Resubmitting to ERIC to acknowledge funding.

Improving Content Knowledge and Comprehension for English Language Learners: Findings From a Randomized Control Trial

Sharon Vaughn and Leticia R. Martinez
The University of Texas at Austin

Jeanne Wanzek
Florida State University

Greg Roberts, Elizabeth Swanson, and Anna-Mária Fall
The University of Texas at Austin

Supporting the reading comprehension and content knowledge acquisition of English language learners (ELs) requires instructional practices that continue beyond developing the foundational skills of reading. In particular, the challenges ELs face highlight the importance of teaching reading comprehension practices in the middle grades through content acquisition. We conducted a randomized control trial to examine the efficacy of a content acquisition and reading comprehension intervention implemented in eighth-grade social studies classrooms with English language learners. Using a within-teacher design, in which 18 eighth-grade teachers' social studies classes were randomly assigned to treatment or comparison conditions. Teachers taught the same instructional content to treatment and comparison classes, but the treatment classes used instructional practices that included comprehension canopy, essential words, knowledge acquisition, and team-based learning. Students in the treatment group ($n = 845$) outperformed students in the comparison group ($n = 784$) on measures of content knowledge acquisition and content reading comprehension but not general reading comprehension. Both ELs and non-ELs who received the treatment outperformed those assigned to the BAU comparison condition on measures of content knowledge acquisition (ES = 0.40) and content-related reading comprehension (ES = 0.20). In addition, the proportion of English language learners in classes moderated outcomes for content knowledge acquisition.

Keywords: English learners, middle school, reading comprehension, social studies

Supplemental materials: http://dx.doi.org/10.1037/edu0000069.supp

Designing and implementing effective instruction for the growing number of English language learners (ELs) in public schools is a significant educational challenge. Approximately 20% of students in the United States are children of immigrant parents (Fix & Passel, 2003), and although they are not all ELs, many are, with approximately 10.5% of all U.S. students identified as ELs (National Clearinghouse for English Language Acquisition, 2006). The expectation is that within the next 10 to 15 years, as many as one in four children enrolled in schools in the United States will be ELs. Although ELs have many assets, such as linguistic and cultural diversity, they commonly face educational challenges, including low achievement across reading, writing, history, mathematics, and other academic areas (Lesaux, Kieffer, Kelley, & Harris, 2014; Snow & Uccelli, 2009), and are at a disproportionately high risk for school dropout (Hernandez, 2012; Kennelly & Monrad, 2007). Not all language-minority students have the same trajectory for school success. Students who begin kindergarten with proficiency in English have academic trajectories similar to non-ELs, whereas students who enter school with limited English proficiency do not fare as well, demonstrating weaker learning trajectories that are quite divergent from their non-EL peers by the end of elementary school (Kieffer, 2008).

Despite the growing number of ELs in schools and the increased attention to improving their academic opportunities, ELs continue to demonstrate difficulties beyond the elementary grades. In particular, they demonstrate difficulty in literacy, with only 26% of ELs in the eighth grade scoring above "basic" on reading achievement tests (National Center for Education Statistics, 2013). Of further concern, these data have not changed significantly in any state in the previous 10 years. Scores for ELs are 35 to 40 scale score points below students who are not ELs (National Center for Education Statistics, 2013). A

similar pattern is observed on the fourth-grade reading test from the National Assessment of Educational Progress.

ELs frequently score lower on achievement tests in part because of their challenges in developing background knowledge and vocabulary in English (National Center for Education Statistics, 2009). ELs also have the dual task of concurrently learning English and content, increasing the likelihood that the rate they learn to read and understand English will influence their content knowledge. Thus, low scores in reading do not bode favorably for middle-school ELs in content classrooms, in which reading for understanding is an integral part of success. As ELs move from elementary to middle school, the demands for sophisticated language, literacy, and background knowledge increase, requiring teachers to access effective instructional practices that are beneficial to a range of learners, including ELs.

Lack of Opportunity

In addition to the heightened challenges of learning content in a language they are simultaneously learning to read and understand, ELs may have restricted opportunities to learn based on their lack of access to high-quality teachers, proficient student learners, and curricula. For example, Callahan (2005) reported that "tracking" ELs played a significant role in their learning achievement. Her analysis revealed that ELs were primarily clustered in classes that were not college preparatory. To the extent that opportunity to learn content and academic vocabulary 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—teachers of classes with high concentrations of ELs may provide fewer opportunities for rich language discourse and content learning.

Importance of Enhancing Reading Comprehension of ELs in Middle Grades

The previously discussed data, as well as the findings from two practice guides (Baker et al., 2014; Francis, Rivera, Lesaux, Kieffer, & Rivera, 2006) that recommend content area instruction as a focus for learning new concepts and knowledge, underscore the importance of teaching academic content and literacy to ELs in the middle grades. The urgent need to improve instructional practices for ELs in the middle grades is demonstrated in these students' slow development of reading comprehension (Mancilla-Martinez, Kieffer, Biancarosa, Christodoulou, & Snow, 2011). Further, substantial numbers of ELs demonstrate "late-emerging" reading difficulties, or reading problems that emerge after fourth grade (Kieffer, 2010). This finding suggests that ELs can often master the foundational skills of word reading with adequate fluency, but that as the syntax, vocabulary, and background knowledge of texts become more complex, ELs' reading difficulties manifest. At each developmental period, as determined by grade level (Grades 3, 5, and 8), ELs were found to be at substantially greater risk than native English speakers for reading difficulties that were not recognized prior to Grade 3 (Kieffer, 2010). Thus, supporting the reading comprehension of ELs requires instructional practices that continue beyond developing the foundational skills of reading. In particular, the challenges ELs face highlight the importance of teaching reading comprehension practices in the middle grades.

Purpose of the Study

Considering the high need for effective instructional practices that enhance both reading comprehension and knowledge acquisition for ELs, we modified Promoting Adolescents' Comprehension of Text (PACT)—a previously developed package of instructional practices—by interweaving features of instruction associated with improved outcomes for ELs (e.g., additional focus on academic vocabulary and peer discourse). We selected PACT for several reasons. First, PACT has demonstrated efficacy through previous randomized control trials with eighth-grade students and students with disabilities (Swanson, Wanzek, Vaughn, Roberts, & Fall, 2015; Vaughn et al., 2013, 2015; Wanzek et al., 2015). In previous studies examining overall effects for all learners, PACT was associated with improved outcomes in reading comprehension (Vaughn et al., 2013), content acquisition and vocabulary knowledge (Vaughn et al., 2013, 2015), and sustained content and vocabulary knowledge at multiple points through follow-up measures (Vaughn et al., 2015). Second, research suggests that ELs and below-grade-level readers exhibit many of the same learning challenges in the middle grades and that similar instruction may be necessary for both groups (Lesaux & Kieffer, 2010). Thus, many of the instructional practices of PACT held promise for ELs, particularly if instructional enhancements were added. Third, the platform of PACT instructional practices is well aligned with best practices for teaching ELs, leading us to hypothesize that with appropriate modifications (described in the next section), PACT would yield positive outcomes for ELs.

This study represents a randomized control trial of the efficacy of a PACT treatment modified for ELs with eighth-grade students in schools with moderate to high concentrations of ELs, ensuring that ELs would be included in all participating classes. We hypothesized that students who were not ELs would perform similarly to students in previous studies (Vaughn et al., 2013, 2015), with the treatment students demonstrating statistically significantly higher scores than comparison students on content knowledge acquisition and content-related reading comprehension. We also hypothesized that ELs in the treatment condition would outperform ELs in the comparison condition on content acquisition and content reading comprehension. Thus, we hypothesized that the modified version of PACT would have a universally positive effect on all learners. We further hypothesized that participants in the treatment condition would not outperform participants in the comparison condition on the distal measure of reading comprehension. Finally, we acknowledge the important influence of classmates on a given student's outcomes. The considerable 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 language-related peer variables, certainly not in the context of a discourse-based intervention designed to improve reading comprehension and content knowledge. We hypothesized that PACT's effect would depend in part on class levels of English academic language proficiency, which we operationalized as the proportion of ELs (or non-ELs) in the classroom. We expected that the proportion of ELs in a given class would moderate PACT's effect, with increasing class-level prevalence of ELs corresponding to diminishing treatment effects for all students—particularly for ELs.

Features of Effective Instructional Practices for ELs

The set of instructional practices tested in this randomized control trial can be woven into content area instruction (i.e., social studies) to enhance content learning and comprehension for all learners, with a specific focus on ELs. 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 et al., 2006), to enhance the features of instruction and promote best practice for teaching ELs. Although many of these practices were already part of the foundation of PACT, we enhanced the focus on academic vocabulary by teaching theme-related vocabulary words across time and activity, integrating oral and written instruction into content learning, and using both paired learning and team-based learning (TBL; Michaelsen & Sweet, 2011) to provide peer interaction and extended practice with feedback (Vaughn et al., 2013, 2015; Wanzek et al., 2015; see sample lessons in the online supplemental materials).

Reviews of the research on effective instruction for ELs (August & Shanahan, 2006; Baker et al., 2014; Francis et al., 2006) emphasize the importance of addressing academic language by providing direct and systematic instruction of the English language while teaching content across the disciplines. 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 engaging students in reading, speaking, and writing activities, in which students applied the meaning of the words in multiple and meaningful contexts. Academic vocabulary teaching was enhanced in the modified version of PACT by providing instruction on more abstract terms that students need to communicate across the disciplines and that are needed for school tests and tasks—for example, academic expressions for comparing and contrasting and using cause and effect (Lesaux, Kieffer, Faller, & Kelley, 2010).

In addition to the TBL in the original versions of PACT, we provided structured opportunities for ELs to participate in academic discussions and writing that supported the use of learned content vocabulary (August, Branum-Martin, Cardenas-Hagan, & Francis, 2009; Lesaux et al., 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.

Based on intervention studies with ELs, we incorporated additional features of instruction associated with improved outcomes for ELs (August et al., 2009; Vaughn et al., 2009). Students worked in pairs or small groups during most PACT components to prepare for discussing and writing responses to inferential questions and summaries that 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, one of the principles of the TBL comprehension checks and knowledge application activities was continuous, targeted feedback, in which teachers affirmed or corrected students' understanding of the content.

Method

Research Design

We conducted a randomized control trial to test the efficacy of a modified version of the PACT reading comprehension and content acquisition intervention in eighth-grade social studies classes. Participants included English-speaking students, ELs, and former ELs. To be selected, schools (and their districts) had to serve high numbers of ELs—and each class had to have at least one identified EL. In the selected schools, all eighth-grade social studies teachers participated, and their class sections were randomly assigned to the treatment or comparison condition. Each teacher taught both PACT treatment classes and comparison classes, and the same social studies content was delivered to students in both conditions, albeit using the interrelated components of PACT in treatment classes only.

Setting and Participants

School sites. 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 of the schools were in the southwestern United States—two in a large, diverse urban district, and another in a smaller, predominantly Hispanic suburban district. Four of the schools were in one district in the southeastern United States—about 40% of the students in these schools were Hispanic. The proportion of students identified as ELs in the schools ranged from 15.4% to 44.5%. Although we recruited school districts that served the highest numbers of ELs in our surrounding areas, districts ultimately dictated which schools could participate in the study. Furthermore, principals ultimately decided whether to participate, which resulted in the wide range of EL proportions across schools. We report district-identified EL classification, but the criteria that state departments of education use may differ. In the five schools for which such data were available, the proportion of students who qualified for free or reduced-price meals ranged from 48.8% to 82.6%. Additional school-level demographic information is displayed in Table 1.

Teachers. The 18 teacher participants (nine women and nine men) 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 of the teachers had a bachelor's degree and five had a master's degree. Their teaching experience ranged from less than 1 year to 34 years ($M = 10.13$, $SD = 10.48$). Teachers' ethnicities included 83.3% White, 16.7% Hispanic, and 5.6% Asian.

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 (845 students) and 45 comparison (784 students) classes. When teachers had an odd number of classes, randomization assigned extra classes to treatment. Of the participants, 26.7% were current ELs or held an EL status within the last 2 years. Students' EL designation was determined in part by district language-proficiency tests. Similar to other content area studies with secondary ELs (Snow, Lawrence, & White, 2009; Vaughn et al., 2009), recently exited EL students were included in the EL sample because they require continued

Table 1
School-Level Demographics for All Participants

School descriptives	School A District 1	School B District 1	School C District 2	School D District 3	School E District 3	School F District 3	School G District 3
Gender							
Male	43.2%	42.0%	43%	45.2%	50.6%	44.9%	52.2%
Female	52.7%	50.7%	41.8%	54.2%	47.0%	52.9%	47.8%
Missing	4.1%	7.2%	15.2%	.6%	2.4%	2.2%	.0%
Race							
Hispanic	45.42%	44.72%	85.81%	40.82%	36.9%	38.11%	26.07%
African American	7.25%	7.32%	8.21%	9.55%	10.85%	11.23%	5.39%
Caucasian	33.21%	27.63%	4.48%	38.01%	41.22%	36.56%	60.0%
American Indian	12.98%	17.07%	.75%	9.53%	8.28%	10.57%	6.74%
Asian	.76%	1.63%	.75%	1.12%	2.56%	2.87%	1.57%
Pacific Islander	.38%	1.63%	.0%	.94%	.99%	.66%	.22%
Special education	14.4%	20.3%	12.0%	8.4%	10.7%	8.3%	6.8%
Home language							
English	20.5%	52.2%	53.2%	36.8%	43.3%	29.7%	65.4%
Spanish	74.0%	39.1%	30.4%	57.1%	47.0%	58.3%	29.3%
Other	.7%	1.4%	1.3%	3.2%	5.2%	6.9%	3.1%
Missing	4.8%	7.2%	15.2%	2.9%	4.6%	5.1%	2.21%
English language learner	44.5%	33.3%	15.8%	27.4%	25.6%	37.3%	15.4%
Free or reduced-price meals	n/a	n/a	74.1%	82.6%	68.9%	67.8%	48.8%
Participating students							
n	146	69	158	310	328	276	324
%	9%	4.2%	9.7%	19%	20.1%	16.9%	19.9%
Participating teachers							
n	2	2	2	3	3	3	3
%	11.12%	11.12%	11.12%	16.16%	16.16%	16.16%	16.16%

Note. n/a = not available.

academic language support while they face increasingly demanding academic tasks in mainstream middle-school classrooms. Furthermore, students who were exited from EL status in middle school were included in the EL sample when the variability and subjectivity of criteria used to reclassify ELs as English proficient was considered (Office of Planning, Evaluation and Policy Development, Policy & Program Studies Service, 2012). Additionally, on a student survey administered at pretest, 50.5% of the students reported that a language other than English (mostly Spanish) was spoken at home. Additional student-level demographic information is displayed in Table 2. Note that many students identified themselves as multiracial; therefore, the number of participants represented across racial categories adds up to more than 1,629 (the total number of participants).

Differential attrition. Of the 1,629 participants, 224 did not have a posttest score on the Assessment of Social Studies Knowledge (ASK; all student measures are described in detail later in the article), yielding an overall attrition rate of 14.1% and a differential attrition rate of 1.6%. On the Modified Assessment of Social Studies Knowledge and Reading Comprehension (MASK), 305 students did not have a posttest score, yielding an overall attrition rate of 19.4% and a differential attrition rate of .2%. For the Gates–MacGinitie Reading Comprehension Subtest (4th ed.; MacGinitie, MacGinitie, Maria, Dreyer, & Hughes, 2006), 327 students did not have a posttest score, yielding an overall attrition rate of 20.1% and a differential attrition rate of .4%. To establish whether differential attrition was evident across the groups, a two-way analysis of variance was conducted on the primary outcome variables. The factors in the analysis were treatment condition, completer status at posttest, and the interaction of condition

and completer status. A significant interaction signifies systematic group differences in the characteristics of students who remained in the study. Data revealed no significant condition by completer status interaction effect (p values ranged from .13 to .61). These findings indicate that attrition among groups was unlikely to bias the observed effects of the intervention.

Table 2
Student-Level Demographics by Group

Student descriptives	Comparison		Treatment	
	n	%	n	%
Gender				
Male	333	42.5	426	50.4
Female	416	53.1	385	45.6
Missing	35	4.4	34	4.0
Race				
Caucasian	490	41.11	482	37.73
African American	115	9.65	105	8.20
Hispanic	467	39.18	523	40.92
Asian	14	1.17	30	2.35
American Indian	98	8.22	129	10.10
Pacific Islander	8	.67	9	.70
Special education	65	8.3	95	11.2
Home language				
English	345	44	355	42.0
Spanish	363	46.3	407	48.2
Other	23	2.9	37	4.4
Missing	53	6.8	46	5.4
English language learner	190	24.2	245	29

Intervention Procedures

For approximately 20 weeks, the PACT intervention was delivered in treatment classes during regularly scheduled social studies time. During the first 6 to 8 weeks, three consecutive units were taught in classes that were either approximately 45 min daily or took place every other day for 90 min. For the next 12 weeks, teachers implemented only one of the PACT components, the knowledge acquisition through text reading routine described in Figure 1, three times a week for approximately 15 min per session. Students in comparison sections received instruction on the same content over the same amount of time as students in the PACT treatment classes, but delivery of the content differed, as teachers took a business-as-usual (BAU) approach in the comparison classes.

Description of the Treatment Intervention

The PACT intervention aligned with participating districts' standards and the Common Core Standards (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010). Using the set of instructional practices from the previous PACT intervention studies (e.g., Vaughn et al., 2013), we integrated enhancements for ELs (see the earlier section Features of Effective Instructional Practices for ELs). The treatment features three units, blending five components that work together and complement one another. Following is a description of the five PACT components.

Comprehension canopy. The comprehension canopy starts and guides every unit. It is a 10- to 15-min routine to engage students in a purpose for reading while integrating new content to build students' background knowledge. Teachers initiate a concise introduction and then show a brief video clip that provides students with requisite background information before encountering the unit material. After students discuss video-related questions with a partner and share with the class, the teacher introduces an overarching comprehension question that is revisited and extended throughout the unit. Each comprehension canopy question is designed to develop students' academic language in social studies by focusing on compare and contrast, cause and effect, or perspective taking. For example, students are asked, "How did the colonial regions develop differently?"

Essential words. Following the comprehension canopy, five key words are introduced on the first day of the unit. The purpose of the essential words component is to teach the meaning of concepts that are tightly connected with the content and to support new learning by having students engage with the words in multiple contexts. Each word is taught by using a student-friendly definition, visual representation, related words, examples and nonexamples, and a turn-and-talk prompt that asks students to discuss in pairs an activity related to the essential word(s). As students move through the unit, they have recurrent exposure to each word in warm-up activities, reading of varied texts, and team-based activities. Students are afforded multiple opportunities to apply the meaning of the essential words and to use them orally and in writing. For example, during a warm-up activity, students are required to revisit an illustration of the essential word *mercantilism* and to think about how it makes some people wealthy. Students then have to answer questions about who benefits the least from mercantilism and whether it is fair.

Knowledge acquisition through text reading. Three times a week, teachers lead students through a critical reading routine that lasts approximately 15 min and requires students to read informational text related to the topic. Teachers guide the process by providing a brief introduction, sharing a video clip or a geographical map to set up the context for the content to be read. During this introduction, the teacher also reinforces the essential words that students will encounter and connects the reading to the comprehension canopy question. Students read the text as a whole class with the teacher, in pairs, in small groups, or independently. Additionally, students address a variety of content- and inference-based questions verbally and in writing intermittently throughout the reading.

TBL comprehension check. TBL is based on a university-level practice adapted for use in middle-school classrooms to provide opportunities for text-based discussions and justifications for ideas (Michaelsen & Sweet, 2011). Two key elements were integrated into the TBL comprehension checks: (a) heterogeneous teams of students, and (b) a process in which students work individually and with a team to ensure accountability for learning and understanding the content.

Twice during each unit, teachers administer a short comprehension check to examine students' understanding of unit content and to inform further instruction. This check has 10 comprehension questions, with five focusing on vocabulary. First, students individually complete the comprehension check and turn it in to the teacher. The teacher monitors individual students' comprehension of content through this initial comprehension check. Next, students complete the same comprehension check as a two-person team, but using their texts and notes during this second round to justify their answers. Students use scratch-off cards to mark their answers and receive immediate feedback on accuracy. A correct answer reveals a star. If the team answer is incorrect, the team revisits notes and text, discusses, and selects an alternative answer supported with text evidence. Finally, the teacher provides whole-class targeted instruction to address gaps in students' understanding.

TBL knowledge application. Knowledge application in PACT requires teams to apply the newly learned content of the unit through a problem-solving activity—for example, addressing a question such as, "What might have happened to prevent the Revolutionary War?" At the conclusion of every unit, teams of four students participate in a discussion that is facilitated by sharing ideas and using text evidence to address the task assigned. Students must listen to team members' contributions and think critically before presenting a response to the class. The teacher monitors progress while students work in their teams, provides feedback to teams as they demonstrate their understanding of the content, and facilitates students' extended thinking about the content and evidence. At the end of the activity, student teams share their responses and reasons with the class. Moreover, the teacher ends the activity by connecting the knowledge application work to the comprehension canopy question that started the unit. The teacher synthesizes key information learned over the entire unit and prepares the class for an end of unit assessment.

Implementation Support

The research team provided professional development to participating teachers during two sessions and provided ongoing in-class

support in treatment classes. Prior to launching the intervention, a 1-day professional development workshop trained teachers to implement the intervention in treatment classes and stressed the significance of using the PACT components in treatment classes exclusively. The training day was devoted to (a) providing an overview and explaining the design of the study, (b) discussing relevant research in reading comprehension and content teaching and learning, (c) explaining and modeling each PACT component, and (d) allowing hands-on practice with Unit 1 lessons and materials. (The Meadows Center for Preventing Educational Risk, 2013). An additional 3-hr professional development session was held for teachers after the completion of the first unit. During this second session, research support personnel reviewed the elements of the PACT components and discussed areas for improving PACT implementation.

Each teacher was assigned one research support person, who provided in-class support as needed throughout the units for the first 6 to 8 weeks. Research personnel filled various roles to ensure high levels of PACT implementation. During implementation of the first unit, research personnel were present in treatment classes daily and modeled the first occurrence of each PACT component (e.g., comprehension canopy, essential words). Research personnel scaled back their presence in treatment classes to two to three times a week during the second unit, and to one to two times a week during the third unit. They made more visits to teachers who required further coaching and feedback. Personnel also answered teachers' questions regarding PACT and assisted them in integrating the PACT components into their instructional planning for their treatment classes. During the 12 weeks that teachers continued to implement the knowledge acquisition through text reading routine, research personnel visited with teachers three to four times to keep track of the readings.

Implementation Fidelity

Fidelity data were collected by means of audio recordings in all participating teachers' classes throughout the implementation of the PACT intervention to measure adherence to the PACT components in treatment classes and determine whether any components were present in comparison classes. Before the intervention started, a database manager on the research team randomly selected one treatment and one BAU class period to be recorded per teacher. Next, each teacher audio recorded the randomly selected class periods daily for the duration of the three 10-day units of intervention in the identified classes. Each teacher submitted approximately 30 treatment audio recordings from one class period and 30 BAU audio recordings from another to be coded by the research team. Research personnel then coded two units (about 18 recordings) of instruction for each condition per teacher. Figure 1 provides the number of opportunities to observe each PACT component. For example, in two units, the comprehension canopy is observed two times per teacher; totaling across 18 teachers, this provides 36 possible opportunities. Other total possible opportunities are as follows: warm-up = 144; TBL comprehension check = 72; essential words = 36; knowledge acquisition through text reading = 108; and TBL knowledge application = 36.

The fidelity measure used to code the audio recordings aligned with the PACT components: (a) comprehension canopy, (b) essential words, (c) warm-up, (d) knowledge acquisition through text reading, (e) TBL comprehension check, and (f) TBL knowledge application. Coders rated the extent to which an individual teacher implemented required elements for each component, using a scale from 1 to 4, with 4 indicating that the teacher completed all of the expected aspects of the component. If the component was not required or expected for the day, a not applicable (0) rating was assigned.

As in prior PACT studies, interrater reliability on the fidelity instrument was established by using a gold-standard method (Gwet, 2012). A senior member of the research team held a 3-hr training session on the use of the fidelity instrument for a team of six research support personnel. The team examined the codebook and coding form, reviewed indicators of teacher behaviors associated with each PACT component, discussed comparison class coding, and practiced coding with videos and audio recordings. Two senior researchers then served as the gold standard and coded a set of treatment and comparison audio recordings. The coders individually coded the same audio-recorded lessons, using the fidelity instrument, and additional audio recordings were coded until interrater agreement of 90% or higher was reached. To avoid observer drift, the coding team reestablished reliability coding with two additional interrater checkpoints, using the same interrater agreement of 90% or higher. Although simple percent agreement is popular, it can be inflated because of chance (Hintze, 2005). Cohen's kappa (κ) is a more conservative measure of interrater agreement in that it takes into account chance agreement (Landis & Koch, 1977; Suen & Ary, 1989). Coefficients can range from -1.0 to 1.0. The interrater reliability for the raters was found to be $\kappa = 0.87$ ($p < .001$), 95% confidence interval [0.63, 1.0], which is considered "substantial" (Gelfand & Hartmann, 1975; Landis & Koch, 1977).

Table 3 presents fidelity data for each of the components in treatment and comparison classes. In treatment classes, PACT components were implemented with generally high levels of fidelity. Teachers struggled the most with implementing the knowledge acquisition through text reading component, with about 25% of the component implemented with a low or mid-low rating. Nonetheless, overall, teachers delivered the PACT intervention with consistently high levels of procedural fidelity.

Coders rated audio recordings of BAU instruction in comparison classes by using the same protocol used for treatment audio recordings to determine whether there was contamination of the BAU comparison condition. As displayed in Table 3, differential instruction for treatment and control students with respect to PACT was accomplished. Research support personnel frequently stressed the importance of avoiding PACT spillover into BAU class sections with participating teachers, and coders identified limited evidence of PACT in the BAU sections. Elements of the warm-up, knowledge acquisition, and essential words routines appeared in some BAU audio recordings, but at low rates compared with treatment classes. For example, in occurrences of knowledge acquisition through reading in comparison classes, a teacher might assign independent reading without some of the PACT reading routine elements, like providing context for the reading through video and making connections to essential unit vocabulary. Similarly, warm-up activities observed in BAU recordings rarely connected to and reinforced the unit content like PACT warm-up activities did. However, the PACT intervention was also made up of less common components, such as

10-Day Lesson Cycle

Figure 1. Frequency and duration of each PACT component during one unit. The PACT instructional components are embedded into the teachers' content instruction. PACT = Promoting Adolescents' Comprehension of Text; TBL = Team Based Learning; ELs = English Learners.

TBL comprehension check, which resembles a quiz. This component was rarely detected in BAU audio recordings because some of the elements exclusive to PACT, such as students working with a team to justify their answers and receiving immediate feedback from scratch-off forms, were almost never observed in BAU classes. This trend aligns with findings from prior PACT studies (Vaughn et al., 2013, 2015), which also reported intervention components observed to a minimal degree in comparison classrooms.

Student Measures

The same measures of impact employed in the prior two PACT studies (Vaughn et al., 2013, 2015) were used. 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 (4th ed.; 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.

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 that composed the intervention (Colonial America, Road to Revolution, and 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 amount of 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;

Table 3

Frequency of Fidelity Observations in Treatment and Comparison Classrooms

Fidelity rating scale	CC (N = 31)	%	WU (N = 130)	%	TBLC (N = 64)	%	EW (N = 32)	%	KA (N = 108)	%	TBLK (N = 34)	%
Treatment classes												
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												

	CC (N = 274)	%	WU (N = 274)	%	TBLC (N = 274)	%	EW (N = 274)	%	KA (N = 274)	%	TBLK (N = 274)	%
Comparison classes												
4 = high			26	9.5			2	.7				
3 = mid-high							2	.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; WU = warm-up; TBLC = team-based learning comprehension check; EW = essential words; KA = knowledge acquisition through text reading; TBLK = team-based learning knowledge application.

Wanzek et al., 2015). The Cronbach's alpha with the current sample was .93.

MASK and reading comprehension. 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 1,090 to 1,140, and the word count range was 312 to 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 was administered at pretest and posttest. The alpha coefficient with the current sample was .92.

Results

We fit a series of three-level regression models to estimate parameters and evaluate differences. Students were nested in classes, and classes were nested in teachers. Classes were randomized to condition within teachers. The following is the reduced form equation for the model:

$$\text{Reading Outcome}_{ijk} = \gamma_{000} + \gamma_{010} * (\text{Intervention})_{jk} + \gamma_{020} * (\text{Class \% of ELs})_{jk}$$
$$+ \gamma_{030} * (\text{Intervention} * \text{Class \% of ELs})_{jk}$$
$$+ \gamma_{100} * (\text{Pretest-gm})_{ijk} + \gamma_{200} * (\text{EL})_{ijk}$$
$$+ \gamma_{300} * (\text{Intervention} * \text{EL})_{ijk}$$
$$+ \gamma_{400} * (\text{EL} * \text{Class \% of ELs})_{ijk}$$
$$+ \gamma_{500} * (\text{Intervention} * \text{EL} * \text{Class \% of ELs})_{ijk} + r_{0jk} + u_{00k}$$
$$+ u_{01k}(\text{Intervention})_{jk} + e_{ijk}$$

Reading Outcome$_{ijk}$ represents the posttest score for student i in class j taught by teacher t. In similar, previous analyses (Vaughn et al., 2013; Vaughn et al., 2015) with MASK data and with data

from the related ASK, we included students' item-level responses in the measurement part of a structural equation model, estimating latent outcome scores in a one-parameter item-response model. Given the complexity of the models in this article, we use raw data for the ASK and MASK. Sensitivity analyses that use these earlier data suggest no differences in the direction of findings when using raw versus latent scores. However, differences in precision across the two approaches may be present, with the latent scoring model being more precise. For the Gates–MacGinitie test, we used standard scores. We dummy-coded *Intervention* and *EL*, with the comparison condition and the *non-EL* group at 0; treatment condition and *EL* are coded as 1. We model *Class % of ELs* as the percentage of ELs in each class. Its distribution is relatively normal, though skewed to the right ($M = 28.5$, $SD = 19.5$; skewness = .90). We centered these data on the 10% threshold. We did not use the variable's natural scale because zero is outside the logical range of the moderator (Bauer & Curran, 2005). We also did not transform the data because the variable has substantive meaning in its natural form and because *Class % of ELs* is the focal moderator and the basis for interpreting the results. Finally, we also did not center on the moderator's mean ($M = 28.5$) because the data are somewhat skewed. Instead, we centered *Class % of ELs* at about one standard deviation below its mean (Francis & Vaughn, 2009). We used this value (10% of ELs in the class) to interpret main effects and interaction effects and to calculate regions of statistical significance along the continuum of values for *Class % of ELs*.

Pretest scores on the ASK, MASK, and Gates–MacGinitie are included as grand-mean centered covariates. *Intervention* and *Class % of ELs* are modeled at Level 2 (class level), as is the two-way interaction involving *Intervention* and *Class % of ELs*. Other two-way interactions and the three-way interaction (all involving *EL*) are modeled at Level 1 (student level). Class-level residuals are modeled as r_{0jk}, and student-level residuals are modeled as e_{ijk}. Classes are randomized to condition within teachers, and $u_{00k} + u_{01k}$ (Intervention)$_{jk}$ represents class-level residuals within teachers for the treatment and comparison groups. Effect

sizes for *Intervention* and *EL* status are estimated as the ratio of the model-derived coefficient for *Intervention* (or *EL* status) and the pooled within-group standard deviation across conditions (or *EL* status) at posttest.

Table 4 summarizes pretest and posttest means, standard deviations, and observed score ranges for the *Reading Outcomes* by *Intervention* and by *EL* status. Table 5 summarizes model parameters for each of the three reading outcomes. The reader should use values in Table 5 to interpret main effects and interaction effects, rather than Table 4. The value for any given parameter in the model is conditional on the other parameters in the model. For example, the main effect for treatment in Table 5 is the effect for ELs in classes with 10% ELs (*Class % of ELs*). Predicted treatment effects for ELs and non-ELs at other values of *Class % of ELs* can be calculated and plotted (as described later), but they are based on the equation described in the earlier paragraph and have to be interpreted in terms of the codes for *EL* status (i.e., 0/1) and for treatment condition.

The intercept (γ_{000}) for each model is the mean posttest score when all predictors are at zero or at their centered value if other than zero. Coefficients for the main effects of *Intervention* (γ_{010}), *EL* status (γ_{200}), and *Class % of ELs* (γ_{020}) in a model with a three-way interaction represent deviations from this intercept value (Hoffman, 2014). Our interest is in the main effect of treatment on learning and reading outcomes and the moderating influence of *EL* status and *Class % of ELs*. We interpret the results accordingly.

For the ASK, the estimated coefficient for *Intervention* ($\gamma_{010} = 3.58$, $p < .01$, Effect Size (ES) $= .40$) represents the amount by which the intercept ($\gamma_{000} = 21.74$) increases for non-ELs in treatment classes with 10% ELs. The predicted scores for ELs and non-ELs differ as well. The effect of *EL* status ($\gamma_{200} = -2.65$, $p <$

.001, ES $= -.31$) means that ELs in comparison classrooms with 10% ELs scored 2.65 points lower than non-ELs in the same comparison classroom. Among the two-way interactions involving *Intervention*, the coefficient for the *Intervention* $\times$ *EL* term is positive and differs statistically from zero ($\gamma_{300} = 2.35$, $p < .05$), meaning that the difference in knowledge acquisition between ELs and non-ELs is significantly smaller (less negative, in this case) in treatment classes with 10% ELs than in comparison classes with 10% ELs (again, in the context of significant three-way interaction). The significant three-way interaction ($\gamma_{500} = -.08$, $p < .01$) means that the regression coefficient for the interaction of *Intervention* and *EL*, $\gamma_{300} = 2.35$, is conditional on values of *Class % of ELs*. The coefficient or slope for the two-way interaction changes by $-.08$ units for every change in % of Class EL. Another way of saying this is that the EL/non-EL difference in treatment classes widens as EL becomes more prevalent in a class.

To probe the three-way interaction, we plotted predicted scores for the four groups defined by the interaction of *Intervention* and *EL* status across values of *Class % of ELs* (see Figure 2). The reader can find the corresponding *y* value (adjusted posttest score on the ASK) for a given value of *x* (*Class % of ELs*) for any or all of the four groups (i.e., EL in treatment, EL in comparison, non-EL in treatment, and non-EL in comparison). To evaluate the moderating effect of *Class % of ELs* on variation in the *Intervention* $\times$ *EL* status interaction, we calculated regions of significance along *Class % of ELs* based on fixed-effects estimates and their associated covariance matrices, as described by Bauer and Curran (2005). The regions of significance along *Class % of ELs* are bounded by 8.8 and 11.48. This means that the difference between ELs and non-ELs in treatment classes is significantly smaller than the difference between ELs and non-ELs in comparison classes for

Table 4

Pretest and Posttest Means, Standard Deviations, and Ranges for Reading Outcomes by English Learner Status and Treatment Condition

Measures	Pretest				Posttest			
	M	*SD*	*n*	Range	*M*	*SD*	*n*	Range
ASK								
Non-EL								
Comparison	17.12	6.87	504	1–38	22.14	8.27	474	1–41
Treatment	17.67	7.56	507	0–37	25.97	8.98	477	6–42
EL								
Comparison	14.13	5.43	173	4–37	16.59	7.07	163	4–37
Treatment	14.76	6.15	232	2–37	19.93	8.86	223	4–39
MASK								
Non-EL								
Comparison	8.40	4.43	492	1–20	10.41	5.07	446	0–20
Treatment	8.08	4.77	511	0–20	10.66	5.17	449	1–20
EL								
Comparison	5.66	3.34	161	1–17	8.09	4.09	144	1–19
Treatment	6.28	4.09	235	0–20	8.20	4.67	204	0–20
Gates–MacGinitie								
Non-EL								
Comparison	95.08	14.41	495	65–135	96.48	13.69	461	65–135
Treatment	96.22	15.58	508	65–135	97.66	15.63	465	65–135
EL								
Comparison	83.47	13.36	173	65–131	87.98	12.92	150	65–131
Treatment	87.24	14.31	236	65–135	88.70	14.02	199	65–131

Note. ASK = Assessment of Social Studies Knowledge; EL = English language learner; MASK = Modified Assessment of Social Studies Knowledge and Reading Comprehension; Gates–MacGinitie = Gates–MacGinitie Reading Comprehension Subtest.

Table 5
Multilevel Model Results for Outcomes

Measure	Predictor	b	SE	ES
Assessment of Social Studies Knowledge	Intercept	21.74***	.59	
	Pretest	.78***	.03	
	Intervention	3.58**	1.27	.40
	EL	−2.65***	.90	−.31
	Class % of ELs	−.04	.03	
	Intervention × EL	2.35*	1.04	
	Intervention × Class % of ELs	−.01	.05	
	ELL × Class % of ELs	.02	.03	
	Intervention × EL × Class % of ELs	−.08**	.03	
Modified Assessment of Social Studies Knowledge and Reading Comprehension	Intercept	10.34***	.34	
	Pretest	.62***	.04	
	Intervention	1.00*	.41	.20
	EL	−1.26*	.54	−.26
	Class % of ELs	−.06***	.02	
	Intervention × EL	.35	.91	
	Intervention × Class % of ELs	−.02	.02	
	EL × Class % of ELs	.06***	.02	
	Intervention × EL × Class % of ELs	−.04	.03	
Gates–MacGinitie Reading Comprehension Subtest	Intercept	96.17***	1.02	
	Pretest	.60***	.03	
	Intervention	1.71	1.48	.12
	EL	−1.21	1.81	−.08
	Class % of ELs	−.10**	.04	
	Intervention × EL	−1.21	1.81	
	Intervention × Class % of ELs	−.08	.06	
	EL × Class % of ELs	.03	.06	
	Intervention × EL × Class % of ELs	.04	.06	

Note. SE = Standard Error; EL = English Learner; ES = Effect Size.
* $p < .05$. ** $p < .01$. *** $p < .001$.

values of *Class % of ELs* below 8.8%. Between 8.8% and 11.48%, the difference for ELs and non-ELs in comparison and treatment classes does not differ from zero. Above 11.48%, the difference between ELs and non-ELs in comparison classes is significantly smaller than the difference between ELs and non-ELs in the treatment classes, and the difference is increasingly smaller as *Class % of ELs* increases.

For the MASK, the coefficient for the three-way interaction term is not statistically significant ($\gamma_{500} = -.04$); however, we report the effects for the full model (i.e., the model with the three-way interaction term) so that parameter estimates can be compared across the three reading outcomes, and we interpret only the main effects. The main treatment effect on the MASK differs statistically from zero ($\gamma_{010} = 1.00$, $p < .05$, ES = .20), meaning that non-ELs in treatment classrooms with 10% ELs scored about 1 point higher on the posttest than non-ELs in comparison classrooms with 10% ELs. The significant effect for *EL* status ($\gamma_{200} = -1.26$, $p < .05$, ES = −.26) suggests that ELs performed worse than non-ELs in comparison classes with 10% ELs, and the significant effect for *Class % of ELs* ($\gamma_{020} = -.06$, $p < .001$) indicates that MASK scores for non-ELs in comparison classes decrease by .06 points for each additional percentage point on *Class % of ELs*. On the Gates–MacGinitie test, only the main effect of *Class % of ELs* ($\gamma_{020} = -.10$, $p < .01$) is significant and negative, similar to the trend on the MASK.

Discussion

This study investigated the efficacy of the PACT set of instructional practices (Vaughn et al., 2013) adjusted to meet the needs of

ELs in eighth-grade social studies classes. We hypothesized that students who were not ELs would perform similarly to students in previous studies (Vaughn et al., 2013, 2015). We also hypothesized that ELs who received the PACT treatment would outperform ELs in the BAU comparison condition. In sum, we believed that PACT enhanced with instructional practices for ELs would positively affect all learners participating in the treatment condition.

These hypotheses were confirmed. Both ELs and non-ELs who received the treatment outperformed those assigned to the BAU comparison condition on measures of content knowledge acquisition (ES = 0.40) and content-related reading comprehension (ES = 0.20). We interpret these findings as particularly impactful because there is complete overlap in the content taught to the treatment condition and the BAU comparison condition, with the only variation being the manner in which the content was taught. Furthermore, because randomization was at the class level, teacher effects were controlled, allowing for students in the treatment and comparison conditions to have the same teacher. This design provides a challenging test to the treatment, increasing confidence in the effect for content knowledge acquisition. In addition, these findings align with those of prior studies of PACT's 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 administered a standardized reading comprehension measure to determine whether there were differential effects on reading comprehension for students in the treatment or control condition. We did not
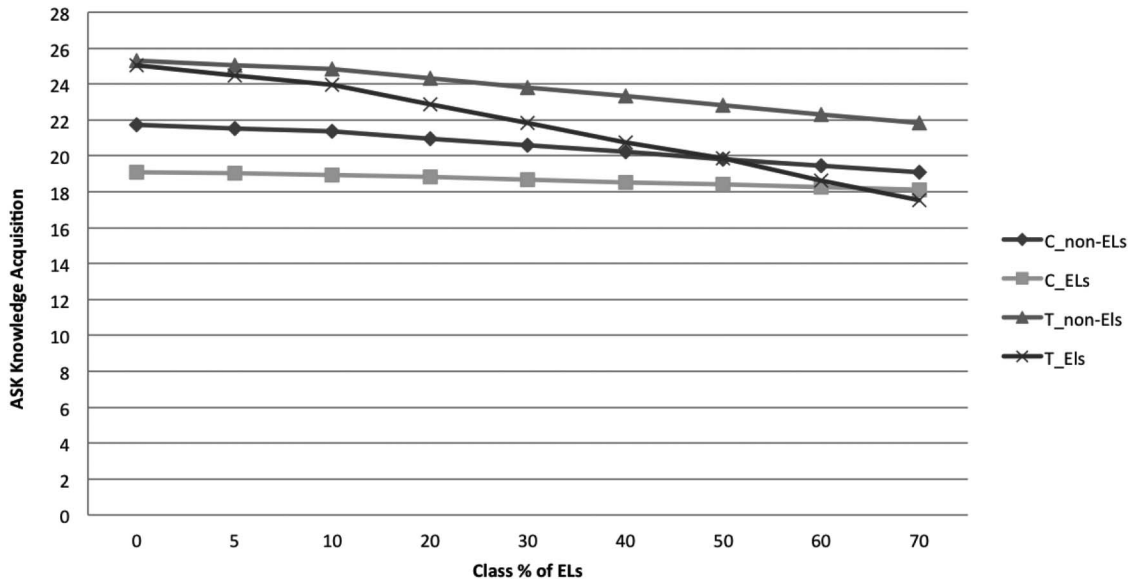
Figure 2. Predicted Assessment of Social Studies Knowledge scores for the four groups as a function of class percentage (%) of ELs. Class % of ELs = percentage of students in each class who are ELs; C_non-ELs = non-ELs randomized to the comparison group; C_ELs = ELs randomized to the comparison condition; T_non-ELs = non-ELs assigned to treatment; T_ELs = ELs randomized to treatment.

hypothesize differences, based on findings from previous PACT studies; however, we also wanted to ensure that participating in the PACT treatment did not negatively affect target students.

We further hypothesized that the benefit of PACT would vary, depending on the class-level prevalence of English academic language, which we defined as the percent of ELs in the classroom. Specifically, we assumed that increases in the class-level percentage of ELs would disadvantage both ELs and non-ELs because sophisticated content-related English academic language is less available to all students under such conditions. Our rationale was that overreliance on discourse-based practices among peers whose language and vocabulary use in English were still developing would reduce the overall effects of the treatment. Our results suggest that the content knowledge acquisition outcomes for ELs and non-ELs are more similar in treatment classes with 10% ELs than in BAU classes with 10% ELs. In other words, at the >0% to 10% EL threshold, ELs and non-ELs respond comparably to the PACT intervention on a measure of content knowledge acquisition.

As predicted, this relative advantage for ELs in PACT diminishes as classrooms become increasingly diverse (increased levels of *Class % of ELs*). Trends for content knowledge acquisition across values of *Class % of ELs* decline for all groups. However, the outcomes for ELs and non-ELs in treatment classrooms become significantly less similar than the relative outcomes for ELs and non-ELs in BAU as classes become increasingly language diverse. In other words, as the percentage of ELs in a class increases, performance on the content knowledge acquisition measure decreases for all students but more dramatically for ELs than for non-ELs. This moderating effect for *Class % of ELs* begins at about 12% ELs and continues across the range of *Class % of ELs*. This finding suggests that the influence of PACT may depend partly on the quality of classroom discourse and that ELs are increasingly more disadvantaged than non-ELs in PACT-like interventions when English academic language is less accessible or less often used by one's classmates.

The question then arises about how to interpret this finding in practice. One interpretation is that discourse-based treatments have a stronger impact on knowledge acquisition for all students when less than 12% of the class are ELs. Conversely, this finding may mean that for a discourse-based treatment to sustain its impact as the proportion of ELs in the class increases, additional supports are necessary. We hesitate to conjecture too much about the practical implications from this study until further studies confirm this finding.

Although students in the treatment classrooms outperformed students in the BAU comparison classrooms on the measure of content knowledge acquisition, it should be noted that the increase in ELs in the classroom in more traditional instruction (i.e., BAU comparison condition) in social studies did not negatively affect the performance of ELs. One possible explanation for this finding is that in traditionally instructed social studies classes (i.e., BAU), ELs spend little time interacting with text and with each other to establish meaning (Swanson et al., 2015).

Examining the moderating effect of proportion of ELs in the class on content comprehension and general comprehension yielded different findings than those for content knowledge acquisition. The absence of a significant two- or three-way interaction with *Intervention* suggests that *EL* status, *Class % of ELs*, and the interaction of *EL* and *Class % of ELs* do not influence the treatment's effect on students' content reading comprehension (MASK) or general reading comprehension (Gates–MacGinitie) outcomes. This finding may be because of the fact that peer discourse and language use occurred largely in the PACT components related to content knowledge (e.g., TBL comprehension check, TBL knowledge application), rather than reading comprehension (e.g., knowledge acquisition through text reading).

Several limitations to the current study should be noted. First, this study was conducted in three school districts in the southeast and southwest United States, limiting generalization to these students and teachers in these regions. In addition, a limitation to most studies investigating ELs in secondary settings, including this one, is the lack of availability of participants' language proficiency in their first and second languages. A standardized measure of English proficiency would be necessary to address whether the "saturation" of ELs in the classroom is a less significant factor when students' proficiency in English is higher. We are unable to address this important question because we had only district-identified classification information on ELs and were not able to obtain a measure of English proficiency.

References

Angrist, J. D., & Lang, K. (2004). Does school integration generate peer effects? Evidence from Boston's Metco program. *The American Economic Review, 94,* 1613–1634. http://dx.doi.org/10.1257/0002828043052169

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,* 345–376. http://dx.doi.org/10.1080/19345740903217623

August, D., & Shanahan, T. (2006). *Developing literacy in second-language learners: Report of the National Literacy Panel on Language-Minority Children and Youth.* Mahwah, NJ: Erlbaum.

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,* 373–400. http://dx.doi.org/10.1207/s15327906mbr4003_5

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

Fix, M., & Passel, J. S. (2003). *U.S. immigration: Trends and implications for schools.* Washington, DC: The Urban Institute.

Francis, D. J., Rivera, M., Lesaux, N., Kieffer, M., & Rivera, H. (2006). *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.centeroninstruction.org/files/ELL1-Interventions.pdf

Francis, D. J., & Vaughn, S. (2009). Effective practices for English language learners in the middle grades: Introduction to the special issue of Journal of Research on Educational Effectiveness. *Journal of Research on Educational Effectiveness, 2,* 289–296. http://dx.doi.org/10.1080/19345740903217664

Gelfand, D. M., & Hartmann, D. P. (1975). *Child behavior analysis and therapy.* New York, NY: Pergamon Press.

Gottfried, M. A. (2014). Classmates with disabilities and students' non-cognitive outcomes. *Educational Evaluation and Policy Analysis, 36,* 20–43. http://dx.doi.org/10.3102/0162373713493130

Gwet, K. L. (2012). *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters* (3rd ed.). Gaithersburg, MD: Advanced Analytics.

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

Hernandez, D. J. (2012). *Double jeopardy: How third-grade reading skills and poverty influence high school graduation.* Baltimore, MD: The Annie E. Casey Foundation.

Hintze, J. M. (2005). Psychometrics of direct observation. *School Psychology Review, 34,* 507–519.

Hoffman, L. (2014). *Longitudinal analysis: Modeling within-person fluctuation and change.* New York, NY: Routledge Academic.

Kennelly, L., & Monrad, M. (2007). *Approaches to dropout prevention: Heeding early warning signs with appropriate interventions.* Washington, DC: National High School Center at the American Institutes for Research.

Kieffer, M. J. (2008). Catching up or falling behind? Initial English proficiency, concentrated poverty, and the reading growth of language minority learners in the United States. *Journal of Educational Psychology, 100,* 851–868. http://dx.doi.org/10.1037/0022-0663.100.4.851

Kieffer, M. J. (2010). Socioeconomic status, English proficiency, and late-emerging reading difficulties. *Educational Researcher, 39,* 484–486. http://dx.doi.org/10.3102/0013189X10378400

Landis, J. R., & Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics, 33,* 363–374. http://dx.doi.org/10.2307/2529786

Lesaux, N. K., & Kieffer, M. J. (2010). Exploring sources of reading comprehension difficulties among language minority learners and their classmates in early adolescence. *American Educational Research Journal, 47,* 596–632. http://dx.doi.org/10.3102/0002831209355469

Lesaux, N. K., Kieffer, 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,* 196–228. http://dx.doi.org/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,* 1159–1194. http://dx.doi.org/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. http://dx.doi.org/10.1007/s11145-009-9215-5

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

National Center for Education Statistics. (2009). *The nation's report card: Reading 2009* (NCES 2010–458). Washington, DC: Institute of Education Sciences.

National Center for Education Statistics. (2013). *The nation's report card: 2013 mathematics and reading—Which student groups are making gains?* Retrieved from http://www.nationsreportcard.gov/reading_math_2013/#/gains-by-group

National Clearinghouse for English Language Acquisition. (2006). *The growing numbers of limited English proficient students: 1993/4–2003/4.* Washington, DC: Office of English Language Acquisition.

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standards.* Washington, DC: Author.

Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service. (2012). *National evaluation of Title III implementation—Report on state and local implementation*. Washington, DC: Author.

Snow, C., Lawrence, J., & White, C. (2009). Generating knowledge of academic language among urban middle school students. *Journal of Research on Educational Effectiveness, 2,* 325–344. http://dx.doi.org/10.1080/19345740903167042 (Corrigendum published 2014, *Journal of Research on Educational Effectiveness, 7,* p. 407. http://dx.doi.org/10.1080/19345747.2014.967563)

Snow, C. E., & Uccelli, P. (2009). The challenge of academic language. In D. Olson & N. Torrance (Eds.), *The Cambridge handbook of literacy* (pp. 112–133). New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511609664.008

Suen, H. K., & Ary, D. (1989). *Analyzing quantitative behavioral observation data*. Hillsdale, NJ: Erlbaum.

Swanson, E. A., 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,* 426–442. http://dx.doi.org/10.1177/0014402914563704

The Meadows Center for Preventing Educational Risk. (2013). *Sample lesson components for PACT Instruction Refined for English Language Learners*. College of Education, The University of Texas at Austin, Austin, TX.

Vaughn, S., Martinez, L. R., Linan-Thompson, S., Reutebuch, C. K., Carlson, C. D., & Francis, D. J. (2009). Enhancing social studies vocabulary and comprehension for seventh-grade English language learners: Findings from two experimental studies. *Journal of Research on Educational Effectiveness, 2,* 297–324. http://dx.doi.org/10.1080/19345740903167018

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,* 31–50.

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,* 77–93.

Wanzek, J., Kent, S. C., Vaughn, S., Swanson, E. A., Roberts, G., & Haynes, M. (2015). Implementing team-based learning in middle school social studies classes. *The Journal of Educational Research, 108,* 331–344.