Class Percentage of Students With Reading Difficulties on Content Knowledge and Comprehension

Sharon Vaughn, PhD[1], Anna-Mária Fall, PhD[1],
Greg Roberts, PhD[1], Jeanne Wanzek, PhD[2],
Elizabeth Swanson, PhD[1], and Leticia R. Martinez, PhD[1]

Abstract

We examined the efficacy of a content acquisition and reading comprehension intervention implemented in eighth-grade social studies classrooms. Using a within-teacher randomized control design, 18 eighth-grade teachers' social studies classes were randomly assigned to a treatment or comparison condition. Teachers taught all their classes (treatment and comparison) using the same content; however, in the treatment classes, teachers used instructional practices that included comprehension canopy, essential words, knowledge acquisition, and team-based learning. Students with reading comprehension difficulties in the treatment classes (*n* = 359) outperformed students with reading comprehension difficulties in the comparison classes (*n* = 331) on measures of content knowledge acquisition and content reading comprehension but not general reading comprehension. In addition, the proportion of students with reading comprehension difficulties in classes moderated outcomes for content knowledge acquisition and content reading comprehension.

Almost a quarter of eighth-grade students do not read well enough to understand important concepts and make simple inferences from text-based materials (National Center for Education Statistics, 2015). These students also struggle to understand the meaning of words well enough to use them to comprehend a sentence or paragraph. Unfortunately, these reading difficulties too often result in a difficult postsecondary college pathway. Approximately 68% of community college students are required to take at least one remedial course (Chen, 2016), termed by some as the "Bermuda Triangle of higher education" because only 28% of remedial course takers earn a college degree within 8 years (Attewell, Lavi, Domina, & Levey, 2006). Based on ACT scores (Noble, Roberts, & Sawyer, 2006), the clearest differentiator between students who are college ready and students who are not is the ability to comprehend complex texts. If the nation's K–12 education goal is to prepare every student for meaningful postsecondary experiences (Executive Office of the President, 2015), designing and implementing effective instruction for middle-grade students requires considering ways to improve their reading comprehension.

Enhancing Reading Comprehension in the Middle Grades

Students who read proficiently read more text and acquire a broader vocabulary and understanding of content-related

constructs, advantaging them not only in knowledge acquisition but also more broadly in general reading comprehension. On the other hand, students who demonstrate reading comprehension difficulties read less text and score lower on achievement tests in part because of underdeveloped background knowledge and vocabulary (Ahmed et al., 2016). As students move from elementary to middle school, the demands for sophisticated language, literacy, and background knowledge increase. This trend combined with underdeveloped background knowledge and vocabulary is particularly foreboding for secondary students, when reading for understanding is such an integral part of learning. These diverse student-learning characteristics require teachers to access instructional practices that are beneficial to a range of learners.

The need to enhance reading comprehension in the middle grades is also documented in students' unsatisfactory trajectory for developing reading comprehension after the elementary grades (Mancilla-Martinez, Kieffer,

[1]University of Texas at Austin, Austin, TX, USA
[2]Vanderbilt University, Nashville, TN, USA

Corresponding Author:
Sharon Vaughn, Meadows Center for Preventing Educational Risk, University of Texas at Austin, 1912 Speedway, D4900, SZB 228, Austin, TX 78712-1284, USA.
Email: srvaughn@austin.utexas.edu

Biancarosa, Christodoulou, & Snow, 2011). Further exacerbating this challenge is that substantial numbers of students demonstrate "late-emerging" reading difficulties, indicating that their reading problems emerge after third grade (Catts, Compton, Tomblin, & Bridges, 2012; Kieffer, 2014). This finding suggests that many students can master the foundational skills of word reading with adequate fluency but that as the syntax, vocabulary, and background knowledge of texts become more complex, their reading difficulties manifest. Supporting the reading comprehension of students with reading difficulties requires instructional practices that move beyond developing the foundational skills of reading and tackle the challenges of text-based content learning in the secondary grades. This lofty goal requires that reading comprehension instruction be integrated into content area classroom practices.

Reading for Understanding Initiative

As part of the Reading for Understanding Initiative (Douglas & Albro, 2014), a network of more than 130 researchers have worked since 2010 to advance knowledge and practice around reading for understanding across the grade span (including students at the secondary level). As part of the Reading for Understanding mechanism, we developed and implemented Promoting Adolescents' Comprehension of Text (PACT), an approach that uses text sources along with team-based learning (TBL; Michaelson & Sweet, 2011) as critical foci for content acquisition, vocabulary growth, and reading comprehension gains (Vaughn et al., 2013). Using a randomized control trial design, we evaluated the efficacy of the multicomponent intervention. We also tested in isolation the efficacy of the TBL components (Wanzek et al., in press). Findings indicate that when eighth-grade teachers' social studies classes are randomly assigned to PACT treatment or typical instruction conditions (i.e., blocking on teacher and content foci), students in treatment classes consistently demonstrate statistically significantly higher scores on content vocabulary and knowledge (effect size [ES] range from .17 to 40; Vaughn et al., 2013, 2017), practically significantly higher scores on reading comprehension in the content area (ES range from .20 to .59; Vaughn et al., 2013, 2015, 2017), and though not statistically significant, higher ES scores on distal measures of reading comprehension (Vaughn et al., 2013, 2017). PACT has demonstrated efficacy through previous randomized control trials with middle and high school students, students with disabilities (Swanson, Wanzek, Vaughn, Roberts, & Fall, 2015), and students who were English learners (ELs). Thus, we suggest that PACT holds promise for universal instructional supports for a range of learners.

Enhancing PACT for All Learners

PACT represents a bundle of instructional practices woven into content area instruction (i.e., social studies) to enhance content learning and comprehension for all learners. Using the PACT instructional practices as a foundation, we integrated research-based knowledge derived from multiple sources, including practice guides (Baker et al., 2014; Francis, Rivera, Lesaux, Kieffer, & Rivera, 2007), to enhance the features of instruction and promote best practices for meeting the needs of a range of learners (e.g., students with reading difficulties, ELs).

In the PACT treatment, informational text reading that included target vocabulary was central to every unit and anchored the instruction of academic vocabulary. Essential words in each unit were taught explicitly and reinforced by reading, speaking, and writing activities where students applied the meaning of the words in multiple and meaningful contexts. Academic vocabulary teaching was enhanced in this version of PACT through instruction on more abstract terms that students need to know to communicate across the disciplines and complete school tests and tasks (e.g., academic expressions for comparing and contrasting and using cause and effect; Lesaux, Keiffer, Faller, & Kelley, 2010).

In addition to the TBL that was used in the original versions of PACT, we provided structured opportunities for students to participate in academic discussions around content and text and writing tasks that supported the use of learned content vocabulary and discourse (August, Branum-Martin, Cardenas-Hagan, & Francis, 2009; Lesaux, Kieffer, Kelley, & Harris, 2014). For example, in knowledge application activities, students were taught and expected to justify their answers by using learned academic vocabulary and citing evidence from informational texts.

Students worked in pairs or small groups during most PACT components to prepare for discussions and to write responses to inferential questions and summaries focused on building knowledge and developing academic language. Instruction on new social studies content was supplemented with brief videos, visuals, and graphic organizers to provide students the necessary background information to participate in academic discourse. Finally, the TBL comprehension checks and knowledge application activities provided continuous targeted feedback, where teachers affirmed or corrected students' understanding of the content and facilitated further discussion.

In addition to knowledge and reading comprehension, the PACT instructional practices were well aligned with best practices for teaching ELs; thus, we conducted a randomized control trial examining the efficacy of PACT in classes with at least one EL. Findings from this study revealed that ELs in the treatment condition performed statistically significantly higher on vocabulary and content acquisition as well as reading comprehension in the

content area than students in traditional social studies instruction (Vaughn et al., 2017). Furthermore, we learned that the proportions of ELs in classes moderated outcomes. When the percentage of ELs exceeded a threshold of about 14%, performance in the treatment condition was negatively affected.

Because the sample from this study included students with significant reading comprehension problems in each class, we sought to examine (a) the extent to which the performance of these students in treatment classes was associated with improved outcomes when compared with students with reading difficulties in comparison classrooms and (b) the extent to which the proportion of students with struggling-reader status in a class moderated outcomes.

Class Composition and Reading Outcomes

Many teachers perceive that student heterogeneity influences the quality of their instruction and that excessive heterogeneity is a problem that substantially affects their instruction and student learning (Markow & Cooper, 2008). Evidence suggests that there is support for teachers' views that peer composition at the class level influences student reading outcomes (Ammermueller & Pischke, 2009) and that more homogeneous grouping within classes can be associated with improved student outcomes (Scammacca et al., 2016).

Although there is considerable speculation about problems associated with too much heterogeneity in a class, there are also challenges when the class composition represents too many students with learning difficulties. There are several possible explanations for why classes that do not have a range of achievement levels can be problematic, including the following: (a) teachers have higher expectations for students in classes with a considerable number of high achievers and alter the quality of their instruction accordingly; (b) all students benefit from the participation of peers representing a range of achievement levels, potentially increasing the quality and range or oral responses and group work; and (c) the academic climate may be improved such that students are more engaged and motivated to learn when there is a range of achievement levels perhaps also attributed to a greater range of responses and even more high-quality responses that engage all students in the class (e.g., Dar & Resh, 1986; Dreeben & Barr, 1988; Good & Marshall, 1984; Hallinan, 1988; Pallas, Entwisle, Alexander, & Stluka, 1994).

Thus, when classes are composed of mostly struggling readers, instruction and learning may be affected. To the extent that students in classes with significant numbers of students with low levels of literacy are perceived by teachers as less capable, it might be expected that opportunities for rich language discourse and content learning are

adjusted downward. Earlier studies support this supposition. The negative consequences associated with low expectations for groups of students has been demonstrated with ELs. For example, Callahan (2005) reported that ELs were primarily clustered in non-college-preparatory classes—with deleterious effects on learning opportunities and, thus, achievement. Within such classes, Vaughn et al. (2017) reported an interaction between EL class composition and reading comprehension outcomes, with student outcomes on content and comprehension measures negatively affected when the composition of ELs in a class increased beyond about 14%.

To date, the literature addressing peer effects on learning (e.g., Angrist & Lang, 2004; Gottfried, 2014; Hanushek, Kain, Markman, & Rivkin, 2003) has focused on socioeconomic status and prior achievement in large-scale extant databases. We are unaware of studies that consider the class-level proportion of struggling readers and its moderating effect on the relationship of treatment and reading comprehension and content knowledge outcomes for individual struggling readers. We model this three-way interaction using extant data from a randomized control trial of the PACT treatment for eighth-grade students (Vaughn et al., 2017). PACT has demonstrated significant main effects in past trials (Vaughn et al., 2013, 2015, 2017). In this paper, we focus on its conditional effects for struggling readers in classrooms that vary in the proportion of struggling readers. We hypothesized that PACT's effect would vary for struggling readers in classrooms that differed in proportion of struggling readers.

Method

Research Design

We conducted a within-teacher randomized block design to study the effects of PACT in eighth-grade social studies classes (Vaughn et al., 2017). Schools were selected to participate if they served a population with large numbers of ELs. The distribution of ELs varied at the class level, with at least one identified EL being required for participation. All eighth-grade social studies teachers' class sections were randomly assigned to the PACT intervention condition or the typical practice condition. Each teacher provided the same U.S. history content to students in the study, but the PACT instructional practices were implemented in treatment classes only. For this study, we examined the outcomes of students identified as struggling readers by scoring 85 or below at pretest on the *Gates-MacGinitie Reading Comprehension subtest* (MacGinitie, MacGinitie, Maria, Dreyer, & Hughes, 2006). We sought to investigate the moderating effects of the class percentage of struggling readers on content acquisition and reading comprehension.

Setting and Participants

Schools. The PACT study was implemented during the 2013–2014 academic year across seven middle schools in three school districts in two distinct areas of the United States. Three schools were in the southwestern United States, with two schools in a large, diverse urban district and another in a smaller, predominantly Hispanic suburban district. The proportion of students in each school who qualified as having reading comprehension problems ranged from 28% to 79%. The proportion of students across all schools who qualified for free or reduced-price lunch ranged from 49.2% to 96.8%.

Teachers. The 18 teacher participants (nine females and nine males) were eighth-grade U.S. history teachers who implemented the intervention with researchers' support in treatment classes and continued with typical instruction in comparison classes. All the teachers had a bachelor's degree, and five of them had a master's degree. Their teaching experience ranged from 0 to 34 years ($M = 10.13$, $SD = 10.48$). Teachers' ethnicities included White (83.3%), Hispanic (16.7%), and Asian (5.6%).

Students. A total of 1,629 eighth-grade students were assigned to 94 U.S. history class sections. Classes were randomly assigned within teacher to 49 treatment classes (845 students) or 45 comparison (784 students) classes. When teachers had an odd number of classes, randomization assigned extra classes to treatment. Of the total participants, 42% were identified as having reading comprehension difficulties (at or below 15 standard score points, one standard deviation, on the *Gates-MacGinitie Reading Comprehension subtest*; MacGinitie et al., 2006); however, the distribution of students with reading difficulties in classes ranged from 0% to 100%. Additional demographic information is displayed in Table 1. Figure 1 depicts the distribution of struggling readers across classrooms.

Intervention Procedures

Students in the treatment classes received the intervention during their regularly scheduled eighth-grade social studies class, which lasted approximately 50 min daily or 90 min every other day. In the first 6 to 8 weeks of the fall semester, three successive units were taught in treatment classes. Following the three units, teachers taught the knowledge acquisition through text reading intervention component for an additional 12 weeks. This extended single component was implemented three times a week for approximately 15 min per session. Typical instruction covering the same content for the same amount of time was provided for students who were randomly assigned to comparison classes.

Table 1. Student-Level Demographics by Group.

Demographic	Comparison		Treatment	
	n	%	*n*	%
Gender				
Male	333	42.5	426	50.4
Female	416	53.1	385	45.6
Missing	35	4.4	34	4
Race				
Caucasian	490	41.11	482	37.73
African American	115	9.65	105	8.2
Hispanic	467	39.18	523	40.92
Asian	14	1.17	30	2.35
American Indian	98	8.22	129	10.1
Pacific Islander	8	0.67	9	0.7
Special education	65	8.90	95	11.90
English language learners	190	26	245	30.7
Struggling readers	331	46.8	359	46.2
Free or reduced-price lunch	457	71.1	487	71.4

Description of the Treatment Intervention

We aligned the PACT intervention practices with the standards outlined by participating districts and the Common Core State Standards. PACT comprises five components implemented throughout a 10-lesson instructional unit, as described below.

Comprehension canopy. The comprehension canopy is introduced on Day 1 of the 10-lesson cycle. It is a roughly 15-min activity in which the teachers identify a purpose for learning and reading about the content. While the students think about new content, teachers build background knowledge. Students view a brief video, shorter than 5 min, that sets the context for what they will read about and discuss during the unit. Following the video, student partners discuss a question that was posed to them prior to watching. After debriefing about the video, teachers introduce an overarching question that guides students' learning throughout the unit. As the focal point of the unit, the question is revisited and built on every day for the remainder of the unit. This question fosters student dialogue and practice using academic language that is specific to social studies (e.g., cause and effect, perspective taking).

Essential words. Five concepts essential to learning the content are explicitly taught on the 1st day of the unit by using a vocabulary instructional routine. Each concept is introduced with a student-friendly description and visual representation. Students are provided with related words, sentences as examples of word use, and nonexamples. A question prompt to guide students to apply their knowledge of the concept to a personal experience or scenario is shared with the class.
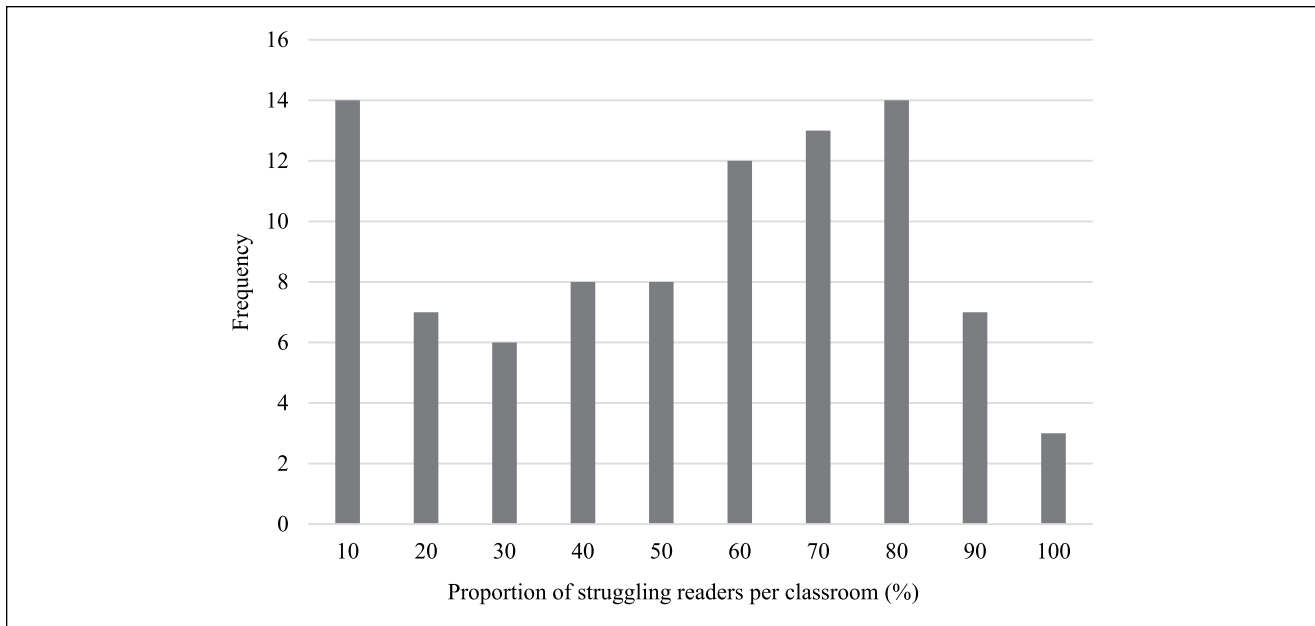
Figure 1. Proportion of struggling readers per classroom.

Student partners discuss responses for about 2 min, and after the teacher listens to a few discussions, students are selected to share their partner discussion with the class. Essential words are revisited in subsequent days during other activities, like daily warm-up exercises, text reading, and TBL knowledge application. Students receive repeated exposure to the words and opportunities to practice applying the meaning of the word in multiple contexts.

Knowledge acquisition through text reading. The intervention supports reading comprehension through the reading of primary and secondary sources three times throughout each unit. Teachers guide students through this 15- to 20-min routine to provide an organizational structure for student learning through informational text. Prior to reading a text, the teacher shares a video or visual that helps students build understanding. The teacher reviews essential words and their meaning in the text as well as relates the text to the comprehension canopy question. During this routine, students are grouped in a variety of formats, depending on the students' needs. During the first unit, the teacher may choose to lead the students through the readings, but as they progress through the units, students can be paired or work in a small group. While reading, students address inferential questions of varying difficulty, and in preparation for peer discussions, students record important information from the text in their learning logs.

TBL comprehension checks. TBL was adapted from a university-level practice to provide students with opportunities for text-based discussions (Michaelsen & Sweet, 2011). Heterogeneous teams work together to extend their thinking by

demonstrating their understanding of the content and thinking critically about other team members' contributions to content discussions.

Each PACT unit has two TBL comprehension checks to confirm that students understand the content to that point and to provide the teacher with information on how students are progressing. Each TBL comprehension check is a short quiz with 10 questions and five vocabulary items. To begin, students take the quiz individually and turn it in to the teacher to be graded. Thereafter, a small team or partners complete the same comprehension check while using their learning logs and unit readings. Students are encouraged to use their materials to justify their answers and discussions with their group. For every quiz item, students must agree on the answer and then indicate their answer on provided scratch-off cards. If the answer is correct, a star is revealed. However, if the answer is incorrect, the team or partners must return to their notes and further discuss and support their new answer with text evidence. Based on students' performance on their individual and group checks, the teacher reteaches or reviews if necessary.

TBL knowledge application. At the close of the unit, groups of students work together on the TBL knowledge application problem-solving activity to apply and extend their understanding of the unit content. For example, in one unit, the groups decide on the three most important causes of the American Revolution. While student teams discuss, they must cite text evidence to support their contributions. The teacher walks around to facilitate discussions and provide feedback to students. The teams then prepare and present their responses and justifications to the class. The teacher

closes the activity by connecting it to the comprehension canopy question that guided the unit and summarizes a response from students' contributions.

Implementation Support

The research team provided professional development and instructional support for participating teachers. Prior to intervention implementation, teachers participated in a 1-day intensive professional development session that prepared them to apply the PACT components in treatment classes. To begin, researchers addressed the purpose of the project and reviewed research procedures, including that teachers were to maintain typical practice in comparison classrooms. The professional development focused on explaining and modeling each instructional component in Unit 1, as well as providing opportunities for teachers to participate as the students and later as instructors. Teachers were provided an additional 3 hr of professional development after the implementation of the first unit. Throughout the booster session, research support personnel revisited each PACT component but targeted the instructional practices they observed as being the most difficult for teachers, such as the critical reading of text.

To further support PACT implementation in the classroom, research support personnel helped teachers develop competence in the application of the intervention components. Every teacher was assigned a research support person, who assisted the teacher for the duration of the 6 to 8 weeks to complete the three PACT units. During the first unit, research support was present daily to model, co-teach, provide feedback, and answer questions. Research support reduced visits to two to three times a week during the second unit and one to two times a week during the last unit. The research support personnel also helped teachers plan on how to embed the intervention components within the extensive amount of content. After the completion of the PACT units, while teachers continued to implement the knowledge-acquisition-through-text-reading component in their treatment classes, research personnel met with teachers at least three times in 12 weeks to inquire about the amount of reading they did weekly.

Implementation Fidelity

Fidelity data were collected throughout the implementation of the PACT intervention. In this study, treatment integrity was measured through audio recordings of teachers' instruction in one randomly selected treatment class. Each teacher was provided a lanyard with an audio recorder that they were required to wear daily during the identified intervention class period to capture instruction of every lesson that spanned the three PACT units. Additionally, we asked the teachers to record their instruction for the same time frame

in one randomly selected comparison class section to evaluate the presence of the intervention features in typical practice. Every teacher submitted approximately 30 treatment audio recordings and 30 typical-practice audio recordings. Research assistants were then trained to code recordings of the last two units of instruction of each condition per teacher.

A fidelity-of-implementation measure was developed to document teachers' adherence to the intervention components and the quality of PACT implementation. The instrument was aligned with the components of the intervention (see Description of the Treatment Intervention section). We coded fidelity by rating each of the instructional components on a 4-point Likert-type rating scale ranging from 1 (*low implementation*) to 4 (*high implementation*) to assess the extent to which the teacher completed each PACT component. A *not applicable* (0) rating was an option for coders because a component was not required every day of the unit. Figure 1 presents how many times each component was required during each unit in treatment classes. Moreover, global quality ratings (1 to 5) were assigned to each audio recording to report on a teacher's overall instructional quality, classroom management, and implementation of the PACT intervention.

Before formal coding occurred, intercoder agreement was established with a team of six coders. Using the gold-standard method, coders were required to meet a 90% interrater agreement score established by senior researchers who developed the PACT intervention. After the initial agreement, the coding team did two additional interrater reliability checks with the gold standard to avoid observer drift.

Table 2 demonstrates the analyzed fidelity data in treatment and comparison classes. Across the intervention, fidelity of implementation was at a mid-high level with mean component scores ranging from 2.76 to 3.63. The one component that was rated low or mid-low most frequently was knowledge acquisition through text reading ($M = 2.76$). Although reading occurred consistently, teachers often left out instructional practices integral to critical reading, such as facilitating discussion through the reading and tying the reading back to the essential words or the comprehension canopy question.

In comparison classes, data indicate that few PACT elements were observed. The team coded audio recordings for the comparison classes by using the same rating scale used for fidelity to treatment to examine the extent to which PACT components were present in typical-practice classes. It was determined that the amount of overlap in intervention components between the two conditions was minimal. As Table 2 shows, PACT components were present at a low rate or not observed at all. When elements of a PACT component were detected in audio recordings, they did not resemble the complete PACT component. For example, a teacher might begin the class with a warm-up activity, but

Table 2. Frequency for Fidelity Observations in Treatment and Comparison Classrooms.

Classroom and frequency	CC		WU		TBLC		EW		KA		TBLK	
	n	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
Treatment classrooms	(31)		(130)		(64)		(32)		(108)		(34)	
4 = High	15	48.4	105	80.8	12	18.8	17	53.1	7	6.5	5	14.7
3 = Mid-high	11	35.5	6	4.6	42	65.6	11	34.4	74	68.5	27	79.4
2 = Mid-low	4	12.96	15	11.5	3	4.7	3	9.4	21	19.4	2	5.9
1 = Low	1	3.2	4	3.1	7	10.9	1	3.1	6	5.6		
0 = Not observed												
Comparison classrooms	(274)		(274)		(274)		(274)		(274)		(274)	
4 = High			26	9.5			2	0.7				
3 = Mid-high							2	0.7	12	4.4		
2 = Mid-low	8	2.9	39	14.2	2	.7	6	2.2	34	12.4	3	1.1
1 = Low									11	4.0		
0 = Not observed	266	97.1	209	76.3	272	99.3	264	96.4	217	79.2	271	98.9

Note: CC = comprehension canopy; EW = essential words; KA = knowledge acquisition; TBLC = team-based learning comprehension check; TBLK = team-based learning knowledge application; WU = warm-up.

the warm-up was not connected to the content that students were studying. This finding is similar to those of prior PACT implementations, where differential instruction for treatment and control sections were noted (Vaughn et al., 2013, 2015).

Student Measures

We used the same measures of impact employed in the prior studies implementing PACT (Vaughn et al., 2013, 2015, 2017). Research personnel uninformed of the condition to which students were assigned administered all three measures to students in the treatment and comparison groups prior to and immediately following treatment.

Gates-MacGinitie Reading Comprehension subtest. The *Gates-MacGinitie Reading Comprehension subtest* (fourth edition; MacGinitie et al., 2006) is a group-administered, timed (35-min) assessment of reading comprehension. The assessment consists of expository and narrative passages ranging in length from three to 15 sentences. Students read each passage silently and answer three to six multiple-choice questions related to the most recently read passage. As students progress through the assessment, items increase in difficulty. Internal consistency reliability ranges from 0.91 to 0.93, and alternate form reliability is reported as 0.80 to 0.87.

Assessment of Social Studies Knowledge (ASK). The researcher-developed ASK (Vaughn et al., 2013) is a 42-item, four-option, untimed multiple-choice test that measures content knowledge in the three units of the intervention ("Colonial America," "Road to Revolution," "The

American Revolution"). Items with known difficulty parameters were collected with permission from released state and Advanced Placement social studies tests from Texas, Massachusetts, and the College Board. Researcher-developed vocabulary items were also included in the item set. The ASK was administered at pretest and posttest.

The items for the ASK were selected after a series of pilot tests to validate the provided difficulty parameters, refine the instructions for test administrators, and estimate the time necessary for administration. The final items were selected following a series of item-level confirmatory factor analyses to evaluate model fit and estimate item parameters (Vaughn et al., 2013; Wanzek et al., in press). The Cronbach's alpha with the current sample was .93.

Modified Assessment of Social Studies Knowledge and Comprehension (MASK). The MASK (Vaughn et al., 2015) is a 21-item, four-option, untimed multiple-choice test that measures reading comprehension in the content area. The assessment consists of three reading passages drawn from the ASK (Vaughn et al., 2013, 2015) used in previous PACT studies but with altered Lexile levels. For this assessment version, the Lexile range was between 1,090 and 1,140, and the word count range was between 312 and 349. Each passage is related to content covered in the three 10-day cycles. Students read each passage silently and immediately answer seven multiple-choice questions about the passage. Reading comprehension items were researcher developed and measured students' ability to identify main ideas, understand vocabulary in context, identify cause and effect, and summarize. The MASK reading comprehension measure was administered at pretest and posttest. The alpha coefficient with the current sample was .92.

Data Analysis

Our primary interest was in treatment's conditional effect on struggling readers in differently composed classrooms (i.e., classwide proportion of struggling readers). In this section, we describe the model-building process. In the next section (Results), we report and interpret the fully conditional model for each outcome. A conditional model in this case included (a) the three-way interaction of intervention status, struggling-reader status, and classwide proportion of struggling students; (b) all possible two-way interactions among the three main effects; and (c) the main effects for treatment, student status as struggling or not struggling, and classwide proportion of struggling students. Note that the main effects are conditional on the two-way and three-way interactions and that the two-way interactions are conditional on the higher-order interaction. Each term is meaningful only in the context of the others.

To begin, we fit multilevel regression models with students in classes and classes nested in teachers. For each of the three reading outcomes, we fit three models. The first, an unconditional model (i.e., empty means, random intercept), estimated the variance in reading outcomes due to (a) differences between students within classrooms (Level 1), (b) differences between classrooms (Level 2), and (c) differences between teachers (Level 3). About one third of the variance in reading outcomes was between classes (from 29% to 35%, depending on the outcome), and another 10% (from 10% to 12%, depending on outcome) was between teachers (see Tables 3 and 4).

Accordingly, we added student- and classroom-level fixed effects in Model 2. We included the grand-mean-centered pretest score and a dummy-coded variable for struggling reader status (1 = struggling, 0 = nonstruggling) at the student level, and we modeled the proportion of struggling readers per classroom at Level 2, centered at its mean (.47) and divided by 10 to aid with interpretation. There was an average of 17 students per class in the study sample, so .10 corresponds to approximately two students. Intervention status was modeled at the class level, with treatment coded as 1 and business as usual (BaU) as 0. In Model 3, we evaluated random components of the model by allowing the intercept and covariates to vary across levels. With the exception of struggling-reader status, the addition of random slopes did not improve the model's fit (*p* values ranged from .74 to .89). Accordingly, we report the model with pretest scores as a fixed effect at Level 1, struggling-reader status as a random effect at Level 1, and intervention status and classwide proportion of struggling readers as Level 2 fixed effects.

To evaluate potential interactions of struggling-reader status, intervention status, and the classwide proportion of struggling readers in predicting reading outcomes, we modeled relevant cross-level and between-class, within-level interaction terms. The cross-level terms included the interaction of (a) struggling-reader status with intervention status, (b) struggling-reader status by classwide proportion of struggling readers, and (c) struggling-reader status by intervention status by classwide proportion of struggling readers. The between-class, within–Level 2 term was the interaction of intervention status and classwide proportion of struggling readers. BaU and nonstruggling students were used as reference groups (i.e., coded as 0). We treated classwide proportion of struggling readers as continuous and calculated regions of significance using methods described by Bauer and Curran (2005). We estimated effect sizes as a ratio of the model-derived coefficient for a given parameter and the pooled within-group standard deviation across conditions at posttest.

Results

Attrition and Missing Data

Descriptive statistics for each outcome by intervention group and by struggling-reader status are presented in Table 5. Differential attrition was less than 2% (1.6% for ASK,

Table 3. Unconditional Multilevel Model Results: Fixed Effects.

Measure	Estimate	SE	p value
ASK			
Intercept	21.65	0.95	<.001
MASK			
Intercept	9.30	0.52	<.001
Gates-MacGinitie			
Intercept	93.07	1.50	<.001

Note: ASK = *Assessment of Social Studies Knowledge*; *Gates-MacGinitie* = *Gates-MacGinitie Reading Comprehension subtest*; MASK = *Modified Assessment of Social Studies Knowledge and Comprehension.*

Table 4. Unconditional Multilevel Model Results: Random Effects.

Measure	Estimate	p value	ICC
ASK			
Level 1: Students	43.63	<.001	.53
Level 2: Classes	28.54	<.001	.35
Level 3: Teachers	9.63	.03	.12
MASK			
Level 1: Students	14.96	<.001	.59
Level 2: Classes	7.38	<.001	.29
Level 3: Teachers	2.90	.03	.11
Gates-MacGinitie			
Level 1: Students	128.82	<.001	.57
Level 2: Classes	74.18	<.001	.33
Level 3: Teachers	21.83	.01	.10

Note: ICASK = *Assessment of Social Studies Knowledge*; *Gates-MacGinitie* = *Gates-MacGinitie Reading Comprehension subtest*; C = intraclass correlation; MASK = *Modified Assessment of Social Studies Knowledge and Comprehension.*

Table 5. Descriptive Statistics for Reading Outcomes.

Measure	Pretest			Posttest		
	n	*M*	*SD*	*n*	*M*	*SD*
ASK						
Nonstruggling readers						
Comparison	362	19.38	6.71	311	25.05	7.84
Treatment	393	20.49	7.17	362	29.19	7.74
Struggling readers						
Comparison	317	12.79	4.32	287	15.93	6.15
Treatment	347	12.30	4.23	312	17.95	7.15
MASK						
Nonstruggling readers						
Comparison	357	9.62	4.57	305	12.34	4.75
Treatment	399	9.74	4.81	352	12.41	4.91
Struggling readers						
Comparison	302	5.24	2.48	254	6.83	3.13
Treatment	343	4.82	2.54	272	6.69	3.34
Gates-MacGinitie						
Nonstruggling readers						
Comparison	376	103.18	10.21	311	101.77	12.20
Treatment	418	104.95	10.58	365	102.81	14.04
Struggling readers						
Comparison	331	79.32	7.43	264	85.46	9.90
Treatment	359	79.36	7.45	271	84.60	10.47

Note: ASK = Assessment of Social Studies Knowledge; Gates-MacGinitie = Gates-MacGinitie Reading Comprehension subtest; MASK = Modified Assessment of Social Studies Knowledge and Comprehension.

0.2% for MASK, and 0.4% for *Gates-MacGinitie*). Pretest variables yielded no significant intervention-by-attrition interactions. Missing data were handled using full-information maximum-likelihood estimation with robust standard errors and a mean-adjusted chi-square test statistic (MLR) in Mplus 7.31 (Muthén & Muthén, 2010).

ASK

The main treatment effect on ASK was statistically significant ($\beta = 3.54$, $p < .001$, ES = .40) meaning that nonstruggling students in treatment classes scored significantly higher than nonstruggling students in BaU classes, when controlling for classwide proportion of struggling students (i.e., when modeling classwide proportion of struggling readers as .47, which is its mean). The predicted scores for struggling and nonstruggling readers differed as well ($b = -3.15$, $p < .001$, ES = -.42). Struggling readers in BaU classrooms scored 3.15 points lower than nonstruggling readers in the same BaU classroom, on average, when modeling classwide proportion of struggling readers as its mean (.47). The main effect for classwide proportion of struggling readers, which differed significantly from 0 ($b = -0.65$, $p < .001$), indicates that for each 10% increase in the proportion of

struggling readers in a typical BaU comparison classroom, posttest scores on the ASK decreased by .65 points.

The coefficient for the two-way interaction of intervention status by struggling-reader status did not differ statistically from 0 ($b = .02$, $p = .98$). This means that differences in the outcomes for struggling and nonstruggling readers did not depend on assignment to treatment or BaU, when controlling for classwide proportion of struggling readers. The interaction of struggling-reader status by classwide proportion of struggling readers was positive and differed significantly from 0 ($b = .37$, $p = .01$), suggesting that the advantage of nonstruggling readers narrows in BaU classrooms as the classwide proportion of struggling readers increases. For the two-way interaction between intervention status and classwide proportion of struggling readers, the coefficient did not differ significantly from zero, indicating that the effect of treatment is the same across values for classwide proportion of struggling readers ($b = -.14$, $p = .41$), but only for the group of nonstruggling students. Finally, the significant three-way interaction ($b = -.72$, $p = .00$) for intervention status by struggling-reader status by classwide proportion of struggling readers means that the interaction of intervention status and struggling-reader status depends on the classwide proportion of struggling students.

Table 6. Fixed and Random Effects for Reading Outcomes.

Effects	Estimate	SE	p value	ES/ICC
Fixed				
ASK				
Intercept	21.86	0.67	<.001	
Pretest	0.64	0.03	<.001	
Intervention	3.54	0.58	<.001	.40
Struggling reader	−3.15	0.56	<.001	−.42
Class % of struggling readers	−0.65	0.17	<.001	
Intervention × Struggling Reader	0.02	0.65	.96	
Intervention × Class % of Struggling Readers	−0.14	0.17	.41	
Struggling Readers × Class % of Struggling Readers	0.37	0.14	.01	
Intervention × Struggling Readers × Class % of Struggling Readers	−0.72	0.24	<.001	
MASK				
Intercept	9.48	0.28	<.001	
Pretest	0.53	0.05	<.001	
Intervention	0.87	0.41	.03	.17
Struggling reader	−0.38	1.10	.73	−.09
Class % of Struggling Readers	−0.78	0.12	<.001	
Intervention × Struggling Reader	−0.90	1.29	.48	
Intervention × Class % of Struggling Readers	0.14	0.12	.26	
Struggling Readers × Class % of Struggling Readers	0.48	0.16	<.001	
Intervention × Struggling Readers × Class % of Struggling Readers	−0.40	0.24	.09	
Gates-MacGinitie				
Intercept	93.99	0.64	<.001	
Pretest	0.59	0.03	<.001	
Intervention	−0.73	0.99	.46	−.05
Struggling reader	−0.23	0.93	.80	.02
Class % of Struggling Readers	−1.16	0.25	<.001	
Intervention × Struggling Reader	1.95	1.75	.27	
Intervention × Class % of Struggling Readers	−0.56	0.35	.11	
Struggling Readers × Class % of Struggling Readers	0.78	0.40	.05	
Intervention × Struggling Readers × Class % of Struggling Readers	−0.83	0.71	.24	
Random				
ASK				
Level 1: Students	26.30	1.48	<.001	.90
Level 2: Classes	1.13	0.53	.03	.04
Level 3: Teachers	1.68	0.50	<.001	.06
MASK				
Level 1: Students	10.05	0.55	<.001	.92
Level 2: Classes	0.84	0.56	.13	.08
Level 3: Teachers	0.00	0.32	1.00	.00
Gates-MacGinitie				
Level 1: Students	78.43	4.56	<.001	.89
Level 2: Classes	8.96	3.61	.01	.10
Level 3: Teachers	0.97	1.40	.49	.01

Note: ASK = *Assessment of Social Studies Knowledge*; ES = effect size, applies to fixed effects; *Gates-MacGinitie* = *Gates-MacGinitie Reading Comprehension subtest*; ICC = intraclass correlation coefficient, applies to random effects; MASK = *Modified Assessment of Social Studies Knowledge and Comprehension*.
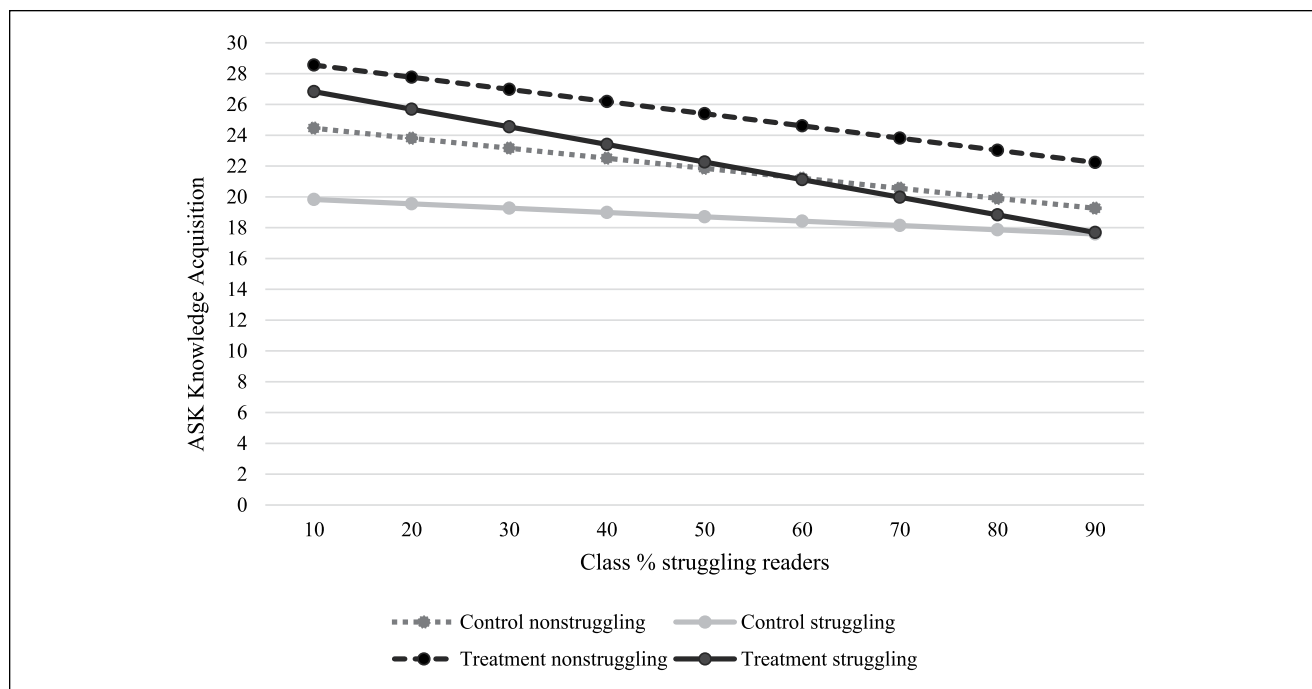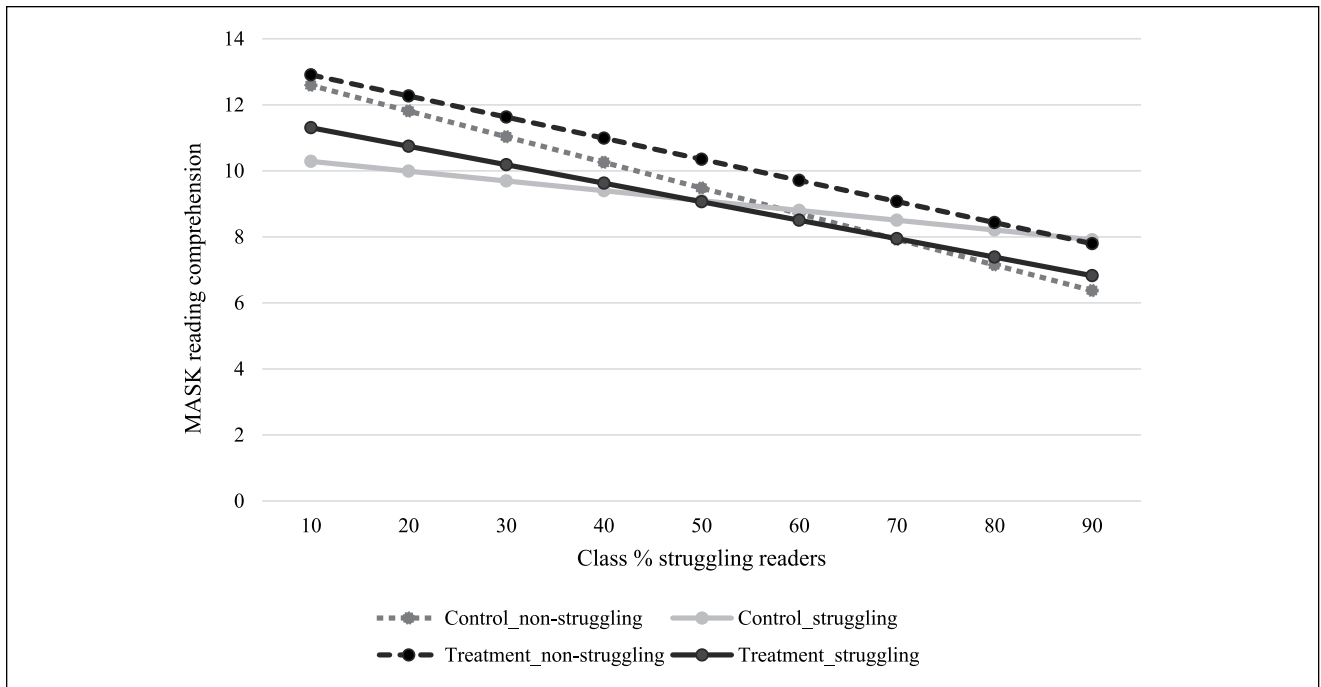
Figure 2. Predicted *Assessment of Social Studies Knowledge* scores for the four groups as a function of class proportion of struggling readers.

To identify areas along the continuum of classwide proportion of struggling readers where intervention status and struggling-reader status interact significantly, we used methods recommend by Bauer and Curran (2005). Outcomes on the ASK were bounded by classwide proportion values of .19 and .65, which indicates that differences between struggling and nonstruggling readers participating in the PACT treatment were significantly smaller than differences between struggling and nonstruggling readers in BaU classes for classrooms with less than .19 (or 19%) struggling readers, which is about three or four students, assuming an average of 17 students per class. For classrooms with a proportion of struggling readers between .19 to .65, differences in ASK outcomes for struggling and nonstruggling students in BaU and treatment classes did not differ from zero. In contrast, for classrooms with more than 65% struggling readers, differences between struggling and nonstruggling students in BaU were significantly smaller than differences between struggling and nonstruggling students in treatment classrooms. Put differently, PACT's effect for struggling readers compared to their nonstruggling classmates lessened as the class proportion of struggling readers exceeded .65. We plot predicted scores across values of class proportion of struggling readers in Figure 2.

MASK

The main treatment effect differed statistically from zero, indicating that nonstruggling students in treatment classrooms (i.e., .47 struggling readers) scored higher, on average, than nonstruggling students in comparison classes ($b = .87$, $p = .03$; ES = .17) when classwide proportion of struggling students was controlled. MASK scores for nonstruggling readers in comparison classes decreased by .78 points for every .10 increase in struggling readers ($b = -.78$, $p < .001$). The statistically significant interaction between struggling-reader status and classwide proportion of struggling readers ($b = .48$, $p < .001$) means that differences between struggling and nonstruggling readers in comparison classes decreased significantly as the proportion of struggling readers in the class increased. The regions of significance for the three-way interaction ($b = -.40$, $p = .09$) are bounded by values of .00 and .46 for class proportion of struggling readers, which means that the difference for struggling and nonstruggling students in comparison and treatment classes does not differ from zero for values between .00 and .45. Above .46, the difference between struggling and nonstruggling readers in treatment classes is significantly larger than differences between struggling and nonstruggling readers in comparison classes. See Figure 3 for plotted MASK values.

Figure 3. Predicted *Modified Assessment of Social Studies Knowledge and Comprehension* for the four groups as a function of class proportion of struggling readers.

Gates-MacGinitie

For the *Gates-MacGinitie*, only the main effect of classwide proportion of struggling readers ($b= -1.16$, $p< .001$) differed from zero. The trend is similar to that for the MASK comprehension outcome (i.e., negatively signed). Table 6 provides fixed and random effects for all reading outcomes.

Discussion

This study addressed the effect of the PACT intervention on comprehension and knowledge acquisition outcomes for struggling readers in classrooms that differed in the proportion of struggling readers. We hypothesized that PACT's effect would vary for struggling readers in classrooms that differed in proportion of struggling readers, with struggling students benefiting more from classroom compositions with a preponderance of nonstruggling students or classrooms with a smaller overall proportion of struggling readers. We interpret these findings as particularly impactful because there is complete overlap in the content taught to students in both conditions, with the only variation being the use of PACT instructional practices in the treatment condition and typical instructional practices in the BaU comparison condition. Furthermore, because randomization was at the class level, teacher effects were also controlled. Controlling teacher effects provides a strong challenge to the treatment increasing confidence in the positive findings for treated students. In addition, these findings align with those from prior

studies of PACT efficacy with general populations of students, which reported effect sizes of 0.17 (Vaughn et al., 2013) and 0.32 (Vaughn et al., 2015) on content knowledge acquisition and an effect size of 0.29 (Vaughn et al., 2013) on content-related reading comprehension. We did not hypothesize statistically significant gains on the distal reading comprehension measure (*Gates-MacGinitie Reading Comprehension subtest*); however, we administered the measure to ensure that participating in the PACT treatment did not negatively affect reading comprehension of target students.

In our recent study from which these data were drawn (Vaughn et al., 2017), we hypothesized that the benefit of PACT would vary depending on the class-level prevalence of ELs. We determined that increases in the class-level percentage of ELs were associated with lower content acquisition scores for both ELs and non-ELs. Our interpretation of the findings, requiring further research and testing, was that all students under such conditions had less access to and use of content-related academic language. Because the PACT treatment relied heavily on discourse-based practices, we interpreted the findings as an indication that the relative advantage for ELs in PACT diminishes as classrooms have an increased percentage of ELs. Trends for content knowledge acquisition across values of class percentage of ELs decline for all groups.

This study reports similar findings for students who are struggling readers. As the percentage of struggling readers

in a class increases, performance on the content knowledge acquisition measure decreases. This moderating effect for class percentage of struggling readers suggests that they may be more disadvantaged in classes with large numbers of struggling readers than when there are fewer struggling readers—at least for PACT-like interventions.

Although we do not know precisely why these findings prevail for ELs and struggling readers, it is possible to consider the components of PACT and their possible impact when class compositions vary. In addition to a discourse-based approach, PACT relies heavily on a text-based approach in which students work with their peers in a team-based process to use text as a resource for asking and answering content questions. This study investigated the extent to which the same moderating effect of prevalence of student group (i.e., struggling readers) would influence content knowledge and reading comprehension for participating students. We hypothesized that when the proportion of struggling readers in the class was high, treatment effects would diminish for all students, particularly the students with reading comprehension difficulties. Students with reading comprehension difficulties may have limited language, vocabulary, or background knowledge that influences student-to-student discourse (Shanahan & Shanahan, 2008). In addition, comprehension difficulties may limit the students' abilities to make key connections among newly learned concepts and other content information.

Implications

These findings have implications for educators. First, the findings confirm what many teachers and policy makers already know, which is that individual outcomes for many students are a complex product of multiple influences. In our data, class-level factors and student-level factors influence the acquisition of history content for both struggling and nonstruggling readers. At the student level, our models suggest that struggling readers achieve at lower levels in traditional instructional settings compared to their classmates, on average. At the class level, the proportion of struggling readers matters. However, these different influences are conditional; the effect of one factor depends on the presence of absence of another. For example, our results indicate that struggling readers benefit most from PACT-like interventions when implemented in classrooms where fewer than 20% of students are struggling readers, though they may not be disadvantaged in settings where up to 65% of students are struggling. That said, it is important to remember that this is a single study that represents a unique set of potential factors. It would be a mistake to generalize these findings to include alternative instructional approaches, for example. However, to the extent that a low-performing (recall that the

classwide mean proportion of struggling students is .47) school or district is implementing discourse- and text-based instructional models in middle school history classes, the findings represent a useful data point for deciding about classwide composition and the distribution of struggling readers.

Second, the findings provide a framework for educators tasked with making policy about how students are scheduled into content area classes (in history classes, specifically). Even in traditional settings, where teacher-focused instructional approaches are prevalent, struggling students appear to perform marginally better when the classwide proportion of struggling readers is in the lower quartile. Our data do not address the question, "How would typical readers have performed in typical classes without struggling readers?" so we do speculate on the advantages or disadvantages for this group. However, for struggling students, there appear to be clear benefits to a relatively equitable distribution of ability levels. This becomes increasingly the case when discourse- and text-based instructional models are implemented. Again, these data need to be considered with caution, particularly by educators working in more affluent or high-achieving schools or districts, a group that may represent a different population than the groups sampled in this study.

Finally, we think these findings raise important questions. We have reported findings similar to these for students who are EL (Vaughn et al., 2017), and we have speculated in several places about the possible causal mechanisms that underlie these effects. However, we do not have the data necessary to address this question directly. We hypothesize, generally, that the language demands inherent in a discourse or text-based intervention may have a role. However, direct empirical support is needed, whether in large-scale trials, like the one described here, or in more controlled studies where the relevant cognitive processes are more easily engaged and manipulated.

Limitations

This study was conducted in three school districts in the southeastern and southwestern United States, limiting generalization to these students and teachers in these regions. Although we collected information on the general types of instruction provided in both the PACT treatment and typical-instruction comparison classrooms, we cannot examine specific instructional supports that may have improved the learning of students in classrooms with high percentages of students with significant reading comprehension difficulties. Similarly, we do not have information on the supports, such as reading interventions, these students were provided throughout the rest of the school day.

Declaration of Conflicting Interests

Funding

References

Ahmed, Y., Francis, D. J., York, M., Fletcher, J. M., Barnes, M., & Kulesz, P. (2016). Validation of the direct and inferential mediation (DIME) model of reading comprehension in Grades 7 through 12. *Contemporary Educational Psychology*, *44*, 68–82.

Ammermueller, A., & Pischke, J. S. (2009). Peer effects in European primary schools: Evidence from the progress in international reading literacy study. *Journal of Labor Economics*, *27*(3), 315–348.

Angrist, J. D., & Lang, K. (2004). Does school integration generate peer effects? Evidence from Boston's Metco program. *American Economic Review*, *94*(5), 1613–1634. doi:10.1257/000282804305216

Attewell, P., Lavin, D., Domina, T., & Levey, T. (2006). New evidence on college remediation. *Journal of Higher Education*, *77*(5), 886–924.

August, D., Branum-Martin, L., Cardenas-Hagan, E., & Francis, D. J. (2009). The impact of an instructional intervention on the science and language learning of middle grade English language learners. *Journal of Research on Educational Effectiveness*, *2*(4), 345–376. doi:10.1080/19345740903217623

Baker, S., Lesaux, N., Jayanthi, M., Dimino, J., Proctor, C. P., Morris, J., . . . Newman-Gonchar, R. (2014). *Teaching academic content and literacy to English learners in elementary and middle school* (NCEE 2014-4012). Washington, DC: National Center for Education Evaluation and Regional Assistance. Retrieved from http://ies.ed.gov/ncee/wwc/publications_reviews.aspx

Bauer, D. J., & Curran, P. J. (2005). Probing interactions in fixed and multilevel regression: Inferential and graphical techniques. *Multivariate Behavioral Research*, *40*(3), 373–400. doi:10.1207/s15327906mbr4003_5

Callahan, R. M. (2005). Tracking and high school English learners: Limiting opportunity to learn. *American Educational Research Journal*, *42*(2), 305–328. doi:10.3102/00028312042002305

Catts, H. W., Compton, D., Tomblin, J. B., & Bridges, M. S. (2012). Prevalence and nature of late-emerging poor readers. *Journal of Educational Psychology*, *104*(1), 166–181.

Chen, X. (2016). *Remedial coursetaking at U.S. public 2- and 4-year institutions: Scope, experiences, and outcomes* (NCES 2016-405). Washington, DC: National Center for Education Statistics. Retrieved from http://nces.ed.gov/pubsearch

Dar, Y., & Resh, N. (1986). Classroom intellectual composition and academic achievement. *American Educational Research Journal*, *23*(3), 357–374.

Douglas, K. M., & Albro, E. R. (2014). The progress and promise of the Reading for Understanding research initiative. *Educational Psychology Review*, *26*, 341–355. doi:10.1007/s10648-014-9278-y

Dreeben, R., & Barr, R. (1988). The formation and instruction of ability groups. *American Journal of Education*, *97*(1), 34–64.

Executive Office of the President. (2015). *Every Student Succeeds Act: A progress report on elementary and secondary education*. Retrieved from https://obamawhitehouse.archives.gov/sites/obamawhitehouse.archives.gov/files/documents/ESSA_Progress_Report.pdf

Francis, D. J., Rivera, M., Lesaux, N., Kieffer, M., & Rivera, H. (2007). *Practical guidelines for the education of English language learners: Research-based recommendations for instruction and academic interventions*. Portsmouth, NH: RMC Research, Center on Instruction. Retrieved from http://www.naesp.org/resources/2/Research_Roundup/2007/RR2007v24n1a4.pdf

Good, T., & Marshall, S. (1984). Do students learn more in heterogeneous or homogeneous groups? In P. Peterson, L. Wilkinson, & M. Hallinan (Eds.), *Student diversity and the organization, process, and use of instructional groups in the classroom* (pp. 15–38). New York, NY: Academic Press.

Gottfried, M. A. (2014). Classmates with disabilities and students' noncognitive outcomes. *Educational Evaluation and Policy Analysis*, *36*(1), 20–43. doi:10.3102/0162373713493130

Hallinan, M. T. (1988). School composition and learning: A critique of the Dreeben-Barr model. *Sociology of Education*, *61*(3), 143–146.

Hanushek, E. A., Kain, J. F., Markman, J. M., & Rivkin, S. G. (2003). Does peer ability affect student achievement? *Journal of Applied Econometrics*, *18*(5), 527–544. doi:10.1002/jae.741

Kieffer, M. J. (2014). Morphological awareness and reading difficulties in adolescent Spanish-speaking language minority learners and their classmates. *Journal of Learning Disabilities*, *47*(1), 44–53.

Lesaux, N. K., Keiffer, M. J., Faller, S. E., & Kelley, J. G. (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. doi:10.1598/RRQ.45.2.3

Lesaux, N. K., Kieffer, M. J., Kelley, J. G., & Harris, J. R. (2014). Effects of academic vocabulary instruction for linguistically diverse adolescents: Evidence from a randomized field trial. *American Educational Research Journal*, *51*(6), 1159–1194. doi:10.3102/0002831214532165

MacGinitie, W. H., MacGinitie, R. K., Maria, K., Dreyer, L. G., & Hughes, K. E. (2006). *Gates–MacGinitie reading test* (4th ed.). Rolling Meadows, IL: Riverside.

Mancilla-Martinez, J., Kieffer, M. J., Biancarosa, G., Christodoulou, J. A., & Snow, C. E. (2011). Investigating English reading comprehension growth in adolescent language minority learners: Some insights from the simple view. *Reading and Writing*, *24*, 339–354. doi:10.1007/s11145-009-9215-5

Markow, D., & Cooper, M. (2008). *The MetLife survey of the American teacher: Past, present, and future*. Retrieved from http://files.eric.ed.gov/fulltext/ED504457.pdf

Michaelson, L. K., & Sweet, M. (2011). Team-based learning. *New Directions for Teaching and Learning*, *128*, 41–51. doi:10.1002/tl.467

Muthén, L. K., & Muthén, B. O. (2010). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

National Center for Education Statistics. (2015). *National assessment of educational progress*. Retrieved from https://nces.ed.gov/nationsreportcard/

Noble, J. P., Roberts, W. L., & Sawyer, R. L. (2006). *Student achievement, behavior, perceptions, and other factors affecting ACT scores*. Retrieved from http://files.eric.ed.gov/fulltext/ED510477.pdf

Pallas, A. M., Entwisle, D. R., Alexander, K. L., & Stluka, M. F. (1994). Ability-group effects: Instructional, social, or institutional? *Sociology of Education*, *67*(1), 27–46.

Scammacca, N. K., Roberts, G. J., Cho, E., Williams, K. J., Roberts, G., Vaughn, S. R., & Carroll, M. (2016). A century of progress: Reading interventions for students in Grades 4–12, 1914–2014. *Review of Educational Research*, *86*(3), 756–800.

Shanahan, T., & Shanahan, C. (2008). Teaching disciplinary literacy to adolescents: Rethinking content-area literacy. *Harvard Educational Review*, *78*(1), 40–59.

Swanson, E., Wanzek, J., Vaughn, S., Roberts, G., & Fall, A.-M. (2015). Improving reading comprehension and social studies knowledge among middle school students with disabilities. *Exceptional Children*, *81*(4), 426–442. doi:10.1177/0014402914563704

Vaughn, S., Martinez, L. R., Wanzek, J., Roberts, G., Swanson, E., & Fall, A.-M. (2017). Improving content knowledge and comprehension for English language learners: Findings from a randomized control trial. *Journal of Educational Psychology*, *109*(1), 22–34. doi:10.1037/edu0000069

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

Vaughn, S., Swanson, E. A., Roberts, G., Wanzek, J., Stillman-Spisak, S. J., Solis, M., & Simmons, D. (2013). Improving reading comprehension and social studies knowledge in middle school. *Reading Research Quarterly*, *48*(1), 77–93. doi:10.1002/rrq.039

Wanzek, J., Kent, S., Vaughn, S., Swanson, E. A., Roberts, G., & Haynes, M. (in press). Implementing team-based learning in middle school social studies classes. *Journal of Educational Research*.