures, was 7% with 6.0% attrition in the treatment group and 7.8% attrition in the control group (differential attrition of 1.77%), meeting criteria set by WWC for a tolerable threat of bias under both optimistic and cautious assumptions (What Works Clearinghouse, 2021). Attrition was also calculated separately for each measure as the number of students who did not have posttest data for that

measure divided by the number of students who were assigned IDs upon entry into the study, which can be found in Appendix E of the supplemental materials.

Data Collection

An individual battery of student-level assessments of approximately 45 minutes in length was administered by the research team in a 7- to 9-week window in the beginning of the kindergarten academic year (prior to random assignment; October-December 2017, September-October 2018, for Study 1, Study 2, respectively) and in the spring of the academic year (April-June 2018, March-May 2019, for Study 1, Study 2, respectively). All measures except the proximal vocabulary (words taught) and proximal knowledge (knowledge taught) measures were administered at both pretest and posttest windows. Assessment administration took place in quiet locations within the student's school. Data collectors were blind to condition. Prior to collecting individual student data in the fall, assessors completed a standardized training that typically included: (1) viewing an online training, (2) passing a quiz, (3) attending an in-person training session, (3) practicing the assessments, (4) observing administration by an experienced assessor, and (5) being observed in the field. Prior to the spring assessment window, new data collectors completed the same process and returning data collectors were provided with a refresher training. Data collected in the field were double checked for completion scoring accuracy by the lead assessor, then mailed to the FSU lab, where scoring re-checks took place and data were double-entered into REDCap software (Harris et al., 2009; Harris et al., 2019).

Measures

Oral Language Skills

Vocabulary. Vocabulary was measured using both proximal measures (i.e., researcher-developed) and distal measures (i.e., standardized tests). The proximal measure was 60-item

receptive measure (Curriculum-Specific Vocabulary Test [CSVT]) for which we drew a stratified random sample of 15 words evenly from the Plants, Farms, and Native American units. These units were selected because all students in the treatment condition would have been exposed to these full units prior to the start of the posttest data collection window. Six of the selected words were domain-specific (e.g., pollination, harvest, wigwam) and nine were general-purpose (e.g., gigantic, valuable, mischief). For each of the 5 words drawn from a given unit, two were defined during a read-aloud and three were both defined and reviewed in the Word Work activity after the read-aloud; three were general-purpose and two were domain-specific words. In this task designed based on work by Beck and McKeown (2007), students were asked to respond by answering yes or no to a series of 4 questions about each word (two sets of two questions, one being true, the other false). Two of the questions situated the word within a context (e.g., Could you harvest corn in a field? Would you harvest pencils in a school?). Two of the questions were definitional (e.g., Does harvest mean to gather a crop that is ready to eat or use? Does harvest mean to spread seeds around so plants grow everywhere?). Internal consistency in this sample was calculated using ordinal alpha as .89.

Standardized measures assessed both receptive and expressive vocabulary. The *Peabody Picture Vocabulary Test- Fourth Edition* (PPVT-IV; Dunn & Dunn, 2007) measured receptive vocabulary, in which students identified one of four illustrations that best represented a target word spoken by the assessor. Internal consistency for the measure is reported in the manual as about .96 and calculated in this sample using ordinal alpha as .99 at pretest and .99 at posttest. Generalized expressive vocabulary was measured by the Picture Vocabulary subtest of the Woodcock-Johnson III Tests of Achievement (WJ PVT; Woodcock et al., 2001). In this measure, students were asked to name objects depicted in a series of illustrations. This subtest

manual has a reported median test reliability of .81. Internal consistency in this sample was calculated using ordinal alpha as .96 at pretest and .96 at posttest.

Listening Comprehension. Listening comprehension was measured with standardized measures, including the Sentence Structure subtest of the *Clinical Evaluation of Language Fundamentals- Fourth Edition* (CELF SS; Semel et al., 2004) and the *Test of Narrative Language* (TNL; Gillam & Pearson, 2004). The CELF SS assessed the child's ability to understand increasingly complex sentences by selecting the picture that most closely represented the sentence provided by the assessor. Internal consistency reliability estimates reported in the manual average .87 across ages and calculated using ordinal alpha in this sample as .87 at pretest and .90 at posttest. Narrative Comprehension, the receptive portion of the TNL, individually assessed children's ability to understand stories in three formats: with no picture cues, with five sequenced pictures, and with a single picture. After the administrator read aloud the text, literal and inferential questions were posed to the student. The TNL administration was audio recorded in the event there were questions about item scoring accuracy.

Content Knowledge: Science and Social Studies

Proximal content knowledge (i.e., domain knowledge) of science and social studies was measured using the Curriculum-Specific Knowledge Test (CSKT), a researcher-developed task adapted from Cervetti et al. (2016). Students responded to two separate prompts: *Tell me everything you know about plants* (science prompt) and *Tell me everything you know about Native Americans* (social studies prompt). Administration was untimed and assessors were instructed to provide students with adequate time to respond. After the assessor gave the initial verbal prompt (e.g., *Tell me everything you know about plants.*), a student provided the response, and was given a wait-time of 5 seconds after they stopped talking prior to a follow-up question

(i.e., *Can you tell me anything else about plants? I want to be sure you're sharing everything you know about plants.*) No corrective feedback was provided to students. The assessment was audio recorded and transcribed for offline scoring. Each prompt was scored on the basis of key idea units that were part of the CKLA: Knowledge curricular units on Plants and Native Americans. First, utterances containing a subject and a verb were identified within the student's talk. Then, novel idea units within each utterance that reflected curricular idea units received credit (e.g., Leaves use the light to make their food.). A student's total number of unique idea units was calculated as their final score. There were two coders for each study, trained by the sixth author (who adapted the tool) using 10-15 audio transcriptions. Twenty percent of the 2,169 transcripts were randomly selected for double coding. Inter-rater reliability was high, with accuracy of parsing utterances ranging from 91.6% to 100% and accuracy of coding idea units ranging from 86.7% to 92.8%.

Generalized content knowledge was measured using the standardized Science and Social Studies subtests of the Academic Knowledge test of the WJ-III Tests of Achievement (Woodcock et al., 2001). The first section of each subtest required response by pointing, and the remainder of the items required oral response. The Science subtest assessed information related to biological and physical sciences; the Social Studies subtest assessed information related to history, geography, government, and economics. Internal consistency for the Science subtest was calculated using ordinal alpha as .96 at pretest and .95 at posttest and for the Social Studies subtest as .94 at pretest and .95 at posttest.

Intervention

Treatment Condition

CKLA: Knowledge was intended to replace the portions of the English Language Arts instructional time that consisted of vocabulary and comprehension instruction, usually through the use of read-alouds. Schools committed to one-hour implementation of CKLA: Knowledge, intended to be in addition to their typical phonics instruction (delivered through programs such as Fundations, Journeys, Words Their Way, and Hooked on Phonics). (Note that a parallel Skills Strand is also available for instruction in foundational skills but was not tested in this study. No treatment classrooms reported using this strand.) The hour of implementation was most often implemented during English Language Arts instruction, but for some schools, time was split between the English Language Arts and content-area instructional times. Although the developers of CKLA: Knowledge (i.e., Core Knowledge Foundation) and the study personnel explicitly asked schools not to replace existing content area instruction with the curriculum, schools nevertheless decided to view CKLA: Knowledge (in some cases) as meeting their content area requirements and thus did not implement additional science or social studies instruction. This was better documented in Study 2, with only 15% of teachers reporting that, in addition to CKLA: Knowledge instruction in the English Language Arts instructional time, science was also taught separately regularly or daily. Fifty percent of teachers reported teaching science occasionally when rotating district science kits were available. Thirty-five percent reported that they rarely taught science separately. Forty-three percent indicated that they rarely taught Social Studies in addition to CKLA: Knowledge. Two of the schools in Study 2 identified as STEM schools.

CKLA: Knowledge Curriculum. The CKLA: Knowledge curriculum has an instructional approach aligned with several key practices found in prior trials of content-integrated literacy instruction in the primary grades (Hwang et al., 2021; Hwang et al., 2022). Specifically, the developers of CKLA: Knowledge employed two design principles that resulted in a curriculum designed to simultaneously develop vocabulary, listening comprehension, and content knowledge (McGinty & Bevilacqua, 2016). First, the curriculum emphasizes an interactive read-aloud lesson format to foster language learning critical to reading comprehension. As part of daily interactive read-alouds, teachers read complex texts that are typically beyond the average kindergarten reading level; display images to support vocabulary and understanding of the texts; and engage students in discussion before, during, and after reading aloud. Explicit vocabulary instruction in two to eight core words occurs during the read-aloud as teachers pause and provide a definition (breadth), as well as after the read-aloud, with specific focus on one target word from the text (i.e., Word Work), extending the vocabulary teaching activities to develop depth of word knowledge (Beck & McKeown, 2007; Beck et al., 2013; Coyne et al., 2010). Interactive discussions occur both during and after reading that encourage both literal and inferential language use, with brief teacher comments and questions interspersed during reading and more in-depth discussion after reading, with questions increasing in difficulty. Peer interaction is also encouraged through "think-pair-share" routines. An application activity occurs after reading that extends students' understanding of the content of the text, sometimes with writing activities.

The second design principle of the curriculum is that a knowledge-building approach to selecting and ordering texts that are read aloud may further facilitate vocabulary and comprehension development (McGinty and Bevilacqua, 2016). Indeed, a differentiating feature

of CKLA: Knowledge is that topics (in the domains of science and social studies) are intentionally sequenced to systematically build children's broad content knowledge over time, thus building the prior knowledge needed to understand subsequent texts. For example, the kindergarten Farms unit builds on the Plants unit, and learning in both topics is reinforced and serves as a foundation for learning about Native Americans. Moreover, the teachers' manual explicitly guides teachers to make connections between prior learning and new learning. Within a given topic (e.g., Plants), narrative and informational texts are combined within coherent text sets designed to incrementally build knowledge of the topic. Implicit vocabulary learning occurs through repeated exposure to academic vocabulary that reflects networks of words related to particular concepts.

CKLA Curriculum Implementation. Due to baseline testing prior to random assignment, curricula implementation was shortened into about 6 months (i.e., December – May). To ensure that the curriculum exposure was in keeping with the developers' intent, the research team worked with the Core Knowledge Foundation (i.e., developers) to determine a logically ordered sequence that would maintain coherence within and across academic years of intervention, as this was originally designed to be a multi-year investigation (K-2). The sequence of units was: The Five Senses (Science), Plants (Science), Farms (Science), Native Americans (Social Studies), Seasons and Weather (Science), and Colonial Towns and Townspeople (Social Studies). It is important to note that, due to the multi-year nature of the intervention, implementation continued while posttesting occurred. The Lexile levels of these read-alouds for these units ranged from 810 to 990 (approximate reading level for grades 3-5). The percentage of informational text ranged from 75% - 90%.

Professional Development. Two full days of professional development were provided to teachers from a Core Knowledge Foundation staff member. One full day workshop, occurring prior to curriculum implementation, consisted of describing how CKLA: Knowledge supports vocabulary and comprehension; connecting the curriculum to the English Language Arts standards; providing an overview of materials in the program; examining all major lesson components; and practicing a sample lesson. The second full day workshop, occurring about one month after the start of curricular implementation, covered the collecting and use of assessment data; implementing scaffolding strategies without modifying the objective or rigor; and preparing for units. Teachers who were absent from the training were provided access to online make-up sessions with optional office hours provided for questions. Examination of videos from PD events indicated that the presenter adhered to the intended components of the training with fidelity.

Teachers received training and ongoing support from a research staff member serving as curriculum facilitator. Every two weeks in the Spring semester, the facilitator met with treatment teachers as a grade level with the goal of planning for the following two weeks of curricular implementation (total number of meetings = 9). The facilitator also collected information from teachers about program implementation. On average, teachers attended seven meetings (SD = 1.5, range = 3 – 10).

Business-as-Usual Control Condition

In both districts, there was a commitment to content knowledge building during English Language Arts instruction with read-alouds recommended by the districts in their district scope and sequence materials. Literacy leaders (at the district level in Study 1 and at the school level in Study 2) supported teachers in their instruction. Control group teachers in Study 1 reported using

the district-created curriculum, with about a third of classrooms also using Wilson Fundations for their phonics instruction. One classroom used EngageNY materials, which included an earlier version of CKLA: Knowledge. In Study 2, all control group teachers reported using Journeys, and most also used the scope and sequence provided by the school district, along with the Heggerty phonological awareness program. Two classrooms reported using the CKLA: Skills strand (i.e., phonics instruction). About half of the Study 1 and all of Study 2 control teachers taught science and social studies as part of English Language Arts instruction. All Study 2 teachers taught additional science lessons at least occasionally with district science kits. Two of the Study 2 schools identified as STEM or STEAM schools. Most Study 2 teachers taught some social studies outside of English Language Arts through Social Studies Weekly materials.

Fidelity of Implementation

Adherence to CKLA: Knowledge Lessons. We measured teachers' adherence to the CKLA curricular lessons across units through a checklist that observed teachers' use of major elements (i.e., introduction, setting a purpose, defining core vocabulary, comprehension questions, word work, application), occurring before, during, and after the reading aloud of the text. Suggested verbal prompts and talking points (i.e., extra-textual talk) were included in these elements. We video-recorded each teacher 1-3 times (mode = 2), double coding a randomly selected 23% of videos for the presence or absence of the elements. Inter-rater reliability was generally high, with intraclass correlation coefficients ranging from .72 to .99 with a mean of .89.

Teachers typically completed the major elements of the lesson. All teachers used flip book images during read aloud and virtually all (96.5%) read aloud to a large group as intended by the curriculum designers. Teachers generally taught the following elements of the observed

lessons: 96.7% included the introduction, 100% included the core vocabulary, 100% included comprehension prompts embedded within the read aloud, 95.1% included post-reading comprehension questions, and 88% included the application activity. [Note that the application activity was captured for only 80.6% of the teachers because videotaping took place within the English Language Arts instructional time, but teachers were given the opportunity to teach the application activity at a different time in the day than the rest of the lesson.] Interestingly, only 57.4% set a purpose for the read aloud and 67.2% engaged in post read-aloud word work.

Exposure. Teachers on average taught 71.77 lessons (SD = 11.38), calculated by fitting a best line to the expected number of lessons taught (76-77 lessons) and actual number of lessons taught reported in the facilitator ratings. Facilitators indicated that teachers were generally reflective of their own practices (M = 2.7, SD = 0.35) and receptive to facilitator feedback (M = 2.7, SD = 0.35) (max = 3).

Program Differentiation. During the Spring semester, teachers were asked monthly via survey to report information about their read-alouds on a given day. Although the reported median of the number of minutes read aloud (16-30 minutes) was the same across condition, treatment teachers reported reading aloud for a longer period of time per day: 6.5% of treatment teachers and 21.9% of control teachers read for less than 1-15 minutes; 59.7% of treatment teachers and 56.3% of control teachers read for 16-30 minutes; 27.4% of treatment teachers and 17.2% of control teachers read for 31-45 minutes; and 6.5% of treatment teachers and 4.7% of control teachers read for more than 45 minutes. The Mantel-Haenszel chi-square test demonstrated a statistically significant difference between groups in median minutes read aloud [$\chi^2(1) = 4.83$, $p = 0.03$].

Results

Missing data on the fall pretest measures ranged from 0.76% to 1.85%, and missing data on the spring outcome measures ranged from 7.14 to 9.15%. Little's MCAR test revealed the data to be missing completely at random [$\chi^2(291) = 371.80$, p=.001]. Full information maximum likelihood was used to account for the missing data. All models were conducted using the lme4 package (Bates et al., 2015) in R with a random intercept for school to account for the clustering present in the data. Unconditional 3-level models (children within classrooms within schools) revealed variance at the classroom level to be <.05 for all outcomes, thus only 2-level models were conducted with children nested within schools. A z-scored version of pretest and posttest measures were used in these models to aid in interpretation of coefficients. Treatment was also coded with 0 representing control and 1 representing treatment. Therefore, treatment coefficients can be interpreted as effect sizes in standard deviation units (Hedge's g). We interpret these effect sizes (ES) using the benchmarks developed by Kraft (2020) for K-12 field interventions in education (.05, .20, and >.20 as small, medium, and large, respectively).

As indicated in Table 2, no statistically significant differences were observed on pretest measures between treatment and control groups within the full sample, or within study 1 and study 2 separately. (See Appendix F of supplemental materials for full baseline equivalence model results.) All treatment group differences on pretest measures were ES<.07 making it appropriate to use pretest as a predictor in the impact models according to the WWC threshold of ES<0.25 (2017). Statistically significant differences were observed between the Study 1 and Study 2 samples on the WJ Science pretest such that the Study 1 sample performed significantly better on average compared to Study 2. All other pretest measures except WJ Social Studies (WJ SS) had pretest differences between studies that may be practically important (ES>.05) although

not statistically reliable. Therefore, we include study sample as a covariate in all of the models as well. Table 3 contains means, standard deviations, and correlations among measures within the full sample.

Interaction models were also run to compare treatment effects between Study 1 and Study 2. No statistically significant interactions were found across outcomes, with the exception of the measure of proximal knowledge of social studies (CSKT NA). Treatment effects for both studies were positive but Study 2's effect was significantly larger in magnitude.

Main Effects on Oral Language and Content Knowledge

Effects on Oral Language

Two-level hierarchical linear models, with children nested within schools, were used to address the confirmatory research question about the intent-to-treat impact of the intervention on children's oral language (i.e., vocabulary, listening comprehension) and content knowledge (i.e., science, social studies). Models controlled for pretest scores, study, and random assignment stratum of school size. Pretest scores were not available for proximal measures (CSVT, CSKT PL, and CSKT NA) and, in those cases, PPVT was used as a covariate to represent baseline language performance. The Benjamini-Hochberg linear step-up procedure (Benjamini & Hochberg, 1995) was used to control Type I error per domain examined (vocabulary, listening comprehension, science, social studies). In this procedure, p-values are rank ordered and assigned a critical value according to their rank. This procedure reduces the false discovery rate by penalizing for having non-significant results, requiring a smaller critical value for the more p-values above its critical value. We also conducted corrections within larger domains of oral language versus content knowledge, and statistical significance did not differ between the two

approaches. Table 4 presents results of the main effects models on standardized outcomes and Table 5 presents results of main effects models on proximal outcomes.

Vocabulary. Three measures of vocabulary were administered including one proximal (CSVT) and two standardized (PPVT and WJ PVT). Results revealed a statistically significant main effect of treatment on the proximal measure of vocabulary on the words that were explicitly taught in the curriculum (CSVT) such that those in the treatment group performed .63 standard deviations higher on average than those in the control group (ES = .63, $p < .001$). According to Kraft's (2020) guidelines, this would be classified as a very large effect. With regard to standardized measures, the intervention did not have a statistically significant effect on children's receptive vocabulary as measured by the PPVT but did exert a significant effect on expressive vocabulary, as measured by the WJ Picture Vocabulary test (ES = 0.09, $p = .016$). This effect was small by Kraft's guidelines.

The vocabulary domain had three treatment effect p-values (PPVT, WJ PVT, and CSVT), and after performing the linear step-up procedure, the critical p-value was determined to be .03, yielding statistically significant effects for CSVT and WJ PVT.

Listening Comprehension. There were no statistically significant effects on standardized measures of listening comprehension for CELF SS (ES = -0.05, p = .301) or TNL (ES = 0.02, p = .725).

Effects on Content Knowledge

Science Knowledge. There were two measures of science knowledge, one proximal (CSKT PL) and one standardized (WJ Science). Results revealed statistically significant treatment effects on the proximal measure of knowledge about plants (CSKT PL) such that students receiving the CKLA intervention performed .26 standard deviations better on average

than those in the control group, a large effect according to Kraft's guidelines (ES = 0.26, p < .001). The standardized measure of science knowledge (WJ Science), however, did not show statistically reliable treatment differences (ES = 0.11, p = .054). (Please note that preliminary models reported in Cabell & Hwang [2020] indicated a statistically significant effect [$p < .05$] when results were meta-analyzed across studies and did not account for missing data.) Controlling for Type I error with the linear step-up procedure, the critical p-value was determined to be .025, yielding statistically significant effects for CSKT PL.

Social Studies Knowledge. Two measures of social studies knowledge were administered, one proximal (CSKT NA) and one standardized (WJ Social Studies). Results revealed statistically significant treatment effects on the proximal measure of knowledge about Native Americans (CSKT NA), with a very large effect size such that on average those receiving the CKLA intervention performed 0.93 standard deviations higher than those in the control group (ES = 0.93, p<.001). The standardized measure (WJ Social Studies), however, did not reveal statistically significant treatment group differences (ES = .00, p = .999). The linear step-up procedure determined the critical p-value for the social studies domain to be .025, maintaining significance for the proximal measure of knowledge about Native Americans.

Interaction Effects

Three sets of exploratory analyses were conducted to explore whether the impact of the intervention was moderated by student's initial skills at kindergarten entry, ELL status, or economic disadvantage (Study 2 only). Data on economic disadvantage were only available for those in the Study 2 sample, thus those analyses are limited to that sample only. Given that these are exploratory analyses and the sample was not powered to explore interactions, we use a p<.10 cut off to represent significance. Statistically significant interactions were followed up with

simple slope analyses examining treatment group differences at the mean and 1SD above and below the mean on initial skill. Simple slopes analyses help to interpret interactions by examining at what point along the moderator are there differences between the treatment and control groups.

Initial Skills

Results of interaction effects by initial skills can be found in Table 6 for standardized outcomes and Table 7 for proximal outcomes. WJ Social Studies was the only standardized measure that had a statistically significant interaction (p=.088) between initial skill and treatment condition; however, simple slopes analyses revealed no significant treatment differences at the mean or 1SD above or below the mean on WJ SS pretest. All three proximal measures revealed statistically significant interactions with initial PPVT performance. CSVT revealed a statistically significant interaction such that treatment effects were larger for those with better initial scores on the PPVT (see Figure 2). Simple slopes analyses revealed statistically significant treatment effects at the mean of PPVT pretest (ES = .63, p<.001), 1SD above the mean (ES = .83, p<.001), as well as those 1SD below the mean on PPVT pretest (ES = .42, p<.001). The proximal measure of science knowledge about plants (CSKT PL) also revealed a significant interaction with PPVT pretest such that larger treatment effects were observed for those who started higher on PPVT (see Figure 3). Statistically significant treatment effects were observed at the mean of PPVT (ES = .25, p<.001) as well as at 1SD above the mean on PPVT pretest (ES = .38, p<.001), but not at 1SD below the mean on PPVT pretest (ES = .13, p=.139). This suggests that students starting low on initial PPVT showed less benefits of treatment on plant knowledge than did students starting with higher PPVT. The proximal measure of social studies knowledge of Native Americans (CSKT NA) also showed a statistically significant interaction with PPVT pretest such

that those with higher PPVT pretest showed greater benefits of treatment (see Figure 4). Significant treatment benefits were observed at the mean of PPVT pretest (ES = .92, p<.001), 1SD above the mean (ES = 1.31, p<.001), and 1SD below the mean (ES = .53, p<.001). The relation of initial language skills to outcome was only statistically significant for the treatment group (ES = .41, p<.001) but not the control group (ES = .01, p=.670), most likely due to the fact that the control group did not report teaching this content.

English Language Learner Status

Results of interaction effects by English Language Learner Status can be found in Table 8 for standardized and Table 9 for proximal outcomes. No significant interactions were observed between ELL status and treatment condition on any of the standardized measures or the proximal science measure (CSKT PL). A statistically significant interaction between ELL status and treatment condition was observed for CSVT (p=.052) such that treatment effects were larger for native English speakers compared to ELLs (see Figure 5). However, significant treatment effects on CSVT were observed for both ELLs (ES = .40, p=.004) and native English speakers (ES = .68, p<.001). Similarly, a statistically significant interaction between ELL status and treatment condition was observed for CSKT NA (p=.009) such that treatment effects were larger for native English speakers compared to ELLs (see Figure 6). Again, statistically significant treatment effects on CSKT NA were observed for both ELLs (ES = .58, p<.001) and native English speakers (ES = .99, p<.001).

Economic Disadvantage (Study 2 Only)

Results of interaction models by economic disadvantage can be found in Tables 10 and 11 for standardized and proximal outcomes respectively. There were no significant differences in

treatment effects between children considered by the district as economically disadvantaged and non-economically disadvantaged on any outcome.

Discussion

This paper presented the combined results of two RCTs—the second being a replication of the first—in large urban districts that examined the impact of a Tier 1, whole-class content-rich literacy curriculum on kindergarten children's vocabulary, listening comprehension, and content knowledge. Secondarily, we examined child-by-instruction interactions with children's initial skills, ELL status, and poverty status. Investigating the impact of content-rich literacy curricula on student outcomes is particularly important in high poverty settings, such as the districts in the present RCTs, given that research demonstrates that children in these settings are more likely to be exposed to lower quality instruction and less content knowledge (LoCasale-Crouch et al., 2008; Schmidt et al., 2015). Moreover, replication of interventions is rare but is critically important for advancing the science in the field of educational psychology. CKLA: Knowledge demonstrated a positive and statistically significant impact on vocabulary, not only for the words taught in the curriculum, but also for general expressive vocabulary, as measured by a standardized test. In addition, there was a main impact on children's proximal science and social studies knowledge. There were no significant main effects of the curriculum on standardized measures of content knowledge or listening comprehension. The pattern of effects across the two studies were remarkably consistent indicating that the intervention had a positive effect on vocabulary and content knowledge outcomes for kindergarten students. Statistically significant interactions were found for proximal measures of vocabulary and knowledge, such that children who began the year with relatively higher receptive vocabulary scores derived a greater benefit of learning the words and content knowledge taught in the curriculum.

Impact of CKLA: Knowledge on Language and Knowledge Outcomes

We tested the impact of the widely used CKLA: Knowledge curriculum, one of two strands of the broader CKLA curriculum designed for K-2 settings, that systematically integrates oral language and content-area instruction as a lever to facilitate growth in both oral language skills and science/social studies knowledge simultaneously. Not only can integrating instruction in these areas make more efficient use of limited instructional time in a day, but prior meta-analyses also show that the integration itself appears to have a positive effect on literacy development (Hwang et al., 2022, 2023).

Indeed, the present study joins a small set of studies in the primary grades (K-2) that have tested Tier 1 whole-class content-rich literacy approaches that replace [rather than supplement] English Language Arts instruction (Gray et al., 2022; Kim, Burkhauser, et al., 2021; Kim, Relyea, et al., 2021), including instruction in both science and social studies domains (Connor et al., 2017; Kim, Burkhauser, et al., 2023). In short, ours is the first experimental report that examines the effect of a content-rich, core literacy curriculum on kindergarteners' oral language and science/social studies knowledge. The interrelated underlying processes specific to the CKLA: Knowledge curriculum include systematically building content knowledge through interacting with texts and strengthening semantic networks. These processes are hypothesized to improve children's oral language and content knowledge through the development of schemata about sequentially introduced topics (i.e., systematically building background knowledge before new knowledge is introduced) via interacting with texts— specifically reading aloud of coherent text sets with discussion and writing designed to build content knowledge—along with attention to knowledge networks with the explicit teaching of vocabulary and exposure to related words and ideas around concepts.

Main Effects on Oral Language Skills

CKLA: Knowledge demonstrated a statistically significant, positive, and very large effect on children's knowledge of words that were taught in the curriculum. This finding is in keeping with the broader oral language literature (Marulis & Neuman, 2010) as well as other content-rich literacy interventions that demonstrated a positive effect on proximal measures of vocabulary (Hwang et al., 2022, 2023; Neuman & Kaefer, 2018; Neuman et al., 2021). It is important to note that words tested were drawn from throughout the curriculum, represented a mix of domain-specific and general-purpose words, and included words for which only definitions were provided along with those for which more intensive word work was conducted. As in the case with the LARRC et al. (2022) study, intervention took place over the course of a semester and, thus, there was a lapse in time between instruction on a given word and end-of-year assessment, representing a sustained effect on word learning. Over time, with additional curricular implementation, these proximal effects could result in more generalized effects on vocabulary. If the CKLA: Knowledge K-2 program were implemented fully, children would have an opportunity to learn the meaning of approximately 2000 words. Approximately half of these words are moderately frequent words (falling into Hiebert's Word Zones 3 and 4; Hiebert, 2005) that, together with the frequent words most children know, account for a large proportion of the words that elementary school children encounter in the texts they read (Hiebert, 2005; Hiebert, 2013; Hiebert, Goodwin, & Cervetti, 2018). Thus, over time, the vocabulary instruction component of CKLA: Knowledge could have a significant effect on children's generalized vocabulary and listening comprehension, and ultimately reading comprehension, even if none of the taught words appeared on a measure of general vocabulary knowledge.

Given that the expectation of impact on generalized oral language skills would be over time, it is surprising that CKLA: Knowledge exerted a statistically significant, positive main effect on standardized expressive vocabulary after only one semester of implementation. Though the effect is small at 0.09, it is meaningful, given the difficulty of interventions in shifting children's generalized vocabulary (Elleman et al., 2009; Marulis & Neuman, 2013; Silverman et al., 2020). Put into perspective, Lortie-Forgues and Inglis (2019) reported that large-scale randomized controlled trials in the U.S. (commissioned by the National Center for Educational Evaluation and Regional Assistance) typically demonstrate an average effect size of only 0.01 on kindergarten outcomes as well as on oral language outcomes across grade levels. Upon further analysis, however, we found that the three content-area words on the WJ Picture Vocabulary explicitly taught in CKLA: Knowledge appeared to drive this effect, such that the effect size was reduced to 0.01 when these items were removed from analyses.

In a recent meta-analysis, Cervetti et al. (2023) reported the difficulty of the effects of vocabulary interventions to transfer to words beyond those taught. Indeed, of the multitude of oral language RCTs (primarily targeting vocabulary outcomes) conducted in early childhood settings, only a handful have demonstrated a statistically significant main effect on a standardized vocabulary measure, primarily for preschool-age children (e.g., Gonzalez et al., 2010 [receptive]; Neuman et al., 2021 [expressive]; Wasik & Hindman, 2020 [receptive]). When examining content-rich literacy approaches specifically, Neuman et al. (2021) reported positive effects on a standardized measure of expressive vocabulary in pre-kindergarten (but not kindergarten or first grade) when examining a structured adaptation of the 20-week supplemental World of Words program that integrates vocabulary and science instruction. Connor et al. (2017) reported null findings in K-2 on the same expressive vocabulary test used in the present RCTs in

her RfU study on CALI, a program that integrated instruction in oral language and both social studies and science. Thus, to our knowledge, the present study is the first content-rich literacy RCT that has demonstrated a statistically significant, positive effect on a standardized vocabulary outcome for kindergarteners.

Since it is so rarely observed in the extant literature on early vocabulary learning (Cervetti et al., 2023), it is important to understand why this transfer into generalized vocabulary occurred in this study. Indeed, as Coyne et al. (2010) stated, "a much stronger case could be made for the worth of vocabulary instruction in the early grades if there was evidence to suggest that there were benefits beyond just learning the meanings of words targeted for instruction" (p. 95). In CKLA: Knowledge, target words related to both content as well as general-purpose words were taught deeply and broadly. The program emphasized strengthening semantic networks through the use of coherent text sets on specific science or social studies topics. The relationships among words were carefully considered in selection and in instruction. Thus, students were exposed to content-area concepts that systematically built knowledge over time and teachers were introducing and repeating related words in a variety of contexts, while expecting students to use new words in spoken and written language. The domain-specific and general-purpose words that were taught might have, in part, formed the basis for incidental learning of related words. Cervetti et al. (2016) reported that fourth-grade students who read a set of conceptually coherent information texts on birds demonstrated significantly greater knowledge of the meanings of non-taught words than students who read a set of texts on a range of science topics. This kind of incidental learning could occur during one of the lessons (lesson-specific effects) or, importantly, in a future lesson in the same knowledge domain where semantically related words occur (transfer effects; see also Wright et al., 2022). It could be said

that CKLA: Knowledge creates opportunities for incidental vocabulary learning by exposing students to content-area concepts that systematically build knowledge and semantic networks over time. This is in keeping with preferential attachment (Steyvers & Tenenbaum, 2005), that students who have a well-connected lexicon should more readily acquire new words that are linked to known words.

Despite the effect on generalized expressive vocabulary, effects on generalized measures of receptive vocabulary and listening comprehension were non-significant. Whereas receptive and expressive modalities of vocabulary represent a unitary construct at this age (Lonigan & Milburn, 2017), past read-aloud interventions in early childhood have demonstrated an impact on only one modality (e.g., Gonzalez et al., 2010; Mol et al., 2008). However, the lack of effects may also be due to measurement issues, as vocabulary word selection on standardized tests is generally unsystematic or unspecified in terms of criteria, and there is an over-representation of rare or infrequent words (Pearson et al., 2007). In terms of listening comprehension, our findings are consistent with the RfU study by Connor et al. (2017) who did not find an impact on a standardized measure of listening comprehension. Moreover, LARRC et al. (2019) reported in another RfU study that the relation between their Tier 1 oral language curriculum and effects on reading comprehension is indirect, mediated through growth in target vocabulary. In the case of a content-rich literacy curriculum, it could be that effects on listening comprehension are also mediated through growth in target knowledge, though it is likely that the target knowledge must in some way relate to the topics on the comprehension measure for far transfer to occur (Kim, Burkhauser, et al., 2023).

The TNL measure used in this study may have been too distal to the intervention, and we did not use any near- or mid-transfer listening comprehension measures better aligned with what

was learned in the intervention. It is also important to note that the TNL contained only narrative passages, while the intervention included a significant amount of informational text which differs in its structure. Effects may have been seen had proximal measures been used, as Kim, Burkhauser, et al. (2023) demonstrated effects of the MORE intervention on near-transfer (proximal) and mid-transfer reading comprehension passages but not on far-transfer passages (distal) for first- and second-grade students. Our second measure of listening comprehension, the CELF SS, measured children's understanding of statements increasing in syntactical complexity. Past oral language studies have also demonstrated a failure to shift children's syntactical skills (Cabell et al., 2011; Silverman et al., 2020), hypothesized to be harder to shift than other areas of language (Vasilyeva & Waterfall, 2011). Perhaps to shift children's syntactical skills meaningfully, an intervention focused on this particular area of oral language may be necessary rather than simply exposure to more complex text via read-alouds and some embedded syntax actvities (Phillips, 2023).

Main Effects on Content Knowledge

Findings indicated that CKLA: Knowledge exerted a significant effect on the knowledge taught in the curriculum across both science and social studies topics. This finding is in keeping with the extant literature on integrated content and literacy approaches, with multiple studies demonstrating a significant impact on researcher-created measures of knowledge (Hwang et al., 2022; 2023). Williams and colleagues (2009) demonstrated that integrating content and literacy instruction improves literacy outcomes while not diminishing content area learning among second graders. Building on this, our study, along with other findings (Hwang et al., 2022), suggests that the integration can actually enhance oral language and content knowledge simultaneously. In the present work, teachers across conditions reported teaching kindergartners

about plants. Yet the children exposed to CKLA: Knowledge exhibited stronger knowledge of plants in the expressive language task, in which they were asked to share everything they knew about the topic of plants. Similar to the proximal vocabulary measure, several months elapsed between the plants unit (January) and the spring assessment window, indicating that the topic knowledge was sustained over time.

The lack of effects on the standardized social studies and science measures was in keeping with other studies, with most content and literacy integrated instructional approaches not examining student outcomes on standardized measures of content knowledge (Hwang et al., 2022). Hwang et al. (2022) reported that, while researcher-developed measures examined differences in highly and less aligned measures of knowledge, effects on highly aligned measures only were significant. Thus effects were not seen on far transfer measures (or distal), in keeping with more recent research (Kim, Burkhauser, et al., 2023). Though not statistically significant ($p = 0.054$), the effect size on the WJ Science measure was meaningful at 0.11, indicating some potential far transfer of science instruction from CKLA: Knowledge. It is important to note that no concepts or words on the WJ Science were explicitly taught in the curriculum.

Interaction Effects

When examining proximal measures of vocabulary and content knowledge, we found evidence of Matthew effects that favored students entering kindergarten with stronger English oral language skills. Specifically, our results indicated statistically significant interaction effects on proximal measures with initial receptive vocabulary or ELL status serving as a moderator. While the intervention exerted effects on all children regardless of initial skill, there was an interaction effect with the curriculum-specific vocabulary test such that children who began the year with relatively stronger receptive vocabulary skills demonstrated better word learning from

the intervention. This finding is in keeping with other research that demonstrates that children with better initial oral language skills benefit more from oral language interventions during the early childhood period (e.g., Cabell et al., 2011; Coyne et al., 2009, 2010).

Initial vocabulary skill also mattered for the impact of the intervention on students' end-of-year knowledge measures, which required expressive language to complete. Children in the treatment group with stronger initial vocabulary outperformed those in the control group in both their responses about plants and Native Americans. Yet, for knowledge about plants, specifically, only those children who had at least average initial vocabulary appeared to benefit from the intervention.

In addition, the treatment had a significant impact on both ELLs and non-ELLs when compared with the control group but was larger for non-ELLs on both the proximal vocabulary (CSVT) and proximal social studies (CSKT NA) measures. This is not in keeping with extant literature suggesting academic language and content-rich literacy interventions confer a greater benefit on ELLs (Neuman & Kaefer, 2018; Truckenmiller et al., 2019). Our findings are partially in keeping with a recent meta-analysis of K-5 language interventions demonstrating no differential impact on vocabulary outcomes for ELLs versus non-ELLs; that is, we found a differential impact on proximal but not generalized vocabulary.

Collectively, the results of moderator analyses support a "rich get richer" effect. Moreover, while it is possible that these findings are primarily a reflection of oral language skills (because of the expressive nature of the knowledge test itself), it is also plausible that children with better initial language skills are more likely to learn the vocabulary and knowledge being taught in the classroom. The development of oral language (e.g., vocabulary, listening comprehension) during the early childhood years is critical for proficiency in later reading

comprehension (e.g., Gough & Tunmer, 1986; Hjetland et al., 2020; Language and Reading Research Consortium [LARRC], 2015). The self-productivity principle suggests that there are "multiplier effects which are the mechanisms through which skills begets skills" (p. 35, Cunha & Heckman, 2007). Thus, quicker development of oral language in the early school years matters for literacy. Children with well-developed oral language skills early on are better equipped to take advantage of classroom learning opportunities, which are verbally mediated through conversations. To ameliorate this learning gap, it is important to identify students with lower language skills who would benefit from intensified supports in addition to the Tier 1 curriculum (e.g., Zucker et al., 2021).

It is important to note that poverty status did not moderate the relation between treatment and outcome in Study 2, with intervention benefiting all children regardless of poverty status with regard to vocabulary and knowledge. Some scholars argue that systematically teaching knowledge can level the playing field by providing necessary background knowledge with which to understand subsequent texts (e.g., Hirsch, 2003). It may be that benefit is only evident on comprehension after a much longer exposure to a knowledge building curriculum.

Limitations and Future Directions

This study examined one instantiation of a widely used content-rich literacy approach, which employed read-alouds as the primary instructional vehicle. Although read-alouds are a quality instructional context for language development in early childhood settings (Mol et al., 2009), it is important to note that the curriculum under study was an English Language Arts program that was not intended to replace content area instruction but rather be used in conjunction with high quality, discipline specific content area instruction. Thus, the curriculum did not feature components that necessarily reflected discipline-specific practices (e.g., hands-on

scientific inquiry). However, one limitation of the present findings is that treatment teachers may have supplanted science and social studies instruction with the CKLA: Knowledge curriculum, as several reported to us that they did not offer any other content-area instruction. Some researchers have been investigating highly integrated and interdisciplinary approaches of content area and literacy instruction which maintain the rigor of the content-area discipline (e.g., Duke et al., 2023; Romance & Vitale, 2017; Wright & Gotwals, 2017), "provid[ing] balanced attention to science and literacy growth rather than using one domain to achieve growth in the other or to reclaim time in the school day" (p. 652; Cervetti et al., 2012). The present studies would have been strengthened had treatment schools not only adopted CKLA: Knowledge as their English Language Arts curriculum but also implemented aligned science and social studies instructional units during content-area instructional time.

Relatedly, CKLA: Knowledge was typically one hour in length, rather than the 20-30 minutes typically devoted to reading aloud in the classroom. Our studies were not designed to isolate the effects of individual features of the program, such as increased read-aloud instructional time, and thus instructional time may be a confounding factor (Dagli, 2019). However, increasing instructional time alone does not necessarily benefit literacy outcomes for kindergarteners and there are important child-by-instruction interactions that affect learning (Simmons et al., 2007; Sonnenschein et al., 2010).

Another limitation of the present work is that we measured outcomes after only one semester of intervention. Although this timeframe is in keeping with or longer than other studies of content-integrated instruction (Hwang et al., 2022) or vocabulary instruction (e.g., Coyne et al., 2010), it takes time to build children's oral language and content knowledge in ways that will impact eventual reading comprehension (Hirsch, 2003). Future research should study multi-year

implementation of commercially available, content-rich literacy interventions to ascertain impact on reading comprehension, as there is impressive emerging evidence of this longitudinal effect (e.g. Kim, Burkhauser, et al., 2023; Kim, Gilbert, et al., 2023). In Study 2 reported here, we were able to provide children the treatment in both kindergarten and first grade yet were precluded from data collection at the end of first grade due to the COVID-19 pandemic's effect on school closures. We are currently analyzing longitudinal data collected by the school district to examine the sustained impact of intervention across two academic years on reading comprehension.

ADD MORE HERE. Future work is also needed to understand how the key processes in content-rich literacy curricula mediate impacts on far transfer outcomes. We need to better understand how and whether gains on proximal knowledge and vocabulary support later improvement on standardized outcomes.

Conclusion

Scholars have long argued that knowledge is critical to reading success. This study suggests that simultaneously improving children's oral language and content knowledge is a tractable educational aim. Given the lack of early content-area instruction in the U.S. (e.g., Tyner & Kabourek, 2020), integrating content knowledge into English Language Arts is a feasible approach for providing more content for children in a way that is both cost-effective (i.e., $316 per student in the present studies) and scalable. Further, as noted by Hirsch (2003), by increasing children's "knowledge of words and the world" (p.10), we might be able to solve the puzzle of the nation's stagnant reading comprehension scores. Systematically teaching young children about that knowledge through engaging, interactive read-alouds of content-rich texts is a theoretically informed and evidence-based approach that can ultimately boost learning for all students.

References

Anderson, R. C., & Pearson, P. D. (1984). A schema-theoretic view of basic processes in reading comprehension. In P. D. Pearson, R. Barr, M. L. Kamil, & P. Mosenthal (Eds.), *Handbook of reading research* (1st ed., pp. 255–291). Longman.

Beck, I. L., & McKeown, M. G. (2007). Increasing young low-income children's oral vocabulary repertoires through rich and focused instruction. *The Elementary School Journal*, *107*, 251-271. https://doi.org/10.1086/511706

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

Brabham, E., & Lynch-Brown, C. (2002). Effects of teachers' reading-aloud styles on vocabulary acquisition and comprehension of students in the early elementary grades. *Journal of Educational Psychology*, *94* (3), 465-473. https://doi.org/10.1037/0022-0663.94.3.465

Cabell, S. Q., DeCoster, J., LoCasale-Crouch, J., Hamre, B. K., & Pianta, R. C. (2013). Variation in the effectiveness of instructional interactions across preschool classroom settings and learning activities. *Early Childhood Research Quarterly, 28*(4), 820-830. https://doi.org/10.1016/j.ecresq.2013.07.007

Cabell, S. Q., Justice, L. M., Piasta, S. B., Curenton, S. M., Wiggins, A., Turnbull, K. P., & Petscher, Y. (2011). The impact of teacher responsivity education on preschoolers' language and literacy skills. *American Journal of Speech-Language Pathology, 20*(4), 315–330. https://doi.org/10.1044/1058-0360(2011/10-0104)

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

Cabell, S. Q., & Hwang, H. (2023). Leveraging content-rich English language arts instruction in the early grades to improve children's language comprehension. In S.Q. Cabell, S.B. Neuman, and N. Patton Terry (Eds.), *Handbook on the science of early literacy* (pp. 175-185). Guilford.

Cervetti, G. N., Barber, J., Dorph, R., Pearson, P. D., & Goldschmidt, P. G. (2012). The impact of an integrated approach to science and literacy in elementary school classrooms. *Journal of Research in Science Teaching*, *49*(5), 631–658. https://doi.org/10.1002/tea.21015

Cervetti, G. N., Fitzgerald, M. S., Hiebert, E. H., & Hebert, M. (2023). Meta-Analysis examining the impact of vocabulary instruction on vocabulary knowledge and skill. *Reading Psychology*, 1-38. https://doi.org/10.1080/02702711.2023.2179146

Cervetti, G. N., & Hiebert, E. H. (2019). Knowledge at the center of English language arts instruction. *Reading Teacher, 72*(4)*,* 499–507. https://www.jstor.org/stable/26801637

Cervetti, G. N., & Wright, T. S. (2020). The role of knowledge in understanding and learning from text. In E.B. Moje, P. Afflerbach, N.K. Leseaux, & P. Enciso (Eds.), *Handbook of reading research* (Vol 5, pp. 237-260). Routledge.

Cervetti, G. N., Wright, T. S., & Hwang, H. (2016). Conceptual coherence, comprehension, and vocabulary acquisition: A knowledge effect? *Reading and Writing: An Interdisciplinary Journal, 29*(1), 1–19. https://doi.org/10.1007/s11145-016-9628-x

Connor, C. M., Dombek, J., Crowe, E. C., Spencer, M., Tighe, E. L., Coffinger, S., Zargar, E., Wood, T., & Petscher, Y. (2017). Acquiring science and social studies knowledge in

kindergarten through fourth grade: Conceptualization, design, implementation, and efficacy testing of content-area literacy instruction (CALI). *Journal of Educational Psychology, 109*(3), 301–320. https://doi.org/10.1037/edu0000128

Connor, C. M., Rice, D. C., Canto, A. I., Southerland, S. A., Underwood, P., Kaya, S., Fishman, B., & Morrison, F. J. (2012). Child characteristics by science instruction interactions in second and third grade and their relation to students' content-area knowledge, vocabulary, and reading skill gains. *The Elementary School Journal*, *113*(1), 52-75. https://doi.org/10.1086/665815

Core Knowledge Foundation and Amplify Education, Inc. (2017). *Core Knowledge Language Arts: Knowledge Strand.* Core Knowledge Foundation and Amplify Education, Inc.

Coyne, M. D., McCoach, D. B., Loftus, S., Zipoli Jr, R., & Kapp, S. (2009). Direct vocabulary instruction in kindergarten: Teaching for breadth versus depth. *The Elementary School Journal*, *110*(1), 1-18. https://doi.org/10.1086/598840

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

Coyne, M. D., McCoach, D. B., Loftus-Rattan, S. M., Baker, D. L., & Ware, S. M. (2022). Long-term effects of kindergarten vocabulary instruction and intervention on target vocabulary knowledge through second grade. *Topics in Language Disorders*, *42*(4), E9-E28. https://doi.org/10.1097/TLD.0000000000000296

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

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. Irvington.

Cunha, F., & Heckman, J. (2007). The technology of skill formation. *American Economic Review*, *97*(2), 31-47. https://doi.org/10.1257/aer.97.2.31

Dağlı, U. Y. (2019) Effect of increased instructional time on student achievement, *Educational Review*, *71*(4), 501-517. https://doi.org/10.1080/00131911.2018.1441808

Dickinson, D. K., Freiberg, J. B., & Barnes, E. M. (2011).Why are so few interventions really effective?: A call for fine-grained research methodology. In S. B. Neuman & D. K. Dickinson (Eds.), *Handbook of early literacy research (*Vol. 3, pp.337-357). Guilford Press.

Dickinson, D. K., & Smith, M. W. (1994). Long-term effects of preschool teachers' book readings on low-income children's vocabulary and story comprehension. *Reading Research Quarterly*, *29*(2), 105-122. https://doi.org/10.2307/747807

Dunn, L. M. & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test* (*Fourth Edition*). Pearson.

Duke, N. K., Lindsey, J. B., & Wise, C. N. (2023). Feeding two birds with one hand: Instructional simultaneity in early literacy education. In S. Q. Cabell, S. B. Neuman, and N. Patton Terry (Eds.), *Handbook on the science of early literacy* (pp. 186-195). Guilford.

EdReports. (2023, April 24). *Explore Reports*. https://edreports.org/reports

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

Feitelson, D., Goldstein, Z., Iraqi, J., & Share, D. L. (1993). Effects of listening to story reading on aspects of literacy acquisition in a diglossic situation. *Reading Research Quarterly*, *28*, 71-79. https://doi.org/10.2307/747817

Gillam, R. B. & Pearson, N. A. (2004). *Test of Narrative Language*. Pro-ed.

Gonzalez, J. E., Pollard-Durodola, S., Simmons, D. C., Taylor, A. B., Davis, M. J., Kim, M., & Simmons, L. (2010). Developing low-income preschoolers' social studies and science vocabulary knowledge through content-focused shared book reading. *Journal of Research on Educational Effectiveness, 4*(1), 25-52. https://doi.org/10.1080/19345747.2010.487927

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and special education*, *7*(1), 6-10.

Gray, A. M., Sirinides, P. M., Fink, R. E., & Bowden, A. B. (2022). Integrating literacy and science instruction in kindergarten: Results from the efficacy study of zoology one. *Journal of Research on Educational Effectiveness*, *15*(1), 1-27. https://doi.org/10.1080/19345747.2021.1938313

Grissmer, D., Buddin, R., Berends, M., Willingham, D., DeCoster, J., Duran, C., Hulleman, C., Murrah, W., & Evans, T. (2023). *A kindergarten lottery evaluation of Core Knowledge charter schools: Should building general knowledge have a central role in educational and*

social science research and policy?. (EdWorkingPaper: 23-755). https://doi.org/10.26300/nsbq-hb21

Grissmer, D., Grimm, K. J., Aiyer, S. M., Murrah, W. M., & Steele, J. S. (2010). Fine motor skills and early comprehension of the world: Two new school readiness indicators. *Developmental Psychology, 46,* 1008–1017. https://doi.org/10.1037/a0020104

Harris, P. A., Taylor, R., Minor, B. L., Elliott, V., Fernandez, M., O'Neal, L., McLeod, L., Delacqua, G., Delacqua, F., Kirby, J., Duda, S. N., & REDCap Consortium. (2019). The REDCap consortium: Building an international community of software platform partners. *Journal of Biomedical Informatics*, *95*, 1-10. https://doi.org/10.1016/j.jbi.2019.103208

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics*, *42*(2), 377-381.

Hiebert, E. H. (2005). In pursuit of an effective, efficient vocabulary curriculum for elementary students. In E. H. Hiebert & M. L. Kamil (Eds.), *Teaching and learning vocabulary: Bringing research to practice* (pp.249-269).

Hiebert, E. H., Goodwin, A. P., & Cervetti, G. N. (2018). Core vocabulary: Its morphological content and presence in exemplar texts. *Reading Research Quarterly, 53*(1), 29-49. https://doi.org/10.1002/rrq.183

Hirsch, E. D. (2003). Reading comprehension requires knowledge—of words and the world. *American Educator*, *27*, 10-13.

Hjetland, H. N., Brinchmann, E. I., Scherer, R., Hulme, C., & Melby-Lervåg, M. (2020). Preschool pathways to reading comprehension: A systematic meta-analytic review. *Educational Research Review*, *30*, 100323. https://doi.org/10.1016/j.edurev.2020.100323

Hoover, W. A., & Tunmer, W. E. (2022). The primacy of science in communicating advances in the science of reading. *Reading Research Quarterly*, *57*(2), 399-408. https://doi.org/10.1002/rrq.446

Hwang, H., Cabell, S. Q., & Joyner, R. E. (2022) Effects of integrated literacy and content-area instruction on vocabulary and comprehension in the elementary years: A meta-analysis. *Scientific Studies of Reading*, *26*(3), 223-249. https://doi.org/10.1080/10888438.2021.1954005

Hwang, H., Cabell, S. Q., & Joyner, R. E. (2023). Does cultivating content knowledge during literacy instruction support vocabulary and comprehension in the elementary school years? A systematic review. *Reading Psychology*, *44*(2), 145-174. https://doi.org/10.1080/02702711.2022.2141397

Hwang, H., & Duke, N. K. (2020). Content counts and motivation matters: Reading comprehension in third-grade students who are English learners. *AERA Open*, *6*(1), 1-17. https://doi.org/ 10.1177/2332858419899075

Hwang, H., Lupo, S. M., Cabell, S. Q., & Wang, S. (2021). What research says about leveraging the literacy block for learning. *Reading in Virginia*. *42*, 35-48.

Justice, L. M., Meier, J., & Walpole, S. (2005). Learning new words from storybooks: An efficacy study with at-risk kindergartners. *Language, Speech, and Hearing Services in Schools*, *36*, 17-32. https://doi.org/10.1044/0161-1461(2005/003)

Kaefer, T., Neuman, S. B., & Pinkham, A. M. (2015). Pre-existing background knowledge influences socioeconomic differences in preschoolers' word learning and comprehension. *Reading Psychology*, *36*(3), 203-231. https://doi.org/10.1080/02702711.2013.843064

Kendeou, P., van den Broek, P., White, M. J., & Lynch, J. S. (2009). Predicting reading comprehension in early elementary school: The independent contributions of oral language and decoding skills. *Journal of Educational Psychology, 101*(4)*,* 765-778. https://psycnet.apa.org/doi/10.1037/a0015956

Kim, J. S., & Burkhauser, M. A. (2022). Teaching for transfer can help young children read for understanding. *Phi Delta Kappan*, *103*(8), 20-24.

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. https://psycnet.apa.org/doi/10.1037/edu0000465

Kim, J. S., Burkhauser, M. A., Relyea, J. E., Gilbert, J. B., Scherer, E., Fitzgerald, J., Mosher, D., & McIntyre, J. (2023). A longitudinal randomized trial of a sustained content literacy intervention from first to second grade: Transfer effects on students' reading comprehension. *Journal of Educational Psychology, 115*(1), 73–98. https://doi.org/10.1037/edu0000751

Kim, J. S., Gilbert, J. B., Relyea, J. E., Rich, P., Scherer, E., Burkhauser, M. A., & Tvedt, J. N. (2023). *Time to transfer: Long-term effects of a sustained and spiraled content literacy intervention in the elementary grades.* (EdWorkingPaper: 23-769). https://doi.org/10.26300/t3c6-xh48

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

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

Kintsch, W. (2005) An overview of top-down and bottom-up effects in comprehension: The CI perspective. *Discourse Processes*, *39:2-3*, 125-128, https://doi.org/10.1080/0163853X.2005.9651676

Kook, J. F., & Greenfield, D. B. (2021). Examining variation in the quality of instructional interaction across teacher-directed activities in head start classrooms. *Journal of Early Childhood Research*, *19*(*2*), 128-144. https://doi.org/10.1177/1476718X20942956

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher, 49*(4), 241–253. https://doi.org/10.3102/0013189X20912798

Language and Reading Research Consortium. (2015). Learning to read: Should we keep things simple? *Reading Research Quarterly, 50*(2), 151-169. https://doi.org/10.1002/rrq.99

Language and Reading Research Consortium, Jiang, H., & Logan, J. (2019). Improving reading comprehension in the primary grades: Mediated effects of a language-focused classroom

intervention. *Journal of Speech, Language, and Hearing Research, 62,* 2812-2828. https://doi.org/10.1044/2019_JSLHR-L-19-0015

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

LoCasale-Crouch, J., Mashburn, A. J., Downer, J. T., & Pianta, R. C. (2008). Pre-kindergarten teachers' use of transition practices and children's adjustment to kindergarten. *Early Childhood Research Quarterly*, *23*(1), 124-139. https://doi.org/10.1016/j.ecresq.2007.06.001

Lonigan, C. J., & Milburn, T. F. (2017). Identifying the dimensionality of oral language skills of children with typical development in preschool through fifth grade. *Journal of Speech, Language, and Hearing Research, 60,* 2185-2198. https://doi.org/10.1044/2017_JSLHR-L-15-0402

Lortie-Forgues, H., & Inglis, M. (2019). Rigorous large-scale educational RCTs are often uninformative: Should we be concerned? *Educational Researcher*, *48*(3), 158–166. https://doi.org/10.3102/0013189X19832850

Marulis, L. M., & Neuman, S. B. (2010). The effects of vocabulary intervention on young Children's word learning: A meta-analysis. *Review of Educational Research, 80*(3), 300–335. https://doi.org/10.3102/0034654310377087

Marulis, L. M., & Neuman, S. B. (2013). How vocabulary interventions affect young children at risk: A meta-analytic review. *Journal of Research on Educational Effectiveness*, *6*(3), 223–262. https://doi.org/10.1080/19345747.2012.755591

McGinty, A. S., & Bevilacqua, L. (2016). *The research foundation for core knowledge language arts (CKLA).* Core Knowledge Foundation.

Mezynski, K. (1983). Issues concerning the acquisition of knowledge: Effects of vocabulary training on reading comprehension. *Review of Educational Research*, *53*(2), 253-279.

Mol, S. E., Bus, A. G., & De Jong, M. T. (2009). Interactive book reading in early education: A tool to stimulate print knowledge as well as oral language. *Review of Educational Research*, *79*(2), 979-1007. https://doi.org/10.3102/0034654309332561

Mol, S. E., Bus, A. G., de Jong, M. T., & Smeets, D. J. H. (2008). Added value of dialogic parent-child book readings: A meta-analysis. *Early Education and Development, 19,* 7-26. https://doi.org/10.1080/10409280701838603

Morrow, L. M. (1992). The impact of a literature-based program on literacy achievement, use of literature, and attitudes of children from minority backgrounds. *Reading Research Quarterly*, *27*, 251-275. https://doi.org/10.2307/747794

National Center for Education Statistics. (2022). *National assessment of educational progress, Reading assessment*. U.S. Department of Education, Institute of Education Sciences. https://www.nationsreportcard.gov/highlights/ltt/2022/

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standards (CCSS) for English language arts and literacy in history/social studies, science, and technical subjects*. Author. http://corestandards.org/

Neuman, S. B. (2006). The knowledge gap: Implications for early education. In D. K. Dickinson & S. B. Neuman (Eds.), *Handbook of early literacy* (Vol. 2, pp. 29–40). Guilford.

Neuman, S. B., & Kaefer, T. (2018). Developing low-income children's vocabulary and content knowledge through a shared book reading program. *Contemporary Educational Psychology*, *52*, 15–24. https://doi.org/10.1016/j.cedpsych.2017.12.001

Neuman, S. B., Newman, E. H., & Dwyer, J. (2011). Educational effects of a vocabulary intervention on preschoolers' word knowledge and conceptual development: A cluster-randomized trial. *Reading Research Quarterly, 46,* 249-272. https://doi.org/10.1598/RRQ.46.3.3

Neuman, S. B., Samudra, P., & Danielson, K. (2021). Effectiveness of scaling up a vocabulary intervention for low-income children, pre-K through first grade. *Elementary School Journal*, *121*(3), 385–409. https://doi.org/10.1086/712492

No Child Left Behind (NCLB) Act of 2001, Pub. L. No. 107-110, § 115, Stat. 1425 (2002).

Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology*, *98*(3), 554. https://doi.org/10.1037/0022-0663.98.3.554

Paige, D. D., Wong Fillmore, L., Cabell, S., Goldenberg, C., & Griffin, A. (2021). *Comparing reading research to program design: An examination of McGraw Hill Education's Wonders, an elementary literacy curriculum.* Student Achievement Partners. Retrieved from https://achievethecore.org/page/3364/comparing-reading-research-to-program-design-an-examination-of-mcgraw-hill-education-s-wonders-an-elementary-literacy-curriculum

Pearson, P. D., Hiebert, E. H., & Kamil, M. L. (2007). Vocabulary assessment: What we know and what we need to learn. *Reading Research Quarterly, 42*(2), 282-296. https://doi.org/10.1598/RRQ.42.2.4

Phillips, B. M. (2023). Language interventions in early childhood: Summary and implications from a multistudy program of research. In S.Q. Cabell, S.B. Neuman, and N. Patton Terry (Eds.), *Handbook on the science of early literacy* (pp. 139-150). Guilford.

Pulido, D. (2003). Modeling the role of second language proficiency and topic familiarity in second language incidental vocabulary acquisition through reading. *Language Learning*, *53*(2), 233-284. https://doi.org/10.1111/1467-9922.00217

Pulido, D. (2007). The effects of topic familiarity and passage sight vocabulary on L2 lexical inferencing and retention through reading. *Applied Linguistics*, *28*(1), 66–86. https://doi.org/10.1093/applin/aml049

Romance, N., & Vitale, M. (2017). Implications of a cognitive science model integrating literacy in science on achievement in science and reading: Direct effects in grades 3–5 with transfer to grades 6–7. *International Journal of Science and Mathematics Education*, *15*, 979-995. https://doi.org/10.1007/s10763-016-9721-2

Santoro, L. E., Baker, S. K., Chard, D. J., Howard, L. (2007). The comprehension conversation: Using intentional and purposeful text-based discourse during read alouds to promote student comprehension and vocabulary. In B. Taylor & J. Ysseldyke (Eds.), *Educational interventions for struggling readers* (pp. 109-137). Teachers College Press.

Scarborough, H. S. (2001) Connecting early language and literacy to later reading (dis)abilities: Evidency, theory, and practice. In S. B. Neuman & D. K. Dickinson (Eds.), *Handbook of early literacy research* (pp. 97-125). Guilford Press.

Schmidt, W. H., Burroughs, N. A., Zoido, P., & Houang, R. T. (2015). The role of schooling in perpetuating educational inequality: An international perspective. *Educational Researcher*, *44*(7), 371-386. https://doi.org/10.3102/0013189X15603982

Semel, E., Wiig, E. H., & Secord, W. A. (2004). *Clinical Evaluation of Language Fundamentals—Fourth Edition* (CELF-4). The Psychological Corporation.

Silverman, R. D., Johnson, E., Keane, K., & Khanna, S. (2020). Beyond decoding: A meta-analysis of the effects of language comprehension interventions on K–5 students' language and literacy outcomes. *Reading Research Quarterly, 55*(S1)*,* S207–S233. https://doi.org/10.1002/rrq.346

Simmons, D. C., Kame'enui, E. J., Harn, B., Coyne, M. D., Stoolmiller, M., Edwards Santoro, L., Smith, S. B., Beck, C. T., & Kaufman, N. K. (2007). Attributes of effective and efficient kindergarten reading intervention: An examination of instructional time and design specificity. *Journal of Learning Disabilities*, *40*(4), 331-347. https://doi.org/10.1177/00222194070400040401.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012) A 21 word solution. *SSRN.* http://dx.doi.org/10.2139/ssrn.2160588

Sonnenschein, S., Stapleton, L. M., & Benson, A. (2010). The relation between the type and amount of instruction and growth in children's reading competencies. American Educational Research Journal, 47(2), 358-389. https://doi.org/10.3102/0002831209349215

Stafura, J. Z., & Perfetti, C. A. (2017). Integrating word processing with text comprehension. In K. Cain, D. L. Compton, & R. K. Parrila (Eds.), *Theories of reading development* (pp. 9-31). John Benjamins Publishing Company.

Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analysis and a model of semantic growth. *Cognitive Science*, *29*, 41-78. https://doi.org/10.1207/s15516709cog2901_3

Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology, 38*(6)*,* 934-947. https://psycnet.apa.org/doi/10.1037/0012-1649.38.6.934

Truckenmiller, A. J., Park, J., Dabo, A., & Wu Newton, Y. C. (2019). Academic language instruction for students in grades 4 through 8: A literature synthesis. *Journal of Research on Educational Effectiveness*, *12*(1), 135–159. https://doi.org/10.1080/19345747.2018.1536773

Vadasy, P. F., & Sanders, E. A. (2012). Two-year follow-up of a code-oriented intervention for lower-skilled first-graders: The influence of language status and word reading skills on third-grade literacy outcomes. *Reading and Writing*, *26*, 821-843. https://doi.org/10.1007/s11145-012-9393-4

van Kleeck, A., Stahl, S. A., & Bauer, E. B. (2003). Reading aloud to young children as a classroom instructional activity: Insights from research and practice. In A. van Kleeck, S. A. Stahl, & E. B. Bauer (Eds.), *On Reading Books to Children* (pp. 123-147). Routledge.

Vasilyeva, M., & Waterfall, H. (2011). Variability in language development: Relation to socioeconomic status and environmental input. In S. B. Neuman, & D. K. Dickinson (Eds.), *Handbook of early literacy research: Volume 3* (pp. 36-48). Guilford Press.

Vitale, M. R., & Romance, N. R. (2012). Using in-depth science instruction to accelerate student achievement in science and reading comprehension in grades 1 - 2. *International Journal of Science and Mathematics Education, 10,* 457–472. https://doi.org/10.1007/s10763-011-9326-8

Wasik, B. A., & Hindman, A. H. (2020). Increasing preschoolers' vocabulary development through a streamlined teacher professional development intervention. *Early Childhood Research Quarterly*, *50*(1), 101-113. https://doi.org/10.1016/j.ecresq.2018.11.001

What Works Clearinghouse. (2017). *What Works Clearinghouse standards handbook, version 4.0*. U. S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance (NCEE). https://ies.ed.gov/ncee/wwc/handbooks

What Works Clearinghouse. (2021). *What Works Clearinghouse standards handbook, version 4.1*. U. S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance (NCEE). https://ies.ed.gov/ncee/wwc/handbooks

Williams, J., Stafford, K. B., Lauer, K., Hall, K., & Pollini, S. (2009). Embedding reading comprehension training in content-area instruction. *Journal of Educational Psychology, 101(1)*, 1–20. https://psycnet.apa.org/doi/10.1037/a0013152

Willingham, D. T. (2006). How knowledge helps: It speeds and strengthens reading comprehension, learning—and thinking. *American Educator, 30,* 1–12.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement*. Riverside Publishing.

Wright, T. S., & Gotwals, A. W. (2017). Supporting kindergartners' science talk in the context of an integrated science and disciplinary literacy curriculum. *The Elementary School Journal, 117*(3), 513–537. https://doi.org/10.1086/690273

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

Zucker, T. A., Cabell, S. Q., Petscher, Y., Mui, H., Landry, S. H., & Tock, J. (2021). Teaching Together: Pilot study of a tiered language and literacy intervention with Head Start teachers and linguistically diverse families. *Early Childhood Research Quarterly*, *54*, 136-152. doi:10.1016/j.ecresq.2020.09.001

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*(2), 162-185. https://psycnet.apa.org/doi/10.1037/0033-2909.123.2.162

Table 1

Demographics by study, treatment condition, and the overall sample

	total study 1 sample (N = 544)		total study 2 sample (N = 647)		total control (N = 626)		total treatment (N = 565)		total sample (N = 1191)	
	n	%	n	%	n	%	n	%	n	%
Race [a]										
White, non-Hispanic	22	4.04	133	20.56*	85	13.58	70	12.39	155	13.01
American Indian/ Alaska Native	5	0.92	5	0.77	4	0.64	6	1.06	10	0.84
Asian	5	0.92	9	1.39	9	1.44	3	0.53	14	1.18
Black	412	75.74	331	51.16*	378	60.38	365	64.60	743	62.38
Hispanic	94	17.28	129	19.94*	119	19.01	104	18.41	223	18.72
Native Hawaiian/ Pacific Islander	4	0.74	2	0.31	5	0.80	1	0.18	6	0.50
Two or More	9	1.65	31	4.79*	23	3.67	17	3.01	40	3.36
Gender										
Female	280	51.47	317	43.28	321	51.28	276	48.85	597	50.13
ELL status [a]										
yes	78	14.34	91	14.06	93	14.86	76	13.45	169	14.19
SPED [a]										
yes	52	9.56	68	10.51	56	8.95	64	11.33	120	10.08
ED [a]										
yes	--	--	379	58.58	--	--	--	--	--	--

Note: [a] 7 students had missing data from study 2; * significant difference between cohorts (there were no significant differences by treatment); ELL status = English language learner status, SPED = qualified for special education; ED = economic disadvantage

Table 2

Pretest measures by study, treatment condition, and the overall sample

	total study 1 sample (N = 544)		total study 2 sample (N = 647)		total control (N = 626)		total treatment (N = 565)		total sample (N = 1191)	
	n	mean (sd)	n	mean (sd)	n	mean (sd)	n	mean (sd)	n	mean (sd)
Age at pretest	543	5.68 (.33)	647	5.64 (.31)	626	5.64 (.32)	564	5.67 (.32)	1190	5.66 (.32)
PPVT pretest	529	84.25 (22.05)	639	80.57 (25.50)	616	83.02 (23.19)	552	81.36 (24.98)	1168	82.24 (24.06)
TNL pretest	539	15.54 (7.47)	639	14.58 (8.06)	618	15.06 (7.63)	560	14.97 (8.01)	1178	15.02 (7.81)
CELF SS pretest	540	16.65 (4.59)	641	16.05 (4.82)	620	16.34 (4.61)	561	16.30 (4.86)	1181	16.32 (4.73)
WJ PVT pretest	537	14.91 (4.15)	642	15.19 (4.60)	618	15.14 (4.31)	561	14.99 (4.91)	1179	15.07 (4.40)
WJ SC pretest	536	10.80 (2.22)	642	10.27 (2.62)*	618	10.60 (2.30)	560	10.42 (2.62)	1178	10.51 (2.46)
WJ SS pretest	537	9.83 (2.57)	642	9.84 (2.82)	618	9.84 (2.68)	561	9.84 (2.75)	1179	9.84 (2.71)

Note: * significant difference between studies (there were no significant differences by treatment)

Note: PPVT = Peabody Picture Vocabulary Test- Fourth Edition, TNL = Test of Narrative Language, CELF SS = Clinical Evaluation of Language Fundamentals- Fourth Edition Sentence Structure subtest; WJ PVT = Woodcock-Johnson III Tests of Achievement Picture Vocabulary subtest; WJ SC = Woodcock-Johnson Tests of Achievement Science subtest; WJ SS = Woodcock-Johnson Tests of Achievement Social Studies subtest

Table 3

Means, standard deviations, and correlations

Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1. age at pretest	5.66	0.32																		
2. TNL pre	15.02	7.81	.17																	
3. TNL post	19.43	7.54	.15	.76																
4. PPVT pre	82.23	24.05	.17	.68	.71															
5. PPVT post	97.15	22.42	.16	.66	.76	.83														
6. CELF pre	16.32	4.73	.13	.68	.66	.67	.66													
7. CELF post	19.66	4.43	.14	.63	.73	.66	.74	.66												
8. WJ PVT pre	15.07	4.40	.10	.64	.67	.76	.76	.60	.61											
9. WJ PVT post	16.74	4.30	.09	.62	.69	.75	.79	.58	.64	.85										
10. WJ Sci pre	10.52	2.46	.11	.64	.64	.73	.69	.62	.58	.74	.70									
11. WJ Sci post	11.81	2.33	.10	.61	.69	.67	.70	.55	.63	.67	.71	.67								
12. WJ SS pre	9.84	2.71	.14	.70	.70	.73	.71	.63	.62	.74	.70	.68	.64							
13. WJ SS post	11.19	2.62	.14	.64	.73	.71	.73	.60	.66	.70	.72	.65	.71	.71						
14. CSVT post	6.98	2.94	.12	.45	.51	.49	.53	.39	.45	.42	.47	.40	.45	.43	.41					
15. CSKT PL post	1.56	1.79	.07	.29	.35	.28	.32	.22	.26	.30	.31	.26	.29	.28	.27	.32				
16. CSKT NA post	0.79	1.66	.08	.20	.28	.20	.21	.18	.20	.20	.24	.18	.23	.19	.19	.41	.33			
17. treatment	0.47	0.50	.04	-.01	.02	-.03	-.02	-.01	-.02	-.02	.02	-.04	.03	.00	-.01	.31	.14	.47		
18. cohort	0.54	0.50	-.06	-.06	.06	-.08	.00	-.06	-.05	.03	.01	-.11	-.04	.00	-.12	.11	.28	.15	.09	
19. ELL	0.14	0.35	.02	-.40	-.43	-.45	-.45	-.39	-.37	-.54	-.50	-.45	-.40	-.50	-.45	-.17	-.10	-.09	-.02	.00

Table 4

Results of main effects models for standardized outcomes

	PPVT			TNL			CELF SS			WJ PVT			WJ SC			WJ SS		
Fixed effects	est	se	p	est	se	p	est	se	p	est	se	p	est	se	p	est	se	p
intercept	-.12	.03	.001	-.13	.06	.031	.03	.04	.472	-.03	.04	.406	-.07	.06	.190	.08	.05	.095
pretest	.83	.02	<.001	.76	.02	<.001	.66	.02	<.001	.85	.02	<.001	.67	.02	<.001	.72	.02	<.001
treatment	.01	.04	.874	.02	.06	.725	-.05	.05	.301	.09	.04	.016	.11	.06	.054	.00	.05	.999
study	.16	.04	<.001	.22	.06	.001	-.01	.05	.870	-.04	.04	.243	.05	.06	.414	-.23	.05	<.001
school size	.05	.04	.220	.00	.07	.981	-.03	.05	.534	-.01	.04	.806	-.04	.06	.467	.10	.05	.047
Random effects	Variance			Variance			Variance			Variance			Variance			Variance		
School	.002			.027			.000			.003			.015			.006		
Residual	.296			.379			.565			.278			.524			.469		

Note: est = estimate, se = standard error, PPVT = Peabody Picture Vocabulary Test- Fourth Edition, TNL = Test of Narrative Language, CELF SS = Clinical Evaluation of Language Fundamentals- Fourth Edition Sentence Structure subtest; WJ PVT = Woodcock-Johnson III Tests of Achievement Picture Vocabulary subtest; WJ SC = Woodcock-Johnson Tests of Achievement Science subtest; WJ SS = Woodcock-Johnson Tests of Achievement Social Studies subtest

Table 5

Results of main effects models for proximal outcomes

	CSVT			CSKT PL			CSKT NA		
Fixed effects	est	se	p	est	se	p	est	se	p
intercept	-.46	.06	<.001	-.49	.06	<.001	-.55	.09	<.001
PPVT pretest	.51	.02	<.001	.32	.03	<.001	.22	.03	<.001
treatment	.63	.06	<.001	.26	.07	<.001	.93	.09	<.001
study	.26	.06	<.001	.60	.07	<.001	.23	.09	.016
school size	.03	.07	.686	.09	.07	.189	-.09	.10	.362
Random effects	Variance			Variance			Variance		
School	.020			.014			.067		
Residual	.619			.803			.654		

Note: est = estimate, se = standard error, CSVT = Curriculum-Specific Vocabulary Test; CSKT PL = Curriculum-Specific Knowledge Test – Plants; CSKT NA = Curriculum-Specific Knowledge Test – Native Americans

Table 6

Results of interaction effects by initial skill on standardized outcomes

	PPVT			TNL			CELF			WJPVT			WJSC			WJSS		
Fixed effects	est	se	p	est	se	p	est	se	p	est	se	p	est	se	p	est	se	p
intercept	-.12	.04	.001	-.13	.06	.032	.03	.04	.480	-.03	.04	.412	-.08	.06	.178	.08	.05	.097
pretest	.85	.02	<.001	.77	.03	<.001	.69	.03	<.001	.83	.02	<.001	.71	.03	<.001	.76	.03	<.001
treatment	.01	.04	.863	.02	.06	.718	-.05	.05	.307	.09	.04	.017	.11	.06	.053	.00	.05	.978
study	.16	.04	<.001	.22	.06	.001	-.01	.05	.877	-.04	.04	.250	.05	.06	.405	-.23	.05	<.001
school size	.05	.04	.232	.00	.07	.970	-.03	.05	.539	-.01	.04	.803	-.04	.06	.468	.10	.05	.044
pretest x treatment	-.03	.03	.418	-.03	.04	.431	-.06	.05	.195	.03	.03	.292	-.07	.04	.135	-.07	.04	.088
Random effects	Variance			Variance			Variance			Variance			Variance			Variance		
School	.002			.028			.000			.003			.015			.006		
Residual	.296			.379			.564			.278			.523			.468		

Note: est = estimate, se = standard error, PPVT = Peabody Picture Vocabulary Test- Fourth Edition, TNL = Test of Narrative Language, CELF SS = Clinical Evaluation of Language Fundamentals- Fourth Edition Sentence Structure subtest; WJ PVT = Woodcock-Johnson III Tests of Achievement Picture Vocabulary subtest; WJ SC = Woodcock-Johnson Tests of Achievement Science subtest; WJ SS = Woodcock-Johnson Tests of Achievement Social Studies subtest

Table 7

Results of interaction effects by initial skill on proximal outcomes

	CSVT			CSKT PL			CSKT NA		
Fixed effects	est	se	p	est	se	p	est	se	p
Intercept	-.46	.06	<.001	-.48	.06	<.001	-.54	.08	<.001
PPVT pretest	.40	.04	<.001	.25	.04	<.001	.02	.04	.670
treatment	.63	.06	<.001	.25	.06	<.001	.92	.09	<.001
cohort	.26	.06	<.001	.60	.07	<.001	.23	.09	.013
School size	.03	.07	.642	.09	.07	.174	-.08	.10	.396
PPVT pretest x treatment	.21	.05	<.001	.13	.06	.023	.39	.05	<.001
Random effects		Variance			Variance			Variance	
School		.019			.013			.064	
Residual		.609			.800			.619	

Note: est = estimate, se = standard error, CSVT = Curriculum-Specific Vocabulary Test; CSKT PL = Curriculum-Specific Knowledge Test – Plants; CSKT NA = Curriculum-Specific Knowledge Test – Native Americans

Table 8

Results of interaction effects by English learner status on standardized outcomes

	PPVT			TNL			CELF			WJPVT			WJSC			WJSS		
Fixed effects	est	se	p	est	se	p	est	se	p	est	se	p	est	se	p	est	se	p
Intercept	-.08	.04	.025	-.07	.06	.257	.08	.05	.075	-.02	.04	.585	-.02	.06	.703	.13	.05	.011
pretest	.80	.02	<.001	.70	.02	<.001	.61	.02	<.001	.83	.02	<.001	.63	.03	<.001	.66	.02	<.001
treatment	.00	.04	.927	.01	.06	.897	-.05	.05	.322	.09	.04	.019	.09	.06	.151	-.02	.05	.770
ELL	-.23	.07	.001	-.46	.08	<.001	-.36	.09	<.001	-.07	.07	.291	-.35	.09	<.001	-.36	.09	<.001
cohort	.15	.04	<.001	.22	.06	.001	-.01	.05	.848	-.04	.04	.263	.04	.06	.522	-.23	.05	<.001
size	.06	.04	.121	.03	.06	.698	-.01	.05	.838	-.01	.04	.888	-.02	.06	.755	.12	.05	.025
Treatment x ELL	.05	.09	.585	.08	.11	.490	.01	.13	.933	-.03	.09	.728	.13	.13	.314	.09	.12	.448
Random effects	Variance			Variance			Variance			Variance			Variance			Variance		
School	.002			.026			.000			.003			.016			.007		
Residual	.292			.363			.550			.276			.518			.460		

Note: est = estimate, se = standard error, PPVT = Peabody Picture Vocabulary Test- Fourth Edition, TNL = Test of Narrative Language, CELF SS = Clinical Evaluation of Language Fundamentals- Fourth Edition Sentence Structure subtest; WJ PVT = Woodcock-Johnson III Tests of Achievement Picture Vocabulary subtest; WJ SC = Woodcock-Johnson Tests of Achievement Science subtest; WJ SS = Woodcock-Johnson Tests of Achievement Social Studies subtest

Table 9

Results of interaction effects by English learner status on proximal outcomes

	CSVT			CSKT PL			CSKT NA		
Fixed effects	est	se	p	est	se	p	est	se	p
Intercept	-.52	.06	<.001	-.51	.06	<.001	-.59	.09	<.001
PPVT pretest	.55	.03	<.001	.35	.03	<.001	.23	.03	<.001
treatment	.68	.06	<.001	.25	.07	.001	.99	.09	<.001
ELL	.38	.10	<.001	.16	.12	.180	.30	.11	.008
cohort	.28	.06	<.001	.61	.06	<.001	.25	.09	.009
School size	.00	.06	.993	.08	.07	.229	-.11	.10	.265
Treatment x ELL	-.28	.14	.052	.09	.16	.585	-.41	.16	.009
Random effects		Variance			Variance			Variance	
School		.014			.010			.063	
Residual		.614			.802			.651	

Note: est = estimate, se = standard error, CSVT = Curriculum-Specific Vocabulary Test; CSKT PL = Curriculum-Specific Knowledge Test – Plants; CSKT NA = Curriculum-Specific Knowledge Test – Native Americans

Table 10

Results of interaction effects by economic disadvantage on standardized outcomes (Study 2 only)

	PPVT			TNL			CELF SS			WJ PVT			WJ SC			WJ SS		
Fixed effects	est	se	p	est	se	p	est	se	p	est	se	p	est	se	p	est	se	p
Intercept	.04	.06	.497	-.08	.07	.308	.13	.07	.072	.01	.05	.896	-.02	.08	.772	.05	.06	.397
pretest	.82	.02	<.001	.77	.03	<.001	.66	.03	<.001	.85	.02	<.001	.67	.03	<.001	.75	.03	<.001
treatment	.03	.07	.712	.12	.10	.242	-.03	.09	.729	.16	.07	.026	.21	.10	.039	.03	.09	.749
ED	-.11	.07	.091	.07	.07	.370	-.16	.09	.071	-.11	.06	.075	-.03	.09	.707	-.16	.08	.044
cohort	.04	.06	.463	.11	.08	.208	-.05	.07	.448	-.03	.05	.550	-.08	.08	.321	.06	.06	.346
size	.00	.09	.970	-.17	.10	.103	-.01	.12	.911	-.08	.09	.354	-.17	.12	.168	-.01	.11	.922
Treatment x ED	.04	.06	.497	-.08	.07	.308	.13	.07	.072	.01	.05	.896	-.02	.08	.772	.05	.06	.397
Random effects	Variance			Variance			Variance			Variance			Variance			Variance		
School	.003			.020			.000			.003			.009			.000		
Residual	.297			.359			.555			.270			.529			.448		

Note: est = estimate, se = standard error, PPVT = Peabody Picture Vocabulary Test- Fourth Edition, TNL = Test of Narrative Language, CELF SS = Clinical Evaluation of Language Fundamentals- Fourth Edition Sentence Structure subtest; WJ PVT = Woodcock-Johnson III Tests of Achievement Picture Vocabulary subtest; WJ SC = Woodcock-Johnson Tests of Achievement Science subtest; WJ SS = Woodcock-Johnson Tests of Achievement Social Studies subtest

Table 11

Results of interaction effects by economic disadvantage on proximal outcomes (Study 2 only)

	CSVT			CSKT PL			CSKT NA		
Fixed effects	est	se	p	est	se	p	est	se	p
Intercept	-.27	.09	.004	-.10	.09	.289	-.49	.10	<.001
PPVT pretest	.53	.03	<.001	.34	.04	<.001	.24	.04	<.001
treatment	.80	.12	<.001	.28	.12	.023	1.09	.13	<.001
ED	-.13	.09	.170	-.12	.11	.269	-.01	.10	.960
cohort	.01	.10	.951	.13	.09	.148	-.02	.11	.866
School size	-.21	.13	.106	-.02	.16	.884	-.22	.14	.123
Treatment x ED	-.27	.09	.004	-.10	.09	.289	-.49	.10	<.001
Random effects		Variance			Variance			Variance	
School		.028			.003			.031	
Residual		.566			.853			.684	

Note: est = estimate, se = standard error, CSVT = Curriculum-Specific Vocabulary Test; CSKT PL = Curriculum-Specific Knowledge Test – Plants; CSKT NA = Curriculum-Specific Knowledge Test – Native Americans

Figure Titles

Figure 1. *Theory of Change for CKLA: Knowledge*

Figure 2. *Condition x Initial Language Interaction Effects on Proximal Vocabulary Knowledge*

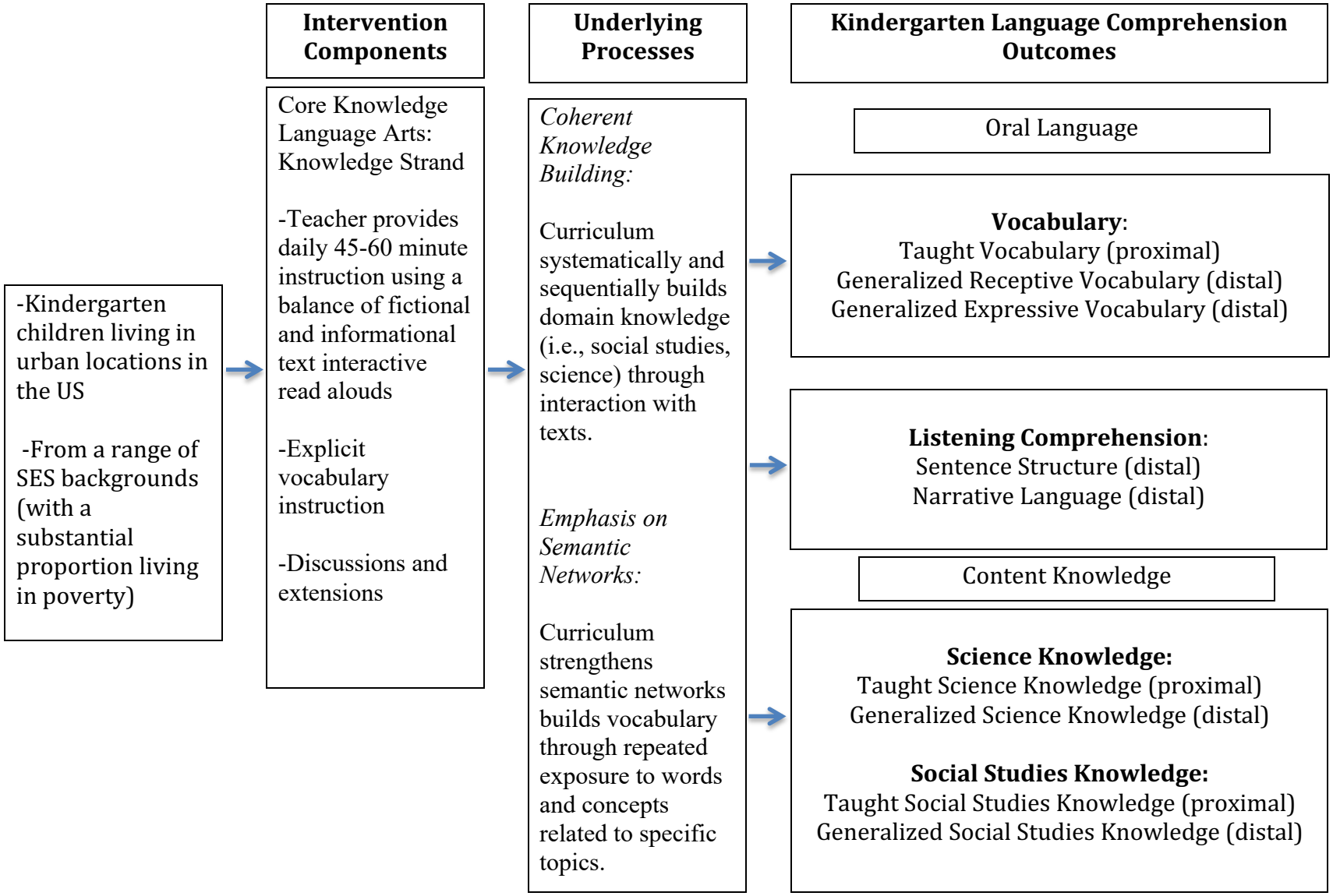
Figure 3. *Condition x Initial Language Interaction Effects on Proximal Science Knowledge*

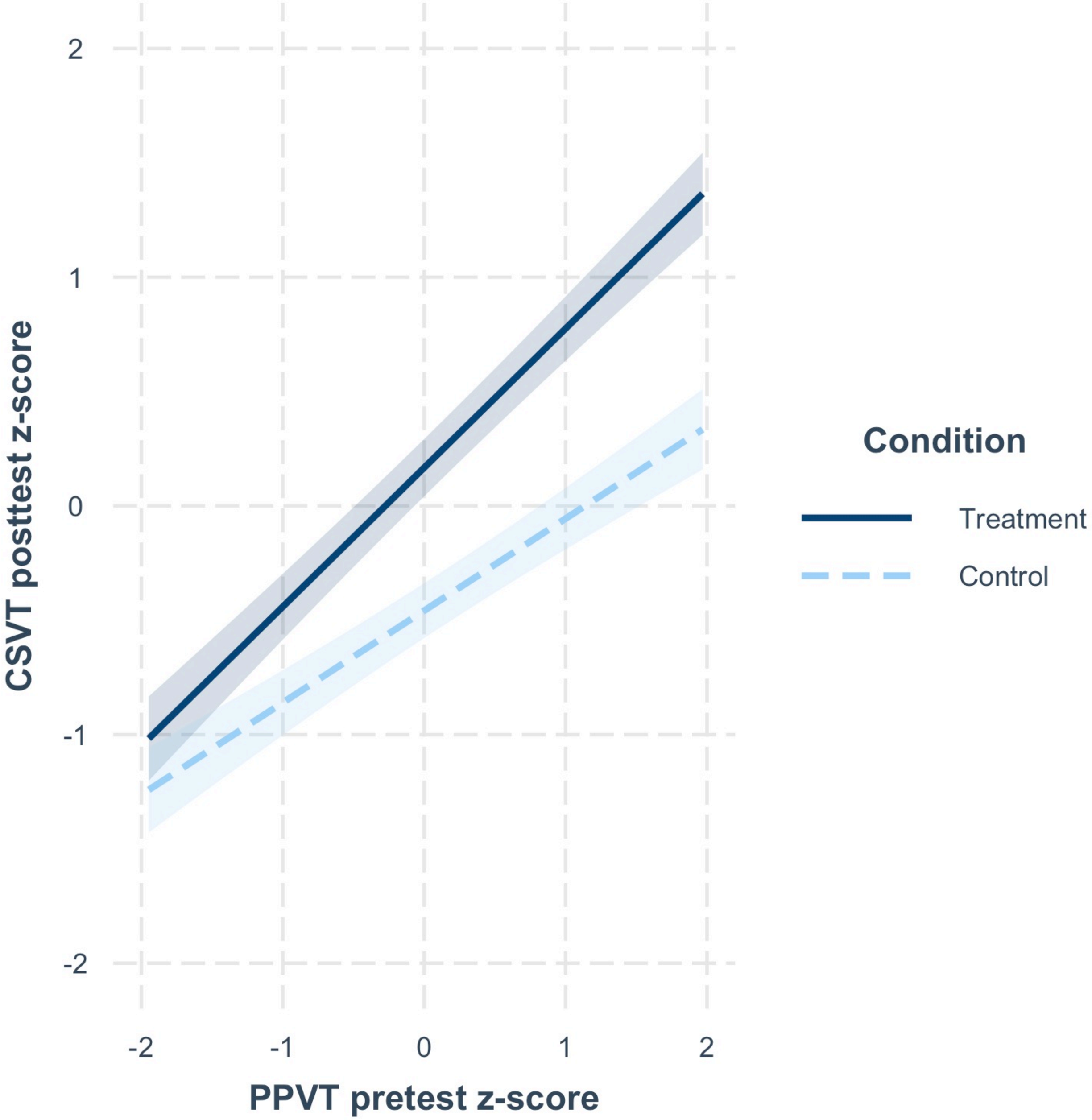
Figure 4. *Condition x Initial Language Interaction Effects on Proximal Social Studies Knowledge*

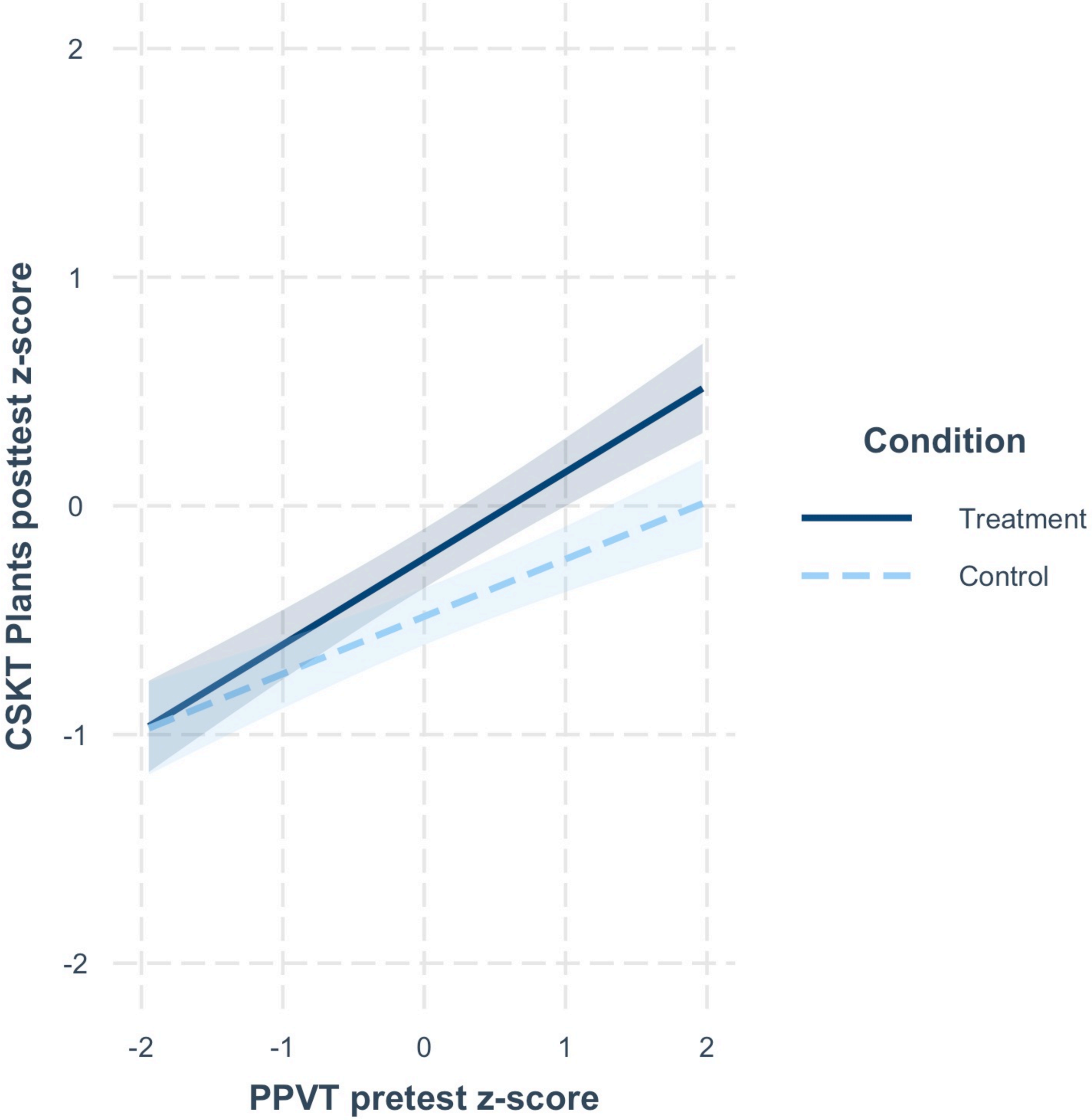
Figure 5. *Condition x English Language Learner Status Effects on Proximal Vocabulary Knowledge*

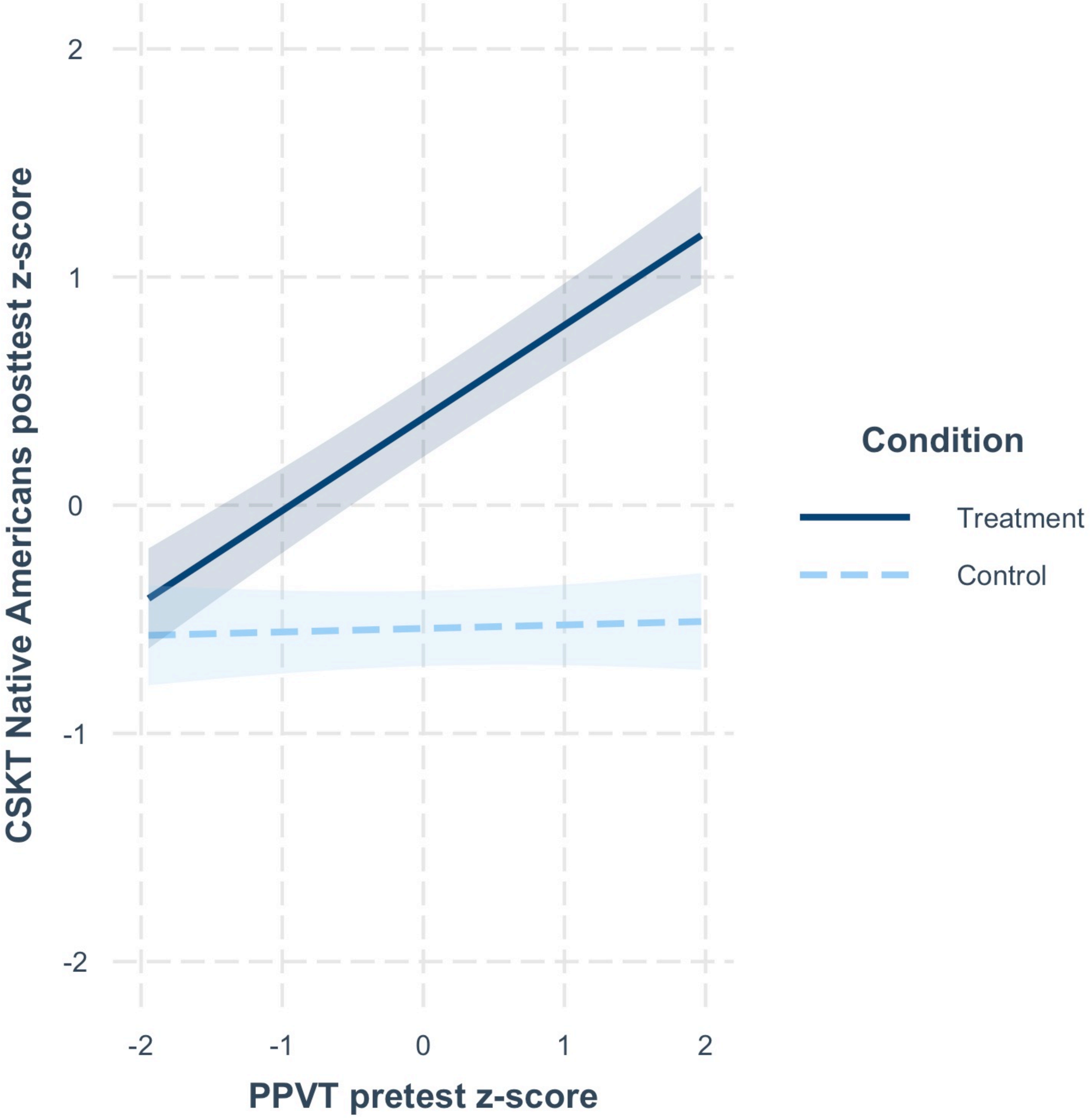
Figure 6. *Condition x Initial Language Interaction Effects on Proximal Social Studies Knowledge*

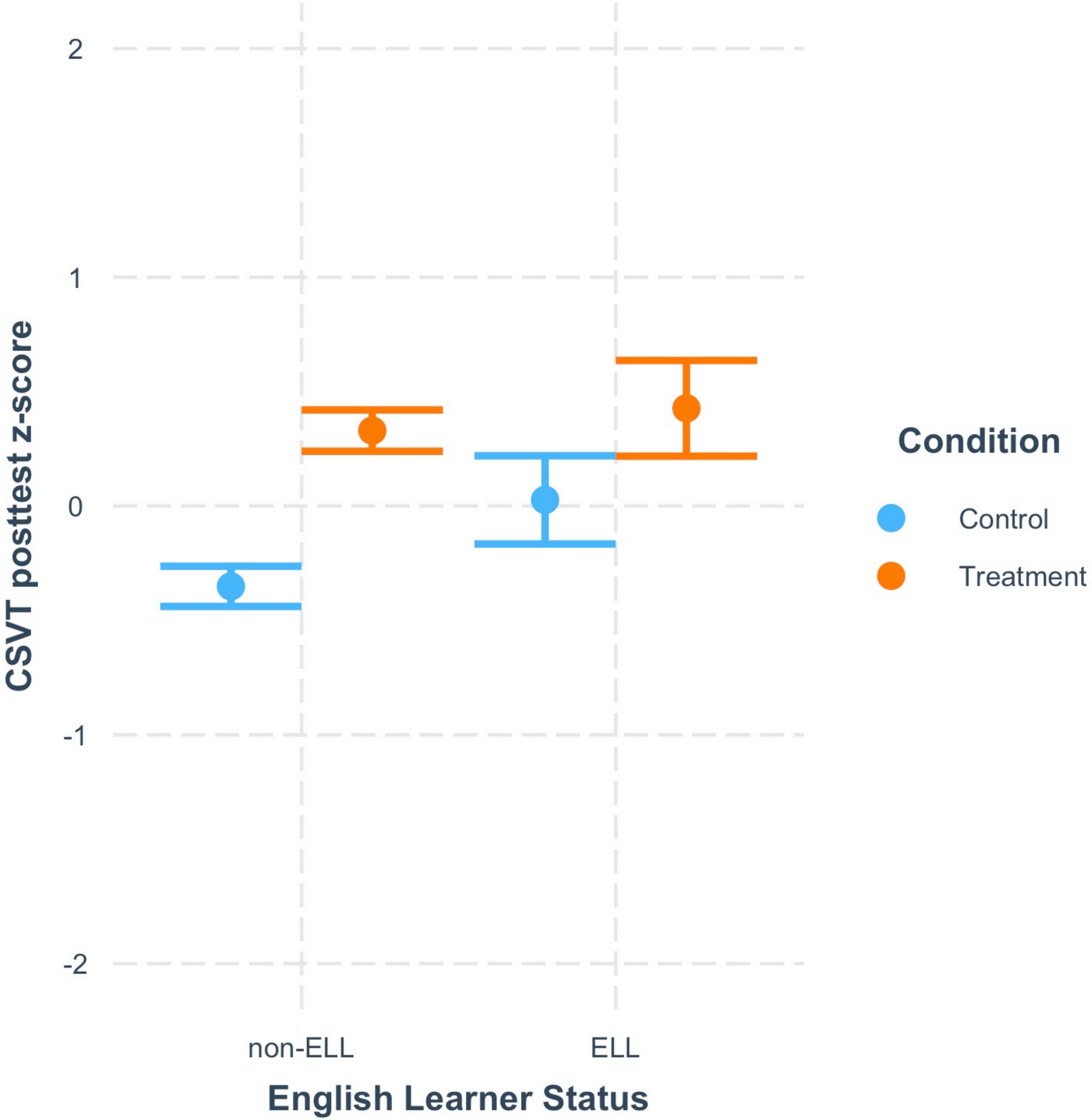
Figure 1. ***Change Model for CKLA: Knowledge***

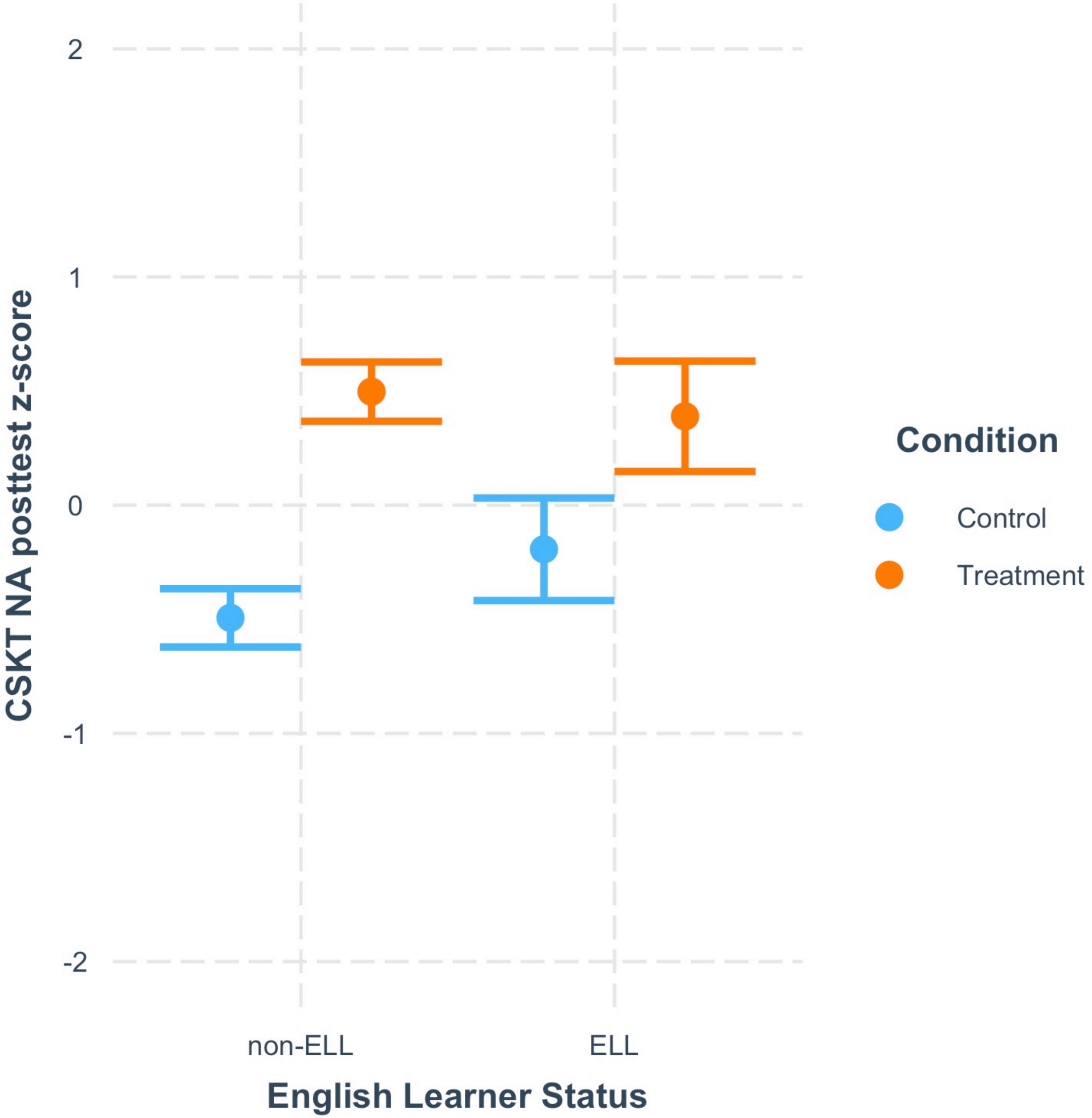



CSVT posttest z-score
2
1
0
-1
-2
-2
-1
0
1
2
PPVT pretest z-score
Condition
Treatment
Control


CSKT Plants posttest z-score
2
1
0
-1
-2
-2
-1
0
1
2
PPVT pretest z-score
Condition
Treatment
Control


CSKT Native Americans posttest z-score
2
1
0
-1
-2
-2
-1
0
1
2
PPVT pretest z-score
Condition
Treatment
Control


2
1
0
-1
-2
CSVT posttest z-score
non-ELL
ELL
English Learner Status
Condition
Control
Treatment


2
1
0
-1
-2
CSKT NA posttest z-score
non-ELL
ELL
English Learner Status
Condition
Control
Treatment