

The Efficacy of Comprehension and Vocabulary Focused Professional Development on English Learners' Literacy

Sharon Vaughn¹, Elizabeth Swanson¹, Anna-Maria Fall¹, Greg Roberts¹, Philip Capin¹,
Elizabeth A. Stevens², and Alicia A. Stewart³

¹ The Meadows Center for Preventing Educational Risk, Department of Special Education, The University of Texas at Austin

² Department of Learning Sciences, Georgia State University

³ Department of Special Education and Interventions, Central Connecticut State University

This study reports the effects of a distributed professional development model emphasizing reading comprehension and vocabulary practices in social studies on the content knowledge, vocabulary, and reading comprehension outcomes of upper elementary students identified as English learners (ELs). Schools were randomly assigned to one of three conditions: researcher-supported professional development (PD), school-supported PD, or business as usual (BAU; typical instruction). Findings from a prior randomized control trial revealed significant effects for both treatment conditions when compared with the BAU group for content knowledge, vocabulary learning and reading comprehension in content (Capin et al., 2021). This analysis addressed three related follow-up questions: (a) Does treatment affect ELs and non-ELs differently? (b) Does treatment affect students differently depending on the school-wide percentage of ELs? (c) Does treatment affect EL students differently from non-ELs depending on the school-wide percentage of ELs? Findings revealed that ELs in treatment conditions outperformed ELs in the BAU condition and school level percentage of ELs had an impact on EL vocabulary performance.

Educational Impact and Implications Statement

This study demonstrates that a distributed professional development model focused on developing fourth-grade teachers' use of integrated vocabulary and reading comprehension practices within social studies instruction significantly improves English learners' (ELs) and non-ELs' performance on measures of social studies knowledge and vocabulary. The study also reported that the effects of professional development on the performance of ELs and non-ELs varied on a general measure of vocabulary based on the school-wide proportion of ELs. These findings suggest that professional development for teachers focused on evidence-based practices for content-area vocabulary and reading comprehension demonstrates positive impact for students though the impact may vary for ELs and non-ELs based on the proportion of ELs at their school site.

Keywords: reading, second language learning, effective teaching

English learners' (ELs) success in school is among the highest of priorities among educational stakeholders. There are currently more than 5 million ELs in schools in the United States and the number is increasing every year with some states identifying as many as one in four elementary students as ELs (Kena et al., 2015). Unfortunately, many teachers report they are underprepared

to teach the range of diverse learners—including ELs—in their classroom (August & Shanahan, 2006; Calderon et al., 2011). ELs present unique assets and challenges and with strategic instruction, their knowledge and skills can be integrated and enhanced. In particular, ELs benefit from practices that develop their vocabulary, language, and knowledge within content instruction, thus helping

This article was published Online First November 11, 2021.

Sharon Vaughn  <https://orcid.org/0000-0001-8305-5549>

Elizabeth Swanson  <https://orcid.org/0000-0002-2716-4078>

Anna-Maria Fall  <https://orcid.org/0000-0002-6257-6684>

Greg Roberts  <https://orcid.org/0000-0001-6333-7442>

Philip Capin  <https://orcid.org/0000-0003-4955-9879>

Elizabeth A. Stevens  <https://orcid.org/0000-0002-8412-1111>

Alicia A. Stewart  <https://orcid.org/0000-0001-6770-5046>

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A150407 to The University of Texas at Austin. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Correspondence concerning this article should be addressed to Sharon Vaughn, The Meadows Center for Preventing Educational Risk, Department of Special Education, The University of Texas at Austin, 1912 Speedway, SZB 5.110, Austin, TX 78712, United States. Email: srvaughn@austin.utexas.edu

ELs meet the dual challenge of language enhancement and knowledge acquisition.

We have been designing approaches to professional development (PD) for teachers that are feasible to implement, based on the most impactful available research, and are integrated within content learning streams that provide a platform for enhancing vocabulary, comprehension, and content-area knowledge of students (Hairrell et al., 2011; Capin et al., 2021; Simmons et al., 2010). We have also been mindful of how to implement these instructional routines in ways that are most responsive to the range of learners in the classroom including ELs. This introduction provides a conceptual background of the research from which this PD and the instructional routines that are highlighted for students who are ELs were derived. We also provide a rationale for the research questions for this study and how these questions are an extension of previous research.

Background

PD for teachers often has a “dead-end” feel in which teachers participate in a learning environment ostensibly designed to enhance their knowledge and skills with the intent that teachers will immediately integrate this learning into their instructional routines (Adey, 2004). While the intent and need for the PD are quite necessary, the intended outcome is rarely reached—that is, instructional adjustments that significantly impact student learning outcomes are rarely attained (Babinski et al., 2018). Identifying a set of instructional practices that teachers can readily assimilate into their instruction that meets the needs of the range of diverse learners in their classrooms is a tall order. While teachers appreciate knowing about instructional practices that are effective for a specific subgroup of their students (e.g., students with reading problems), they consistently report that differentiating instruction for the range of learners in their classrooms is challenging and that they are looking for impactful practices that meet the needs of this increasingly diverse group (Boardman et al., 2005). STRIVE is an evidence-based PD model designed to support teachers to integrate a set of instructional practices associated with improved vocabulary, comprehension, and content learning with upper elementary students (Capin et al., 2021; Simmons et al., 2010).

Professional Development

Though frequently utilized in school settings, a one-time PD delivered to a large group of teachers is unlikely to influence classroom practice (Darling-Hammond et al., 2017; Richardson, 2003). For change in classroom practice to occur that is potent enough to impact student outcomes, teachers benefit from structured opportunities to engage with colleagues around a set of targeted practices (Opfer et al., 2011; Parise & Spillane, 2010). When structured follow-up opportunities are provided, they are associated with beneficial outcomes on teachers’ knowledge, skills, and practices (Garet et al., 2001; Garet et al., 2008) as well as student outcomes (Basma & Savage, 2018; Wasik & Hindman, 2011). Studies reveal that students’ comprehension outcomes are greater when these follow-up opportunities for teachers include distributed practice and engagement on how to use the practices with materials in their classrooms (Capin et al., 2021; Klingner et al., 2004; Porche et al., 2012; Stevens et al., 2020).

Instructional Practices for the Diverse Classroom

Acquiring vocabulary and background knowledge related to disciplinary knowledge (e.g., history, science) is a critical aspect of schooling—particularly in upper elementary and secondary grades. However, understanding these texts and learning from them is often challenging for many students—including ELs (Ahmed et al., 2016; Baumann, 2009; Cervetti et al., 2009; Cunningham & Stanovich, 1998; Elleman et al., 2009; Lesaux et al., 2010). Thus, many teachers bypass text reading in favor of other approaches to teaching content knowledge, including more traditional approaches such as lecturing using multimedia Microsoft PowerPoint presentations (Swanson et al., 2020; Swanson et al., 2016; Wexler et al., 2016). Recognizing the valuable role of text reading in the development of vocabulary, knowledge, and comprehension, we sought to refine a set of text based instructional approaches aimed at enhancing these outcomes in content area instruction, specifically social studies (Hairrell et al., 2011; Simmons et al., 2010).

In selecting and developing the foci for the set of instructional approaches for STRIVE, we were informed by the construction integration model of reading comprehension (Kintsch, 1998) and the content model of reading (McKeown et al., 2009) combined with empirical studies and syntheses (e.g., Gajria et al., 2007) to identify a set of vocabulary and reading comprehension practices that support students learning new knowledge from text reading. The first iteration of STRIVE PD included training fourth-grade social studies teachers in either comprehension practices or vocabulary practices (Simmons et al., 2010). Findings from the first randomized controlled trial (RCT) provided evidence that students whose teachers received PD on the comprehension practices or the vocabulary practices outperformed students whose teachers did not attend STRIVE PD on a measure of social studies content knowledge. In addition, students whose teachers received PD on vocabulary practices outperformed those whose teachers received PD on comprehension practices on a measure of vocabulary. From this study, we learned that neither comprehension practices nor vocabulary practices alone was enough to impact more distal reading outcomes. As a result, we combined the vocabulary and comprehension practices into one parsimonious set of lessons. In the next RCT (Hairrell et al., 2011), teachers participated in STRIVE PD focused on implementing the combined set of comprehension plus vocabulary practices and we examined the role of fidelity and teacher quality on student reading outcomes. Findings indicated that fidelity and teacher quality were significantly and positively related to student outcomes on a standardized measure of reading comprehension. In other words, when teachers implemented the practices as they were designed and with high levels of instructional quality, students performed better on a distal reading comprehension measure. These studies combined taught us that not only do theoretically and empirically grounded PD and instructional practice selection matter, but that fidelity and instructional quality matter as well.

The STRIVE Model

In prior studies that informed the current investigation (Hairrell et al., 2011; Simmons et al., 2010), researchers used a distributed PD model in which teachers learned and practiced a set of

instructional practices over the course of several meetings with colleagues and researchers with a goal to refine and improve their implementation of the practices. [Simmons et al. \(2010\)](#) referred to this approach to PD and the set of instructional practices as STRIVE. The STRIVE approach provides teachers with an initial workshop focused on evidence and use of vocabulary and reading comprehension practices within social studies texts ([Klingner et al., 2004](#); [Porche et al., 2012](#)). The distributed PD, however, includes follow-up opportunities using small group teacher study team meetings examining the curriculum materials and lessons to produce high levels of confidence in teachers using the practices and fidelity to the practice ([Gersten et al., 2010](#)). In the initial efficacy trial ([Simmons et al., 2010](#)), teachers were assigned to deliver either vocabulary practices, comprehension practices, or BAU practices within social studies across the course of 18 weeks. Findings indicated that students whose teachers participated in either the vocabulary or comprehension PD outperformed students in the BAU condition on multiple reading outcomes.

The STRIVE model reflects current research on PD aimed to (a) increase teachers' knowledge of the instructional practices and classroom materials, (b) impact implementation of instructional practices within content learning, or how to implement a set of evidence-based instructional practices with fidelity within the classroom, and (c) improve outcomes for students in vocabulary and content learning. The instructional practices and how they were taught is described in more detail in the method section of this paper; however, we consider these practices as representing an approach to learning that might be considered "universal" in that it has elements associated with improved outcomes for a range of learners, including ELs ([Vaughn et al., 2017](#)). These approaches to learning that are likely to be beneficial to ELs include: (a) identifying and teaching content meaningful words through pictures and brief videos; (b) providing multiple opportunities for practice using the lesson materials through speaking, reading, writing, and interacting with peers; (c) providing students with models and cues for reflecting, problem solving, and extending learning through instructional activities; and (d) supporting students in linking what they know, what they are learning, and reviewing and building learning constructs over lesson use. Because the STRIVE PD is distributed over time, teachers are frequently supported and guided through implementation. The STRIVE PD model and materials provide teachers with: (a) inclusive lessons that address a limited number of high impact instructional practices, and (b) practice opportunities distributed over time so that the lessons could be implemented readily in their classrooms.

Current Study

In a recent randomized control trial ([Capin et al., 2021](#)), investigators addressed whether the transfer of the distributed STRIVE PD model to a key educator in the school (i.e., school-supported PD [SPD]) would be associated with differential effects from a distributed STRIVE PD model provided by a member of the research team (i.e., research-supported PD [RPD]). The rationale was that if SPD was associated with similar impacts as RPD then the likelihood for sustaining the practices over time might be improved recognizing that impacts often dissipate when the researchers depart the school setting ([Datnow, 2002](#); [Giles, 2006](#); [McLaughlin & Mitra, 2001](#)). An SPD approach might promote

school-based "ownership" of the practices and enhance implementation and perhaps student outcomes ([Coburn, 2003](#)).

Results from the main effects study ([Capin et al., 2021](#)) revealed that students whose teachers participated in STRIVE statistically significantly outperformed students whose teachers did not participate in STRIVE on measures of content knowledge (school level $g = 1.30$ to 1.42), vocabulary learning (school level $g = 1.19$ to 1.32) and content reading comprehension (school level $g = .42$ to $.65$). Statistically significant effects were not observed for the Gates MacGinitie Reading Comprehension (school level $g = .05$ to $.10$), however, the effect size for the Gates MacGinitie Vocabulary test was statistically significant (school level $g = .35$ to $.79$). These findings support the efficacy of a distributed PD model on student reading outcomes and supports the efficacy of using more sustainable methods of PD that feature school-supported follow-up PD. There were no significant differences between the two types of professional development delivery that implemented PACT (RPD or SPD).

We perceived the current study as an opportunity to further understand the effects of this STRIVE PD aimed at vocabulary, reading comprehension, and content knowledge outcomes for ELs. For this study, we describe the treatment as STRIVE whether the professional development was provided by researchers alone or researchers with school personnel. We elected to address questions related to ELs in a separate article rather than the main effects article for two reasons: (a) the main effects article represented a 3-year study of multiple cohorts of teachers and schools and an initial efficacy trial thus requiring extensive explanation of the study context and professional development including the instructional practices taught yielding an already long and complicated study description, and (b) the current literature on ELs related to this topic is underdeveloped and an article describing the effects both relative to the target students (ELs) and the context of their instruction (percent of ELs in the school) could be a focused study yielding potentially sharper analysis and description.

Conceptual Model

Our conceptual model for this study was guided by a theory of change in which teachers' knowledge and skills impact students content knowledge as well as construct-related vocabulary development. We also contextualize these findings for ELs within a model in which the outcomes for students who are ELs are differentially influenced by the percent of ELs in their school.

The rationale for this model is derived from several data sources. One is the pattern of findings from the fourth-grade reading test from the National Assessment of Educational Progress (NAEP) in which ELs frequently score lower on achievement tests in part because of their challenges in developing background knowledge and vocabulary in English ([National Center for Educational Statistics, United States. Office of Educational Research, Improvement. Center for Education Statistics, & Institute of Education Sciences \(US\), 2009](#)). Unlike most first language English students, ELs also have the dual task of concurrently learning English and content. In upper elementary grades the demands for language, literacy, and background knowledge are extensive. Essential to their success is access to high-quality teachers who are able to support their learning needs. Unfortunately, many ELs may have restricted opportunities because of their lack of access to

high-quality teachers, proficient student learners, and effective resources. For example, Callahan (2005) reported that ELs were frequently “tracked” into classes that significantly negatively affected their learning achievement. Her analysis revealed that ELs were primarily clustered in classes that were not college preparatory. To the extent that the opportunity to learn content is related to the curriculum demands of the class, and that teachers are more likely to provide challenging content and discourse opportunities to students who are proficient in English, ELs in classes with significant numbers of non-ELs may be more likely to access high-level academic vocabulary and content learning. The reverse is also likely in that teachers of classes with high concentrations of ELs may provide fewer opportunities for rich language discourse and content learning. In a recent study (Vaughn et al., 2017), these hypotheses were confirmed with the benefit of a treatment provided by middle school social studies teachers resulting in differential benefit depending on the class-level prevalence of English academic language, which was defined as the percent of ELs in the classroom. Increases in the class-level percentage of ELs disadvantaged both ELs and non-ELs with the explanation provided that sophisticated content-related English academic language is less available to all students under such conditions. The authors interpret the findings as suggesting that discourse-based literacy treatments have a stronger impact on knowledge acquisition for all students when less than 12% of the students in the class are ELs. In this study, we intend to follow-up on the authors’ (Vaughn et al., 2017) suggestion that the findings required replication.

The analyses were conducted to address three interrelated questions about the moderating role of EL status: (a) Does treatment (defined as STRIVE professional development) affect ELs and non-ELs differently? (b) Does treatment affect students differently depending on the school-wide percentage of ELs? (c) Does treatment affect ELs differently from non-ELs depending on the school-wide percentage of ELs?

Method

Research Design

This study used data from a multiple-cohort, randomized control trial (Capin et al., 2021) conducted to examine the efficacy of STRIVE in elementary schools. Within each of three cohorts, schools were randomly assigned to one of three conditions: RPD ($n = 26$), SPD ($n = 28$), or BAU ($n = 27$). Two schools that were randomized to participate in Cohort 3 did not complete the study. One of these schools was assigned to the BAU condition and declined to participate after a new principal was appointed. No data were collected at this school. A second school, assigned to the RPD condition, sent teachers to the initial PD session and allowed student data to be collected at pretest and after the first unit of instruction. At that time, the school withdrew from STRIVE in order to focus on improving the school’s state standardized test performance rating. We determined not to include data from this school in the analyses because it likely was not missing at random. Both reasons for withdrawal from the study were exogenous, or unrelated to the intervention. This produced an overall

student attrition rate of 2.5%, well within the limits of attrition deemed tolerable by the What Works Clearinghouse (2020).

As a means of providing an equivalent content-base for all conditions, all schools used the state adopted, fourth-grade social studies curriculum. For the purpose of these analyses, treatment is defined as the provision of professional development (STRIVE) to participating teachers, and we combined the RPD and SPD groups to create the treatment condition.

Setting and Participants

For this study, we examined the differential effect of STRIVE among ELs and non-ELs. We also investigated the moderating effect of school percentage of ELs on vocabulary, content acquisition, and reading comprehension. Therefore, additional information about ELs within the larger sample are provided below. The STRIVE study was implemented in 79 urban and near-urban schools across six school districts in the southwestern United States. The proportion of participating students in each school identified as ELs ranged from 0% to 77.4%.

Teachers

All 235 fourth-grade teachers (210 female and 25 male) from recruited schools consented to participate in the study; teachers received compensation for attending PD meetings and completing project tasks outside of their regularly scheduled workday. Teachers averaged 10 years teaching experience and held teaching certifications (100%). There were no statistically significant differences between teachers in the average years of teaching experience, $F(1, 34) = .583, p = .446$, or the percentage of teachers who held master’s degrees, $\chi^2(2) = 1.607, p = .448$.

Students

Parents of 4,757 students provided consent for their children to participate in data collection efforts (77% return rate). Of these students, 3,594 were identified as non-ELs and 921 were identified as ELs. Students’ EL designation was determined by districts based on students’ performance on their state English language proficiency test. Student demographics are detailed in Table 1. Among non-ELs, most students were of Hispanic or Caucasian ethnicity, and a majority of the students qualified to receive free or reduced-price lunch. Among the EL sample, most students were of Hispanic ethnicity, and the majority of these students qualified for free or reduced-price lunch.

Procedure

STRIVE PD

The PD was distributed over time and was delivered across the course of three units of study. Teachers participated in a PD session prior to the school year. They also participated in teacher study team meetings prior to Unit 2 and Unit 3.

During the initial 8-hr PD session led by researchers, teachers received an overview of all STRIVE instructional practices. They also watched video and in-person models of Unit 1 practices followed by time to practice with their peers and plan for implementation in classrooms. Teachers also learned about the importance of treatment adherence to support student outcomes. After

Table 1
Student Demographics

Variables	STRIVE				BAU			
	Non-ELs		ELs		Non-ELs		ELs	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Gender								
Female	1,323	48.8	319	48.6	453	51.2	129	48.9
Male	1,357	50.1	330	50.2	423	47.9	132	50
Ethnicity								
Asian	21	0.8	9	1.4	1	0.1	0	0
African American	136	5	4	0.6	55	6.2	0	0
Hispanic	1,760	64.9	636	96.8	665	75.2	260	99.6
Native American or Pacific Islander	4	0.1	1	0.2	1	0.1	0	0
Two or more	55	2	1	0.2	12	1.4	0	0
Caucasian	734	27.1	6	0.9	150	17	1	0.4
Identified as economically disadvantaged	1,629	60.1	522	79.5	636	71.9	219	83
Participates in special education	300	11.1	66	10	107	12.1	17	6.5

Note. ELs = English learners; STRIVE = Strategies for Reading Information and Vocabulary Effectively.

attending the initial training, teachers implemented Unit 1 lessons in their classrooms. Prior to Unit 2 and again prior to Unit 3, teachers participated in a teacher study team meeting lasting approximately 2 hr after school. To promote collaboration, meetings included an average of four teacher participants. Teacher study-team leaders followed the same meeting agenda, consisting of three components: (a) reflect on instruction from the previous unit, (b) introduce new practices for the coming unit, and (c) set one major goal for the coming unit.

Vocabulary and Reading Comprehension Instructional Practices

All schools in all conditions utilized the state-adopted social studies curriculum. This included using the same fourth-grade progressive state standards and the same state-developed timeline for delivering social studies. Therefore, the only difference between treatment and comparison groups was inclusion in STRIVE where teachers learned to use a set of vocabulary and reading comprehension instructional practices within social studies. Teachers implemented the instructional practices across three, 6-week units of study. Teachers delivered two, 45-min lessons per week (36 lessons total). The introduction of six vocabulary and reading comprehension instructional practices were also distributed over time so that teachers and students incrementally built knowledge and use over the course of time. During Unit 1, teachers introduced the following components: background knowledge, explicit vocabulary instruction, text-based discussion, and gist statements. During Unit 2, teachers added summary writing, and during Unit 3 teachers added the context clue strategy.

During each lesson, teachers engaged in instructional practices before, during and after reading. Two instructional practices took place before text reading. First, teachers used illustrations within the text selection to prompt students to make connections between prior knowledge and new content and to build background knowledge prior to reading. Second, explicit vocabulary instruction included introducing the word using a student-friendly definition, leading a discussion guided by a visual representation of the word, and providing examples of the word in the appropriate context. In Unit 3, teachers taught students a context clue strategy to derive

the meaning of words rather than providing them with a student-friendly definition.

During text reading, teachers led a text-based discussion framed by various question types to encourage literal and inferential thinking. They also taught students how to use the Get the Gist strategy several times during passage reading, a method for writing main idea statements that support content comprehension (Klingner et al., 2012).

After text reading, teachers returned to explicit vocabulary instruction and asked students to evaluate a list of four words to identify the ones related to the target word. Students also wrote a sentence using the word to demonstrate understanding and engaged in a turn-and-talk activity to apply their understanding of words in a way that connected to their own lives (e.g., If you could go on an expedition, where would you go and why?). A word building activity required students to add prefixes or suffixes to the target word to create new words. In Unit 2, teachers showed students how to use gist statements from sections of the text to write a summary of the entire passage.

Teachers also used several features of effective instruction. When teachers introduced new practices, they engaged in explicit instruction by: (a) telling students about the practice, (b) modeling each step of the practice using a think aloud, (c) engaging in multiple practice opportunities with students, and (d) allowing students to practice tasks independently. In addition, teachers provided two types of feedback to facilitate skill acquisition (Hattie & Timperley, 2007). Process-specific feedback focused on the ways in which students engaged in the process of strategies. Product-specific feedback focused on students' final product.

Fidelity

STRIVE PD Fidelity

All teachers attended the initial PD session and teacher study team meetings. After each, teachers completed a PD fidelity form. In the first section, teachers rated the extent to which they felt prepared to teach each instructional component. In the second section, teachers rated PD quality (e.g., the lesson components were well-described; the materials provided were useful; the opportunities to

Table 2
Professional Development Fidelity

Preparedness to teach each component	Initial PD	Teacher Study Team 1	Teacher Study Team 2
	STRIVE <i>M(SD)</i>	STRIVE <i>M(SD)</i>	STRIVE <i>M(SD)</i>
Before reading			
Background knowledge	3.79 (.43)	n/a	n/a
Explicit vocabulary instruction	3.86 (.36)	n/a	n/a
During reading			
Ask and answer questions	3.78 (.45)	n/a	n/a
“Get the Gist” main idea statements	3.71 (.53)	n/a	n/a
“Get the Gist” routine in collaborative learning pairs	n/a	3.60 (.58)	n/a
After reading			
“Gist to Summary”	n/a	n/a	n/a
Comprehension purpose question	3.78 (.45)	n/a	n/a
Vocabulary maps in collaborative learning pairs	3.79 (.45)	3.75 (.48)	n/a
Overall quality	3.93 (.26)	3.79 (.42)	3.80 (.47)

Note. PD = professional development; STRIVE = Strategies for Reading Information and Vocabulary Effectively. Teachers rated how prepared they felt to teach each practice after the professional development sessions, with 1 = *not prepared* and 4 = *prepared*. Overall quality was rated on a 4-point scale. n/a = not applicable because the instructional practice was not introduced.

practice were helpful). Items in both sections were rated on a scale of 1 to 4 (1 = *strongly disagree* to 4 = *strongly agree*). Table 2 shows that teachers agreed that they were prepared to teach each instructional component across all PD sessions. Teachers rated the PD quality as high.

STRIVE Instructional Practices Fidelity

The instructional practices coding document was adapted from a previous study (e.g., Simmons et al., 2010) and contained two sections: instructional adherence and instructional quality. In the instructional adherence section, items aligned to the STRIVE instructional practices and were coded on a 4-point, Likert-type rating scale ranging from 1 (*low alignment with intended method*) to 4 (*high alignment with intended method*). If a component was not expected during a lesson or there was not enough time for the teacher to implement a specific component (i.e., fire drill or other school-related interruption), coders scored the item as “not applicable.” Instructional quality was coded on a 4-point scale and focused on teacher’s general instructional performance using seven items that addressed lesson pacing, the use of feedback, frequency of practice opportunities, teacher preparedness, clarity of questions, explicit instruction, and enthusiasm.

Fidelity Coding Procedure

The research team asked teachers randomized to the RPD and SPD conditions to audio-record all STRIVE lessons. Teachers assigned to the BAU condition recorded 1 week (numerals should

be used with units) of BAU social studies instruction per six-week period (3 weeks total recorded). A total of 228 lessons (three per school; one from each STRIVE unit) were randomly selected across conditions for coding. Seven members of the research team participated in a 4-hr training conducted by the principal investigator who has extensive experience in observation coding (e.g., Swanson et al., 2012; Swanson et al., 2016). Training included learning the fidelity codebook and practice assigning codes using audio recordings. Following the training and using the gold standard method (Gwet, 2001), the principal investigator assigned one audiotape for the coders to independently score for interrater agreement. The principal investigator, who has more than 20 years teaching experience and was on the design team for both the distributed PD model and STRIVE, served as the gold standard. Once independent coding was completed, the principal investigator reviewed the scores and met with the coders individually to discuss discrepant codes. This process was repeated until the research team members reached agreement of 90% or greater with the gold standard. The mean interrater agreement with the gold standard was 92%. Coding took place over the course of one month. Coders were encouraged to ask questions and discuss with the principal investigator when they were unsure about a specific code. Additionally, to maintain agreement, a second member of the research team independently coded one third of audio recordings. Agreement between coders was above 90% for all double-coded audio recordings.

Fidelity Results

We used fidelity data to identify (a) the extent to which instruction in the treatment conditions aligned with STRIVE instructional practices as intended, and (b) the extent to which STRIVE instructional practices were observed in the BAU condition. Fidelity ratings for the RPD and the SPD conditions were midhigh to high for most practices, indicating that teachers implemented the STRIVE instructional practices as intended (see Table 3). Scores for the

Table 3
Instructional Fidelity

STRIVE components	STRIVE		BAU	
	<i>M(SD)</i>	Times observed	<i>M(SD)</i>	Times observed
Background knowledge	2.57 (1.09)	115	2.28 (.90)	57
Explicit vocabulary instruction				
Before reading	3.64 (.717)	121	1.88 (.97)	40
After reading	3.35 (.924)	81	1.33 (.58)	3
Questions to prompt text based discussion				
Pose comprehension purpose question	2.79 (1.43)	121	2.82 (1.17)	11
Questions during text reading	3.27 (.98)	127	2.77 (1.01)	53
Discuss comprehension purpose question	2.86 (1.31)	76	2.33 (1.53)	3
Gist statements	3.10 (.95)	100	2.14 (.90)	7
Summaries	2.96 (1.12)	50	1.25 (.50)	4
Lesson closure	2.88 (1.17)	60	1.92 (1.02)	24

Note. BAU = business as usual; STRIVE = Strategies for Reading Information and Vocabulary Effectively. In Unit 3, explicit vocabulary instruction included the context clue strategy. Implementation of components was rated on a 4-point scale (1 = *lowest* and 4 = *highest*).

researcher-supported PD group ranged from 2.74 (building background knowledge) to 3.81 (explicit vocabulary instruction before reading). Scores for the SPD group ranged from 2.79 (building background knowledge) to 3.64 (explicit vocabulary instruction before reading).

We were also interested in the extent to which STRIVE instructional practices were observed in the BAU condition. Building background knowledge, explicit vocabulary instruction before reading (this included a student friendly definition, an illustration, and the word used in a sentence), questioning during text reading, and lesson closure were observed frequently in the BAU condition; these practices were delivered with medium ($M = 2.15$) to medium-high ($M = 2.90$) alignment to STRIVE practices. The following practices were rarely observed: vocabulary instruction after reading, posing and discussing a comprehension purpose question, Get the Gist, and summary writing.

The fidelity data were analyzed to determine if significant differences existed in frequency of use of each STRIVE-related practice between teachers in each PD group and in the BAU group. We analyzed the fidelity data using Fisher's Exact Test, comparing the three groups pairwise. Because of the number of group comparisons involved, we implemented the Benjamini-Hochberg (Benjamini & Hochberg, 1995) procedure for controlling for the false discovery rate. Seven of the nine STRIVE instructional practices were observed significantly more often in the STRIVE treatment groups compared with the BAU group. No significant differences in fidelity were found between teachers in the two PD conditions.

Measures

Gates MacGinitie Reading Comprehension Subtest

Administered at pre- and posttest, the Gates MacGinitie reading comprehension subtest (4th edition; MacGinitie et al., 2000) is a group administered, 35-min timed assessment consisting of informational and narrative passages. Students read each passage silently and answer three to six multiple-choice questions related to the passage. Schools in one district placed limits on testing time. As a result, a group of students were not administered this measure. Analysis of outcomes on the reading comprehension subtest was conducted using data from 37 schools in districts that did not place limits on testing time. Internal consistency reliability for the reading comprehension subtest ranges from .91 to .93, and alternative form reliability is .80 to .87.

Gates MacGinitie Vocabulary Subtest

Administered at pre- and posttest, the Gates MacGinitie vocabulary subtest (4th edition; MacGinitie et al., 2000) is a group administered, 45-item, 20-min timed assessment of vocabulary knowledge. Each item presents a word in brief context followed by five word-meaning choices. Items consist of age-appropriate and frequently encountered vocabulary words. The Kuder Richardson 20 reliability for this measure is .90–.92.

Content Measures

Three measures were administered to assess students' content knowledge, vocabulary knowledge, and comprehension of social

studies text. Content knowledge and vocabulary knowledge were measured at the end of each 6-week unit. Content reading comprehension was assessed at posttest. Because these measures were used to assess curriculum mastery and many students in the sample were novice English speakers, Spanish versions of the three curriculum-based measures were made available to teachers in all three conditions. None of the teachers in the BAU condition opted to use the Spanish version. Because there were no students in the STRIVE PD conditions who responded to the Spanish version of the content measures those assessment provided in Spanish were not included before analyses of outcomes were conducted ($n = 96$ EL and 57 non EL).

Content Knowledge Tests

Students in all three conditions responded to three content knowledge assessments. Each unit content test consisted of 20–21 items; each item included a brief sentence stem followed by four answer choices. The tests took approximately 30 min to complete. We utilized data from students included in Cohort 1 to evaluate the item properties of the content tests. Items on the content knowledge and reading comprehension measures that were determined to be too difficult for students (based on a criterion of 40% or fewer students responding correctly) were revised. Cohort 2 and Cohort 3 students in all three conditions responded to this revised assessment. To allow for analysis of scores on the content measures with the three cohorts combined, total scores for each student were computed based on the common items administered to students in both cohorts. Item-total correlations were lower than desired, ranging from .33 to .41. Given these low item-total correlations, internal consistency reliabilities also were lower than desired, ranging from .77 to .84 across the three content tests.

Vocabulary Knowledge Tests

Using Espin et al.'s (2001) procedures, Swanson et al. (2018) developed three vocabulary-matching tests of knowledge of key terms from each of the three units. The tests consisted of 24 items for Unit 1, 16 items for Unit 2, and 19 items for Unit 3. For each assessment, students matched each word with a brief definition. Students completed the measure in approximately 20 min. We utilized data from students included in Cohort 1 to evaluate the item properties of the unit test of vocabulary knowledge. None of the items on the vocabulary measure were determined to be too difficult for students (based on a criterion of 40% or fewer students responding correctly). Espin et al. (2001) reported alternate-form reliability for vocabulary matching as .76 to .88. Within our sample, item-total correlations were better than those for the content knowledge tests, ranging from .54 to .58. Internal consistency reliabilities ranged from .89 to .93. However, the reliability estimates may be inflated because matching tests use a common set of response choices.

Content Reading Comprehension

A content reading comprehension assessment also was developed by Swanson et al. (2018) to test students' ability to understand expository texts containing social studies content. Students read five passages and responded to six multiple-choice items following each passage. Passages ranged from 197 to 233 words and

conformed to a fourth-grade reading level with a Lexile range of 700L to 900L. Internal consistency reliability for the assessment was .89.

Data Analysis Plan

We analyzed extant data collected in a school-randomized control trial to evaluate the effects of distributed PD on the content knowledge, vocabulary, and reading comprehension outcomes of upper elementary EL students (Capin et al., 2021). We evaluated the effects of STRIVE treatment in the group of ELs, by addressing three interrelated questions: (a) Does treatment affect ELs and non-ELs differently? (b) Does treatment affect students differently depending on the school-wide percentage of ELs? (c) Does treatment affect EL and non-EL students differently depending on the school-wide percentage of ELs?

All three questions involve statistical moderation. The first question represents the two-way interaction of treatment and EL status. The second question addresses a two-way interaction, in this case between treatment and school-wide percentage of ELs. Question 3 asks whether treatments' effects on ELs and non-ELs differ depending upon the school-wide percentage of ELs, an important school-level moderator in past studies involving EL students (Vaughn et al., 2017). We fit multilevel regression models (Hoffman, 2015) to account for the nesting of students (Level 1) within teachers (Level 2) and schools (Level 3). We indexed students' EL status (1 = EL, 0 = non-EL) on Level 1 of the model. Average treatment effects were modeled at the school-level (1 = STRIVE, 0 = BAU). Statistical moderation involving EL status (RQ 1) was modeled as a cross-level interaction involving Levels 1 and 3. School-wide percentage of ELs was measured at Level 3, and the two-way interaction involving treatment and school-wide percentage of EL students was modeled on Level 3 of the model (RQ 2). The moderating effect of school-wide percentage of ELs on the interaction of treatment and EL-status was modeled as a three-way, cross-level interaction (RQ 3), with EL status at Level 1 and the other two factors at Level 3 of the model.

The Gates MacGinitie reading comprehension and vocabulary subtests (MacGinitie et al., 2000) were administered to students within 2 weeks prior to and 2 weeks immediately following treatment implementation. The content measures, including the unit tests of content knowledge and vocabulary knowledge and the content reading comprehension measure were administered at posttest only. We modeled "Reading Outcomes" for content measures according to the reduced-form equation in Equation 1:

$$\begin{aligned} \text{Reading Outcome}_{ijk} = & \gamma_{000} + \gamma_{001}(\text{STRIVE}_k) \\ & + \gamma_{100}(\text{EL}_{ijk}) + \gamma_{002}(\text{Sch.prc.EL}_k) + \gamma_{101} \\ & (\text{EL}_{ijk} \times \text{STRIVE}_k) + \gamma_{102}(\text{EL}_{ijk} \times \text{Sch.prc.EL}_k) \quad (1) \\ & + \gamma_{003}(\text{Sch.prc.EL}_k \times \text{STRIVE}_k) + \gamma_{102} \\ & (\text{EL}_{ijk} \times \text{Sch.prc.EL}_k \times \text{STRIVE}_k) + e_{ijk} + r_{ojk} + u_{00k} \end{aligned}$$

where Reading Outcome_{ijk} represents the posttest score for Student *i* in Teacher *j* in School *k* and STRIVE_k indicates treatment assignment. EL_{ijk} is EL status with the non-EL group coded as 0 and EL coded as 1. Sch.prc.EL_k is a school-level moderator and represents the percentage of students who are EL in a given school. Other effects in the fixed effects portion of the model represent the two-way interactions and the three-way interaction. Random effects,

represented by the latter three terms, denote student-, teacher-, and school-level variances, respectively.

The reduced form equation for the models predicting outcomes on the Gates-MacGinitie reading comprehension and Gates-MacGinitie vocabulary subtests was:

$$\begin{aligned} \text{Reading Outcome}_{ijk} = & \gamma_{000} + \gamma_{100}(\text{Pretest}_{ijk}) \\ & + \gamma_{010}(\text{Pretest}_{jk}) + \gamma_{001}(\text{Pretest}_{tk}) + \gamma_{002} \\ & (\text{STRIVE}_k) + \gamma_{200}(\text{EL}_{ijk}) + \gamma_{003}(\text{Sch.prc.EL}_k) \\ & + \gamma_{202}(\text{EL}_{ijk} \times \text{STRIVE}_k) + \gamma_{203}(\text{EL}_{ijk} \times \\ & \text{Sch.prc.EL}_k) + \gamma_{004}(\text{Sch.prc.EL}_k \times \text{STRIVE}_k) \\ & + \gamma_{204}(\text{EL}_{ijk} \times \text{Sch.prc.EL}_k \times \text{STRIVE}_k) + e_{ijk} \\ & + r_{ojk} + u_{00k} \end{aligned} \quad (2)$$

where Reading Outcome_{ijk} represents the posttest score for Student *i* in Teacher *j* in School *k*, as in Equation 1. However, unlike the content measures, for which only posttest data were collected, the Gates-MacGinitie subtests were administered at pretest and posttest. Accordingly, Equation 2 includes pretest scores at the model's three levels. We centered Level-1 student scores—(Pretest_{ijk})—on teacher means and teacher-level aggregates—(Pretest_{jk})—on school means. The other terms in Equation 2, including STRIVE_k, the two-way interactions and the three-way interaction, represent the same parameters as their analogs in Equation 1. Random effects are represented, as well.

School-wide EL is expressed as a percentage, bounded by 0 and 100. Note, however, that the percentages are not binomial random variables resulting from a finite number of trials (e.g., Bernoulli trials). Instead, they represent true percentages and to the extent that they distribute normally, we elected to treat them as integer-like data to aid with interpretability. The values ranged from 0 to 77, with a mean of 19.5, standard deviation of 14.3, and median of 20.8. Because the skewness (.68) and kurtosis (1.35) indices were within acceptable bounds, we treated school-wide percentage of EL as interval-level data and centered it on the moderator's mean which was 20%.

Interactions involving continuous moderators were further analyzed using the Johnson-Neyman (J-N) technique (Preacher et al., 2006) to identify values along the moderator's range at which treatment's effect transitions from statistically significant to non-significant. The J-N technique establishes the subset of moderator values for which treatment's effect is statistically significant. We used the "lme4" package in R (Bates et al., 2015) to fit models. Two-way interactions were further decomposed and contrasts were computed using the "emmeans" package (Lenth et al., 2020) in R. We calculated student-level effect sizes using Hedges' (2011) equations for cluster-randomized three-level models. This represents the effect for the average student in the average classroom in the average school. It is analogous to the effect sizes typically reported in the educational sciences literature and aligns with recommendations of the What Works Clearinghouse.

Results

Table 4 summarizes observed pretest and posttest means and standard deviations for Gates-MacGinitie reading comprehension and reading vocabulary subtests (MacGinitie et al., 2000) and the posttest means and standard deviations for the content measures. Values are provided for the two conditions and for

Table 4
Student Level Descriptive Statistics for Reading Outcomes

Reading Outcomes		Pretest		Posttest		
Gates comprehension						
Non-ELs						
BaU	332	472.03	37.95	334	488.25	36.93
STRIVE	1,527	473.48	42.73	1,477	488.97	42.44
ELs						
BaU	77	460.04	32.82	78	465.26	33.66
STRIVE	233	447.18	43.37	232	462.92	44.83
Gates vocabulary						
Non-ELs						
BaU	839	459.84	36.15	780	476.64	39.60
STRIVE	2,560	462.56	40.71	2,434	482.22	43.33
ELs						
BaU	247	441.06	33.37	248	450.75	35.77
STRIVE	622	436.27	40.11	627	451.97	41.41
Benchmark multiple choice						
Non-ELs						
BaU				884	20.02	8.49
STRIVE				2,614	29.50	11.33
ELs						
BaU				264	19.05	7.51
STRIVE				600	27.38	10.03
Benchmark reading comprehension						
Non-ELs						
BaU				884	6.27	4.38
STRIVE				2,614	7.88	4.59
ELs						
BaU				264	5.75	3.80
STRIVE				600	7.04	4.18
Benchmark vocabulary						
Non-ELs						
BaU				884	25.63	14.33
STRIVE				2,614	39.39	16.45
ELs						
BaU				264	22.63	12.03
STRIVE				600	35.18	15.36

Note. ELs = English learners; BAU = business as usual; STRIVE = Strategies for Reading Information and Vocabulary Effectively.

ELs and non-ELs. Tables 5 and 6 summarize model parameters for the content measures and for the Gates-MacGinitie reading subtests, respectively. We summarize the findings according to the three research questions, fitting fully conditional models, with estimates for all possible two-way interactions and the relevant three-way interaction for each outcome, and with lower-order effects interpreted in the context of higher-order interactions.

Research Question 1: Does Treatment Affect ELs and Non-ELs Differently?

Question 1 concerns treatment's moderation by EL status. Because the interpretation of lower-order effects is conditional on the statistical significance of higher-order interaction terms, we present findings for three groups of outcomes—those with simple main effects (i.e., no significant higher order effects), those with simple two-way moderated effects (i.e., no statistically significant three-way interactions), and those with two-way effects “nested” in a statistically significant three-way interaction effect. Content knowledge and the Gates-MacGinitie reading comprehension subtest comprise the first group of outcomes. The Content Reading

Comprehension test represents the second group, and the third group includes the content vocabulary measure and the Gates-MacGinitie vocabulary subtest.

As shown in Tables 5 and 6, on the content knowledge measure, students in STRIVE schools ($\gamma_{001} = 8.32$, p -value $< .001$) outperformed students in the BAU, on average, but treatment's effect did not differ for EL and non-EL students. Both groups performed equally well at posttest; the effect sizes for the STRIVE in the non-EL and the EL groups were .56, 95% CI [.38, .73] and .55, 95% CI [.38, .72], respectively.

On the Gates-MacGinitie reading comprehension subtest, students' performance in STRIVE schools ($\gamma_{001} = -.40$, p -value = .92) did not differ, on average, from students in BAU schools and treatment's effect did not differ for EL and non-EL students (see Table 7). The effect sizes for the STRIVE in the non-EL and the EL groups were $-.04$, 95% CI $[-.29, .20]$ and $.12$, 95% CI $[-.12, .37]$, respectively.

The main effect for STRIVE on the content reading comprehension test differed from 0 ($\gamma_{001} = 1.33$, $SE = .50$, p -value $< .01$), meaning that non-EL students in STRIVE schools with 20% ELs (the mean for school-wide EL) scored about 1.3 point higher at posttest than non-EL students assigned to similar BAU schools

Table 5
Model Estimates for Content Measures

Fixed effects	Unit test of content knowledge			Content reading comprehension			Unit test of content vocabulary		
	Estimate	SE	p-value	Estimate	SE	p-value	Estimate	SE	p-value
Intercept	19.63	1.09	.00	6.01	0.41	.00	24.85	1.48	.00
Simple effects									
STRIVE	8.32	1.31	.00	1.33	0.50	.01	12.28	1.79	.00
Sch. Prc. ELs	-0.05	0.09	.55	-0.01	0.03	.73	-0.15	0.12	.23
EL	-1.14	0.99	.25	-0.94	0.42	.03	-3.48	1.50	.02
Two-way interactions									
EL × STRIVE	1.43	1.12	.20	0.99	0.48	.04	3.27	1.70	.05
Sch. Prc. ELs × STRIVE	-0.14	0.11	.21	-0.05	0.04	.20	-0.16	0.14	.27
Sch.Prc. ELs × EL	0.06	0.09	.46	0.05	0.04	.14	0.24	0.13	.07
Three-way interactions									
Sch. Prc. ELs × STRIVE × EL	-0.09	0.10	.36	-0.05	0.04	.20	-0.30	0.15	.04
Random effects	Variance	ICC		Variance	ICC		Variance	ICC	
Student-level	79.72	0.73		15.75	0.80		184.12	0.77	
Teacher-level	7.33	0.07		0.80	0.04		16.85	0.07	
School-level	21.42	0.20		3.06	0.16		37.27	0.16	

Note. Sch. = school; Prc. = percent; Els = English learners; STRIVE = Strategies for Reading Information and Vocabulary Effectively; ICC = Intraclass Correlation Coefficient.

(similar in terms of school wide percent EL).The significant main effect for EL status ($\gamma_{100} = -.94, SE = .42, p\text{-value} < .05$) suggests that EL students performed less well than non-ELs in average BAU schools. The positive and statistically significant cross-level interaction between EL status and STRIVE indicates that treatment’s effect depended on students’ EL status (see Figure 1). For non-EL students, the effect was .20, 95% CI [.06, .34]; for ELs, STRIVE’s effect was .27, 95% CI [.12, .40].

The main effect on the test of content vocabulary differed significantly from 0 ($\gamma_{001} = 12.28, SE = 1.79, p\text{-value} < .01$) for STRIVE. The main effect of EL status was significant, as well, indicating that EL students in BAU schools with 20% ELs scored 3.48 points lower than non-ELs in the same BAU schools. Among the two-way interactions, STRIVE * EL was positive and differed statistically from 0 ($\gamma_{101} = 3.27, SE = 1.70, p = .05$), indicating that the difference between ELs and non-ELs was significantly smaller (less negative, in this case) in STRIVE schools with 20% ELs than in BAU

schools with 20% ELs (see Figure 2). The effect size in the non-EL group was .51, 95% CI [.35, .66]. In the EL group, the STRIVE effect was .50, 95% CI [.35, .66]. In sum, both ELs and non-ELs benefited from STRIVE; however, non-EL students outperformed ELs on average. It is important to note that the differences in effect sizes between ELs and non-ELs is not large; nonetheless, the two effects differ significantly from one another. Estimates for these contrasts were based on t-values (the ratio of a model-based coefficient and the associated standard error) rather than model-based coefficients. The coefficient for the group of EL students ($\gamma_{001} = 16.3, SE = 2.36, p\text{-value} < .001$) was greater than that for the group of non-ELs ($\gamma_{001} = 12.50, SE = 1.80, p\text{-value} < .001$); however, its standard error was greater as well, resulting in a smaller effect for the EL group.

Finally, on the Gates-MacGinitie vocabulary subtest, the main effect for STRIVE treatment did not differ from 0 ($\gamma_{003} = 2.45, SE = 1.83, p\text{-value} = .19$), although there was a positive and statistically significant cross-level interaction between EL status and treatment ($\gamma_{203} = 5.79, p\text{-value} = .05$), suggesting that ELs in STRIVE schools outperformed non-ELs on the Gates-MacGinitie vocabulary measure (see Figure 3). The effect size for STRIVE was .11, 95% CI [.03, .20] among ELs, on average, whereas treatment’s effect in the non-EL group was .06, 95% CI [-.03, .14]. STRIVE did not have an overall effect; however, it was significantly more effective for ELs than non-ELs in sites with average school-wide percent ELs.

Research Question 2: Does Treatment Affect Students Differently Depending on the School-Wide Percentage of ELs?

Treatment’s effect was nondifferent across values of school-wide percent EL for all outcome measures. Main effects for school-wide percent did not differ from 0 and the two-way interactions involving school-wide percent were not statistically significant.

Table 6
Effect Size Estimates for Reading Outcomes

Reading Outcomes	ES [95% CI]
Unit test of content knowledge	
Non-ELs	0.56 [0.38, 0.73]
ELs	0.55 [0.38, 0.72]
Content reading comprehension	
Non-ELs	0.20 [0.06, 0.34]
ELs	0.27 [0.12, 0.40]
Unit test of content vocabulary	
Non-ELs	0.51 [0.35, 0.66]
ELs	0.50 [0.35, 0.66]
Gates MacGinitie vocabulary	
Non-ELs	0.06 [-0.03, 0.14]
ELs	0.11 [0.03, 0.20]
Gates MacGinitie reading comprehension	
Non-ELs	-0.04 [-0.29, 0.20]
ELs	0.12 [-0.12, 0.37]

Note. ES = effect sizes; Els = English learners.

Table 7
Model Estimates for Gates-MacGinitie Reading Test

Fixed effects	Gates comprehension			Gates vocabulary		
	Estimate	SE	p-value	Estimate	SE	p-value
Intercept	484.74	3.45	.00	471.10	1.53	.00
Simple effects						
STRIVE	-0.40	4.10	.92	2.45	1.83	.19
Sch. Prc. ELs	-0.32	0.33	.34	0.03	0.13	.83
EL	-8.61	4.45	.05	-7.10	2.65	.01
Two-way interactions						
EL × STRIVE	7.33	4.99	.14	5.79	2.99	.05
Sch. Prc. ELs × STRIVE	0.17	0.35	.64	0.01	0.15	.92
Sch. Prc. ELs × EL	-0.03	0.44	.95	0.14	0.22	.53
Three-way interactions						
Sch. Prc. ELs × STRIVE × EL	-0.20	0.47	.68	-0.51	0.25	.04
Control variables						
Pretest L1	0.70	0.02	.00	0.77	0.01	.00
Pretest L2	0.73	0.07	.00	0.91	0.06	.00
Pretest L3	0.79	0.12	.00	1.08	0.05	.00
Random effects	Variance	ICC		Variance	ICC	
Level 1	713.28	0.93		560.40	0.92	
Level 2				31.39	0.05	
Level 3	55.10	0.07		20.21	0.03	

Note. Sch. = school; Prc. = percent; ELs = English learners; STRIVE = Strategies for Reading Information and Vocabulary Effectively; L1-L3 = Levels 1-3; ICC = Intraclass Correlation Coefficient.

Research Question 3: Does Treatment Affect EL and non-ELs Students Differently Depending on the School-Wide Percentage of ELs?

Here, we extend the findings for Research Questions 1 and 2 to include the statistically significant three-way interaction for content vocabulary and the Gates-MacGinitie vocabulary subtest. The statistically significant (p -value = .05) three-way interaction for the measure of content vocabulary means that non-ELs outperformed ELs when the percent of ELs in STRIVE and BAU schools was below 38 (based on the J-N technique). The performance of ELs and non-ELs did not differ in schools with more than 37% EL students. Figure 4 depicts the posttest performance for ELs in BAU schools, for ELs in STRIVE schools, and for non-EL students in BAU and treatment schools across the range of values for school-wide percent EL. The significant three-way interaction for the Gates-MacGinitie vocabulary subtest means that the cross-level interaction between STRIVE and EL was conditional on values of school-wide percent of ELs. Specifically, STRIVE was more effective for ELs than it was for non-ELs when the percentage of ELs was below 25%. At or above 25%, treatment's effect did not differ in the two groups. Figure 5 displays the interrelationships of condition, EL status, and school-wide percent EL.

Discussion

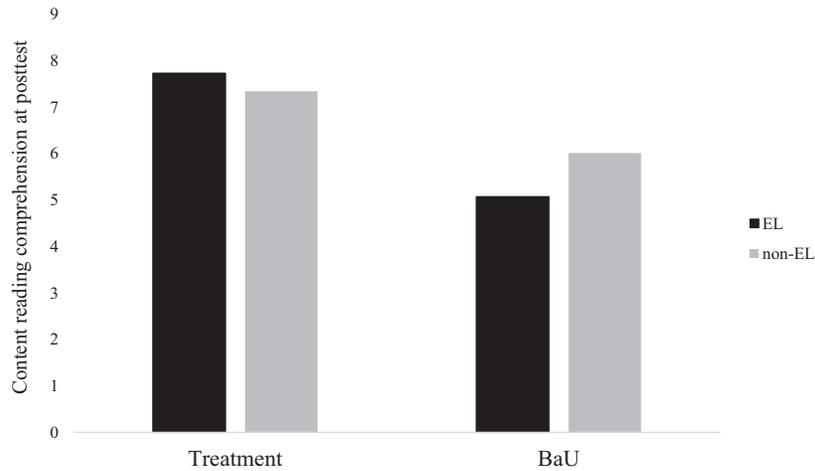
Building from a large-scale intervention study that provided evidence of efficacy for the STRIVE PD model on student content knowledge, content vocabulary, content reading comprehension, and general vocabulary outcomes (Capin et al., 2021), this article addressed critical questions about the relative effects for students who are ELs and how their performance varied across conditions from non-ELs. The set of moderation questions also addressed

how treatment effects differed for ELs based upon the percentage of ELs in the school, a policy and practice issue that has been evaluated in previous studies and found to have a conditioned effect on the performance of ELs based on the percentage of ELs in a class (Vaughn et al., 2017).

The first question informed whether there were differential treatment effects for ELs and non-ELs. This question directly addresses for whom the treatments were effective and whether findings differentially impact students who are ELs compared with those who are non-ELs. On content knowledge acquisition, both the ELs and non-ELs in the treatment conditions outperformed the ELs and non-ELs in the BAU condition. Treatment students who were ELs were not statistically significantly different from treatment non-ELs and these scores were significantly greater than students who received traditional social studies instruction (i.e., BAU). This favorable outcome for ELs in treatment conditions is impactful in light of the fact that content in the BAU was the same as the content in the treatment conditions. Thus, the significant gains on content knowledge for students who are ELs is associated with the reading comprehension and vocabulary practices their teachers learned during STRIVE PD and utilized in the classroom. The effect sizes for ELs and non-ELs in the treatment conditions were not trivial for ELs and non-ELs on content knowledge.

The highest priority for students who are ELs is improving their language development in English. In this study, as well as many others, we use vocabulary development as a proxy for word meaning knowledge and language development. Results with the Gates-MacGinitie vocabulary subtest (MacGinitie et al., 2000) demonstrated that for students in the BAU condition, ELs performed less well than non-ELs by an average of 7.08 points with both groups making positive gains. The need to provide teachers with effective practices for enhancing ELs' reading comprehension, content knowledge, and vocabulary knowledge outcomes across subjects/

Figure 1
EL × STRIVE Interaction on the Test of Content Reading Comprehension



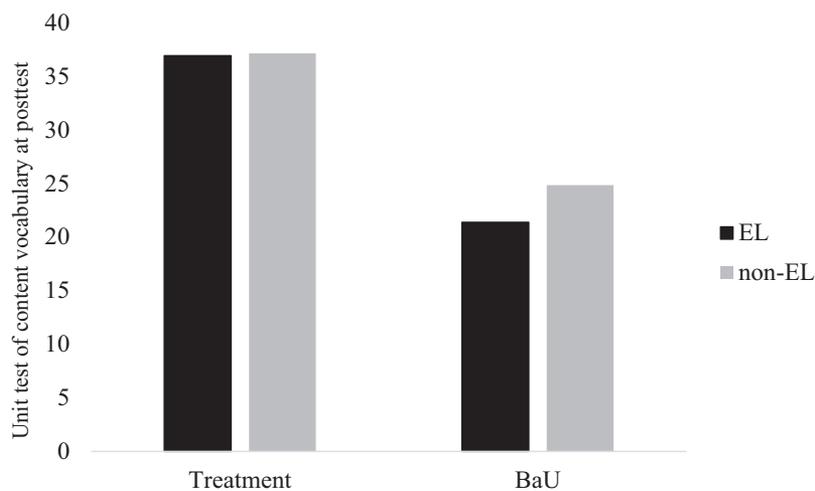
Note. This figure provides a visualization of EL as a moderator of treatment's effect on content reading comprehension at posttest. Els = English learners; BAU = business as usual; STRIVE = Strategies for Reading Information and Vocabulary Effectively.

content areas is of high importance and this finding reveals that what teachers perceived as a feasible set of instructional practices (Capin et al., 2021) positively impacted a range of students in their classes including those who are ELs.

This study also investigated whether treatment differentially affects students depending on the percentage of ELs in the school. Treatment's effect was not significantly different across the range of percent of ELs in the school for treatment conditions. Relatedly, we addressed whether the treatment affects EL and non-EL students differently depending upon the school-wide percentage of

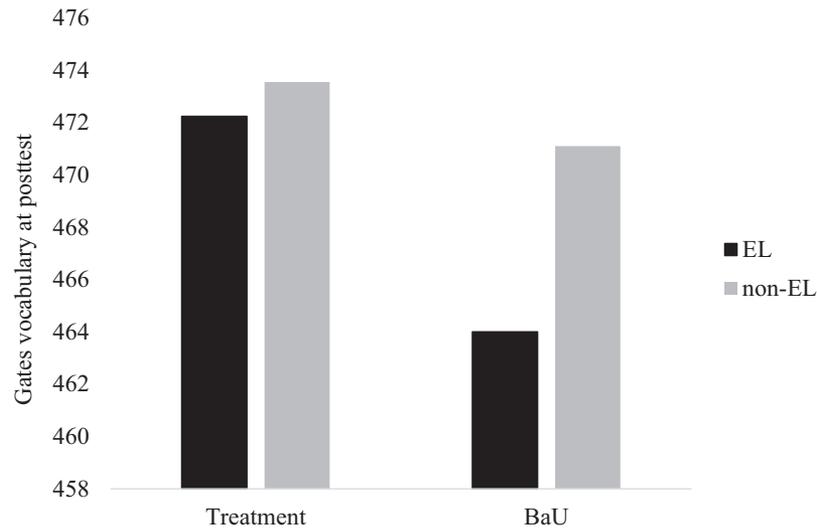
ELs. Addressing this question resulted in a significant three-way interactive effect for vocabulary—both content vocabulary and the Gates-MacGinitie vocabulary subtest. For the measure of content vocabulary, non-ELs outperformed ELs when the percent of ELs in STRIVE and BAU schools was below 38 (based on the J-N technique); however, non-ELs and ELs were not significantly different when more than 37% of the school population were ELs. Interpreting the three-way interaction for the Gates-MacGinitie vocabulary subtest differed by treatment group. Specifically, STRIVE was more effective for ELs than it was for non-ELs when

Figure 2
EL × STRIVE Interaction on the Unit Test of Content Vocabulary



Note. This figure provides a visualization of EL as a moderator of intervention effect between the STRIVE and BAU interventions on the unit test of content vocabulary at posttest. Els = English learners; BAU = business as usual; STRIVE = Strategies for Reading Information and Vocabulary Effectively.

Figure 3
EL × STRIVE Interaction on Gates-MacGinitie Vocabulary

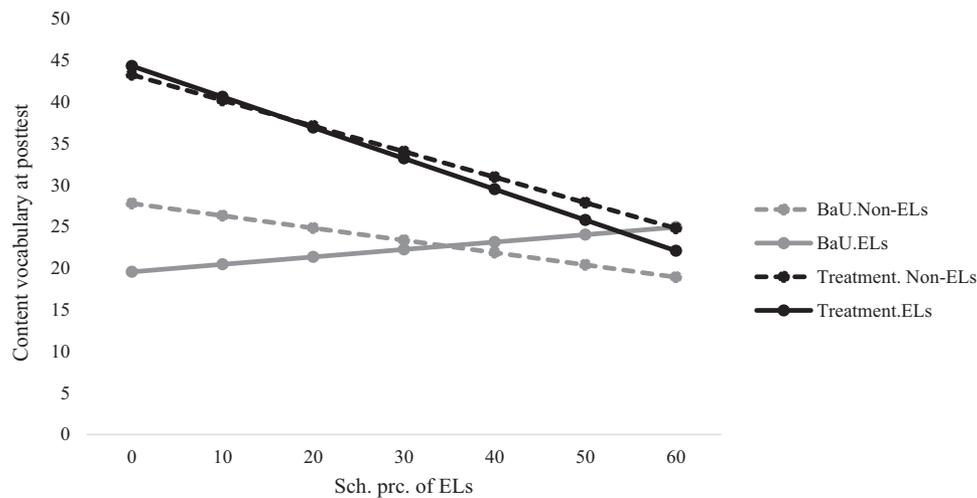


Note. This figure provides a visualization of EL as a moderator of treatment’s effect on Gates-MacGinitie Vocabulary extended scale score at posttest. STRIVE = Strategies for Reading Information and Vocabulary Effectively; Els = English learners; BAU = business as usual.

the percentage of ELs was below 25%. At or above 25%, treatment’s effect did not differ among ELs and non-ELs. We do not interpret these findings as suggesting that there are school level conditions, such as percentage of non-ELs or ELs in a school that are associated with differential performance on vocabulary. This may be because the instruction and/or exposure to vocabulary

learning, when the supports are provided by the research team, is altered in schools where the range of ELs is less than 25%. Of course, with three-way interactions such as these, there are several possible explanations including the vocabulary learning opportunities that occur informally between students in schools where significant numbers of students are ELs. We are not suggesting that

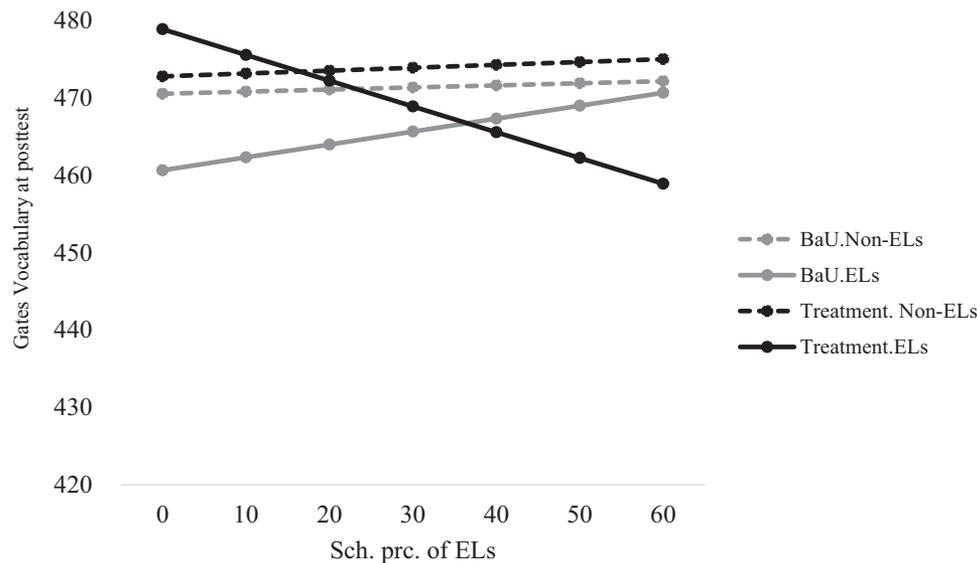
Figure 4
Three-Way Interaction on the Unit Test of Content Vocabulary



Note. This figure provides a visualization of treatment effects on the unit test of content vocabulary for ELs and non-ELs in STRIVE and BAU schools at different values of school-wide percentage of ELs. STRIVE = Strategies for Reading Information and Vocabulary Effectively; Els = English learners; BAU = business as usual.

This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Figure 5
Three-Way Interaction on Gates-MacGinitie Vocabulary



Note. This figure provides a visualization of differential treatment effects between ELs and non-ELs based on school percentage of ELs on Gates-MacGinitie vocabulary at posttest. ELs = English learners; BAU = business as usual; Sch. Prc. = school percent.

this data implies policy action but may serve as guidance for generating future research questions about the conditions under which ELs and non-ELs demonstrate language and literacy growth. We appreciate that these findings compare non-English to English speakers only and that findings related to home-language may vary considerably and in different ways.

In a previous study conducted with middle school students in which a set of instructional practices aimed at improving content learning and reading comprehension were implemented in social studies classes (PACT; Vaughn et al., 2017), both ELs and non-ELs in the treatment condition outperformed students in the BAU condition on content acquisition and reading comprehension. Because randomization was at the class level, the investigators were able to investigate the percentage of ELs in the class as a moderator of treatment effects. Interestingly, percentage of ELs in a class was a significant moderator for EL performance yielding declining performance as the percentage of ELs increased above approximately 15% only for knowledge acquisition but not for reading comprehension. We consider the knowledge acquisition measure as used in the PACT study as a similar construct to the vocabulary measure used in the study reported in this article because they both measure “knowledge” of the content and are associated with performance changes for ELs and non-ELs. The Vaughn et al. (2017) study provides evidence that the percentage of ELs within a class has an impact on the success of ELs within those classes. The current study was randomized at the school level and provides some support for Vaughn et al.’s (2017) findings but leaves some question about the role of school percentage of ELs in student outcomes that should be investigated empirically. It may be that the class level, rather than the school level, is a better setting for investigating the impact of the percent of ELs

because Vaughn et al. (2017) hypothesized that the effects demonstrated in their study may be a consequence of teachers potentially reducing the amount and type of challenging information, decreasing the opportunities to interact purposefully using peer, group, or class discussion, and overall lower expectations. These hypothesized instructional practices would be more accurately measured at the class rather than school level.

Limitations

Interpreting the overall promising findings requires consideration of several potential limitations. This study was conducted within a relatively small geographic area (about 70-mile radius) in the Southwestern United States, thus, potentially limiting the generalizability of the findings. Also, the vast majority of students identified as ELs spoke Spanish in the home, thus, generalizing to other EL groups needs to be considered. We also recognize that the study would be advantaged with a measure of English and a measure of home-language proficiency. Unfortunately, due to limitations imposed by the school districts who were participating in the study, we were unable to collect this data. Also, one school left the study prior to posttest. We dropped this school from the analytic sample rather than analyzing the data according to intent-to-treat assumptions. The bias associated with losing one school posed a very minimal threat to the study’s validity, as described elsewhere in this article, and we opted to report the findings in the most parsimonious way possible (the findings for the ITT version using a FIML estimator are very similar to those on the trimmed sample; they do not suggest a different pattern of conclusions).

Implications for Practice

This study contributes to the development of an evidence-base for improving instructional practices for students who are ELs in upper elementary grades. Our research team prioritizes developing and evaluating literacy-related practices that are beneficial to all learners in the classroom. The assumption is that teachers are more likely to implement instructional practices that are feasible and integrated into their content learning priorities if these practices are associated with improved outcomes for all of the learners in their classroom. This requires robust instructional practices that demonstrate universal impact on learners and are not unduly taxing to teachers (Vaughn & Wanzek, 2014). The PD model investigated in this study targets codeveloping (teachers and researchers) a set of literacy-related practices implemented during their content area social studies classes. Our confidence in the findings from this study are supported by the quality of the design and the relatively large sample size.

References

- Adey, P. (2004). *The professional development of teachers: Practice and theory*. Springer Science & Business Media.
- Ahmed, Y., Francis, D. J., York, M., Fletcher, J. M., Barnes, M., & Kulesz, P. (2016). Validation of the direct and inferential mediation (DIME) model of reading comprehension in Grades 7 through 12. *Contemporary Educational Psychology, 44*, 68–82. <https://doi.org/10.1016/j.cedpsych.2016.02.002>
- August, D. E., & Shanahan, T. E. (2006). *Developing literacy in second-language learners: Report of the National Literacy Panel on Language-Minority Children and Youth*. Erlbaum Publishers.
- Babinski, L., Amendum, S., Knotek, S., Sánchez, M., & Malone, P. (2018). Improving young English learners' language and literacy skills through teacher professional development: A randomized controlled trial. *American Educational Research Journal, 55*(1), 117–143. <https://doi.org/10.3102/0002831217732335>
- Basma, B., & Savage, R. (2018). Teacher professional development and student literacy growth: A systematic review and meta-analysis. *Educational Psychology Review, 30*(2), 457–481. <https://doi.org/10.1007/s10648-017-9416-4>
- Bates, D., Machler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 48. <https://doi.org/10.18637/jss.v067.i01>
- Baumann, J. F. (2009). Intensity in vocabulary instruction and effects on reading comprehension. *Topics in Language Disorders, 29*(4), 312–328. <https://doi.org/10.1097/TLD.0b013e3181c29e22>
- 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>
- Boardman, A. G., Argüelles, M. E., Vaughn, S., Hughes, M. T., & Klingner, J. (2005). Special education teachers' views of research-based practices. *The Journal of Special Education, 39*(3), 168–180. <https://doi.org/10.1177/00224669050390030401>
- Calderon, M., Slavin, R., & Sánchez, M. (2011). Effective instruction for English learners. *The Future of Children, 21*(1), 103–127. <https://doi.org/10.1353/foc.2011.0007>
- Callahan, R. M. (2005). Tracking and high school English learners: Limiting opportunity to learn. *American Educational Research Journal, 42*(2), 305–328. <https://doi.org/10.3102/00028312042002305>
- Capin, P., Stevens, E. A., Stewart, A. A., Swanson, E., & Vaughn, S. (2021). Examining vocabulary, reading comprehension, and content knowledge instruction during fourth grade social studies teaching. *Reading and Writing, 34*, 1143–1170. <https://doi.org/10.1007/s11145-020-10106-5>
- Cervetti, G. N., Jaynes, C. A., & Hiebert, E. H. (2009). Increasing opportunities to acquire knowledge through reading. In E. H. Hiebert (Ed.), *Reading more, reading better* (pp. 79–100). Guilford Press.
- Coburn, C. (2003). Rethinking scale: Moving beyond numbers to deep and lasting change. *Educational Researcher, 32*(6), 3–12. <https://doi.org/10.3102/0013189X032006003>
- Cunningham, A. E., & Stanovich, K. E. (1998). What reading does for the mind. *American Educator, 22*(1–2), 8–15.
- Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). *Effective teacher professional development*. Learning Policy Institute.
- Datnow, A. (2002). Can we transplant educational reform, and does it last? *Journal of Educational Change, 3*(3), 215–239. <https://doi.org/10.1023/A:1021221627854>
- 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), 1–44. <https://doi.org/10.1080/19345740802539200>
- Espin, C. A., Busch, T. W., Shin, J., & Kruschwitz, R. (2001). Curriculum-based measurement in the content areas: Validity of vocabulary-matching as an indicator of performance in social studies. *Learning Disabilities Research & Practice, 16*(3), 142–151. <https://doi.org/10.1111/0938-8982.00015>
- Gajria, M., Jitendra, A. K., Sood, S., & Sacks, G. (2007). Improving comprehension of expository text in students with LD: A research synthesis. *Journal of Learning Disabilities, 40*(3), 210–225. <https://doi.org/10.1177/00222194070400030301>
- Garet, M. S., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, W., Uekawa, K., Falk, A., Bloom, H., Doolittle, F., Zhu, P., & Szejnberg, L. (2008). *The impact of two professional development interventions on early reading instruction and achievement (NCEE 2008-4030)*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. <https://ies.ed.gov/ncee/pdf/20084030.pdf>
- Garet, M., Porter, A., Desimone, L., Birman, B., & Yoon, K. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal, 38*(4), 915–945. <https://doi.org/10.3102/00028312038004915>
- Gersten, R., Dimino, J., Jayanthi, M., Kim, J., & Santoro, L. (2010). Teacher study group: Impact of the professional development model on reading instruction and student outcomes in first grade classrooms. *American Educational Research Journal, 47*(3), 694–739. <https://doi.org/10.3102/0002831209361208>
- Giles, C. (2006). Sustaining secondary school visions over time: Resistance, resilience and educational reform. *Journal of Educational Change, 7*(3), 179–208. <https://doi.org/10.1007/s10833-005-5727-1>
- Gwet, K. (2001). *Handbook of inter-rater reliability: How to estimate the level of agreement between two or multiple raters*. STATAXIS Publishing Company.
- Hairrell, A., Rupley, W., Edmonds, M., Larsen, R., Simmons, D., Willson, V., Byrns, G., & Vaughn, S. (2011). Examining the impact of teacher quality on fourth-grade students' comprehension and content-area achievement. *Reading & Writing Quarterly: Research on Teacher Quality: Improving Reading and Writing Instruction, 27*(3), 239–260. <https://doi.org/10.1080/10573569.2011.560486>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Hedges, L. V. (2011). Effect sizes in three-level cluster-randomized experiments. *Journal of Educational and Behavioral Statistics, 36*(3), 346–380. <https://doi.org/10.3102/1076998610376617>

- Hoffman, L. (2015). *Longitudinal analysis: Modeling within-person fluctuation and change*. Routledge. <https://doi.org/10.4324/9781315744094>
- Kena, G., Musu-Gillette, L., Robinson, J., Wang, X., Rathbun, A., Zhang, J., Wilkinson-Flicker, S., Barmer, A., & Dunlop Velez, E. (2015). *The Condition of Education 2015 (NCES 2015-144)*. U.S. Department of Education, National Center for Education Statistics. <https://nces.ed.gov/pubs2015/2015144.pdf>
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.
- Klingner, J. K., Vaughn, S., Arguelles, M. E., Tejero Hughes, M., & Ahwee Leftwich, S. (2004). Collaborative strategic reading: "Real-world" lessons from classroom teachers. *Remedial and Special Education, 25*(5), 291–302. <https://doi.org/10.1177/07419325040250050301>
- Klingner, J. K., Vaughn, S., Boardman, A., & Swanson, E. (2012). *Now we get it!: Boosting comprehension with collaborative strategic reading*. Jossey-Bass.
- Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2020). emmeans: Estimated marginal means (R Package version 1.3.4). <https://CRAN.R-project.org/package=emmeans>
- Lesaux, N., Kieffer, M., Faller, S., & Kelley, J. (2010). The effectiveness and ease of implementation of an academic vocabulary intervention for linguistically diverse students in urban middle schools. *Reading Research Quarterly, 45*(2), 196–228. <https://doi.org/10.1598/RRQ.45.2.3>
- MacGinitie, W. H., MacGinitie, R. K., Maria, K., Dreyer, L. G., & Hughes, K. E. (2000). *Gates-MacGinitie Reading Tests (4th ed.)*. Riverside Insights.
- McKeown, M. G., Beck, I. L., & Blake, R. G. (2009). Rethinking reading comprehension instruction: A comparison of instruction for strategies and content approaches. *Reading Research Quarterly, 44*(3), 218–253. <https://doi.org/10.1598/RRQ.44.3.1>
- McLaughlin, M., & Mitra, D. (2001). Theory-based change and change-based theory: Going deeper, going broader. *Journal of Educational Change, 2*(4), 301–323. <https://doi.org/10.1023/A:1014616908334>
- National Center for Educational Statistics, United States. Office of Educational Research, Improvement. Center for Education Statistics, & Institute of Education Sciences (US). (2009). *The condition of education*. US Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.
- Opfer, V. D., Pedder, D. G., & Lavicza, Z. (2011). The role of teachers' orientation to learning in professional development and change: A national study of teachers in England. *Teaching and Teacher Education: An International Journal of Research and Studies, 27*(2), 443–453. <https://doi.org/10.1016/j.tate.2010.09.014>
- Parise, L. M., & Spillane, J. P. (2010). Teacher learning and instructional change: How formal and on-the-job learning opportunities predict change in elementary school teachers' practice. *The Elementary School Journal, 110*(3), 323–346. <https://doi.org/10.1086/648981>
- Porche, M., Pallante, D., & Snow, C. (2012). Professional development for reading achievement: Results from the Collaborative Language and Literacy Instruction Project (CLLIP). *The Elementary School Journal, 112*(4), 649–671. <https://doi.org/10.1086/665008>
- Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interactions in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics, 31*(4), 437–448. <https://doi.org/10.3102/10769986031004437>
- Richardson, V. (2003). The dilemmas of professional development. *Phi Delta Kappan, 84*(5), 401–406. <https://doi.org/10.1177/003172170308400515>
- Simmons, D., Hairrell, A., Edmonds, M., Vaughn, S., Larsen, R., Willson, V., Rupley, W., & Byrns, G. (2010). A comparison of multiple-strategy methods: Effects on fourth-grade students' general and content-specific reading comprehension and vocabulary development. *Journal of Research on Educational Effectiveness, 3*(2), 121–156. <https://doi.org/10.1080/19345741003596890>
- Stevens, E. A., Vaughn, S., Swanson, E., & Scammacca, N. (2020). Examining the effects of a Tier 2 reading comprehension intervention aligned to tier 1 instruction for fourth-grade struggling readers. *Exceptional Children, 86*(4), 430–448. <https://doi.org/10.1177/0014402919893710>
- Swanson, E., Barnes, M., Fall, A., & Roberts, G. (2018). Predictors of reading comprehension among struggling readers who exhibit differing levels of inattention and hyperactivity. *Reading & Writing Quarterly, 34*(2), 132–146. <https://doi.org/10.1080/10573569.2017.1359712>
- Swanson, E., Solis, M., Ciullo, S., & McKenna, J. W. (2012). Special education teachers' perceptions and instructional practices in response to intervention implementation. *Learning Disability Quarterly, 35*(2), 115–126. <https://doi.org/10.1177/0731948711432510>
- Swanson, E., Stewart, A., Stevens, E. A., Scammacca, N., Capin, P., Hamilton, B. J., Roberts, G., & Vaughn, S. (2020). *Investigating the role of professional development in impacting upper elementary student reading outcomes* [Manuscript submitted for publication]. Department of Special Education, The University of Texas at Austin.
- Swanson, E., Wanzek, J., Mcculley, L., Stillman-Spisak, S., Vaughn, S., Simmons, D., Fogarty, M., & Hairrell, A. (2016). Literacy and text reading in middle and high school social studies and English language arts classrooms. *Reading & Writing Quarterly, 32*(3), 199–222. <https://doi.org/10.1080/10573569.2014.910718>
- Vaughn, S., Martinez, L. R., Wanzek, J., Roberts, G., Swanson, E., & Fall, A. (2017). Improving content knowledge and comprehension for English language learners: Findings from a randomized control trial. *Journal of Educational Psychology, 109*(1), 22–34. <https://doi.org/10.1037/edu0000069>
- Vaughn, S., & Wanzek, J. (2014). Intensive interventions in reading for students with reading disabilities: Meaningful impacts. *Learning Disabilities Research & Practice, 29*(2), 46–53. <https://doi.org/10.1111/ldrp.12031>
- Wasik, B. A., & Hindman, A. H. (2011). Improving vocabulary and pre-literacy skills of at-risk preschoolers through teacher professional development. *Journal of Educational Psychology, 103*(2), 455–469. <https://doi.org/10.1037/a0023067>
- Wexler, J., Mitchell, M. A., Clancy, E., & Silverman, R. D. (2016). An investigation of literacy practices in science classrooms. *Reading & Writing Quarterly, 33*(3), 258–277. <https://doi.org/10.1080/10573569.2016.1193832>
- What Works Clearinghouse. (2020). *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.

Received June 24, 2020

Revision received March 24, 2021

Accepted March 26, 2021 ■