

Teaching Bilingual Learners: Effects of a Language-Based Reading Intervention on Academic Language and Reading Comprehension in Grades 4 and 5

C. Patrick Proctor

*Boston College, Chestnut Hill,
Massachusetts, USA*

Rebecca D. Silverman

Stanford University, California, USA

Jeffrey R. Harring

*University of Maryland, College Park,
USA*

Renata Love Jones

*Boston College, Chestnut Hill,
Massachusetts, USA*

Anna M. Hartranft

*University of Maryland, College Park,
USA*

ABSTRACT

Students are expected to comprehend and produce increasingly complex texts in upper elementary school, and academic language and literacy skills are considered critical to meeting these expectations. Notions of academic language are also controversial and require careful deliberation when applied to traditionally minoritized populations, including bilingual learners who negotiate more than one language in their daily lives and have varied linguistic repertoires. In the present study, the authors report on a quasi-experimental field trial of a theoretically grounded and language-based reading intervention framed around language components (semantics, syntax, and morphology), language functions, discussion, and reading comprehension. A sample of 239 Portuguese-English and Spanish-English bilingual students in grades 4 and 5 worked in small instructional groups to explore language, apply reading strategies, and discuss and write about big ideas in text. Half of the students were assigned to the intervention group ($n = 119$) and the other half ($n = 120$) to a business-as-usual control group. Classroom teachers ($n = 12$) and specialists ($n = 10$) implemented the intervention with small groups of four to six students. Results showed practically meaningful effects of the intervention on standardized measures of both academic language (Hedges's $g = 0.248$) and reading comprehension (Hedges's $g = 0.166$), with implications for theory, research, and classroom practice.

Approximately one out of every five children in the United States grows up speaking a language other than English at home (Ryan, 2013). These children and youths are, to varying degrees, bilingual (Grosjean, 2010) and constitute the fastest growing population in U.S. schools (Shin, 2013). This growth has been propelled in large measure through increases in the enrollment of Latinx students, which rose from 9.0 million to 12.5 million between 2003 and 2013 and is projected to increase to 14.7 million by 2025 (National Center for Education Statistics, 2017). As schools become ever more linguistically diverse, attention to the language and literacy skills of bilingual students is critical. However, teachers continue to report feeling underprepared to support bilingual learners, particularly in academic language and literacy instruction (Gándara & Santibañez, 2016; Samson & Collins, 2012).

Research over the past 10–15 years has converged on some basic understandings about academic language and literacy outcomes for bilingual learners. Specifically, we have learned that code-based skills (e.g.,

phonological awareness, decoding, word reading) do not present as unique or outsized challenges for bilingual learners as compared with their monolingual counterparts (e.g., Lesaux, Crosson, Kieffer, & Pierce, 2010). However, many students, including bilingual learners, struggle with the academic language skills (e.g., academic vocabulary, morphology, syntax) needed to understand complex texts across content areas in school (Goldenberg, 2013). Academic language is the unique register endemic to much schooling in the United States. It is “a constellation of the high-utility language skills that correspond to linguistic features that are prevalent in academic discourse across school content areas and infrequent in colloquial conversations” (Uccelli, Phillips Galloway, Barr, Meneses, & Dobbs, 2015, p. 338).

The research base on providing academic language support for bilingual students has grown in recent years (see Baker et al., 2014; Brisk & Zhang-Wu, 2017). However, despite the growth, relatively few studies have investigated interventions that include attention to academic language and literacy skills for fourth- and fifth-grade bilingual learners. For example, a meta-analysis of studies of reading instruction with bilingual learners in grades 4–8 found 11 studies, only one of which included students in fourth grade and six of which included students in fifth grade (Hall, Roberts, Cho, McCulley, Carroll, & Vaughn, 2017). Overall, Hall et al. (2017) found that effect size outcomes were stronger when interventions attended to both vocabulary and comprehension (Hedges’s $g = 0.39$) rather than instruction on vocabulary alone ($g = 0.08$). Mean effect sizes across standardized and unstandardized vocabulary and comprehension measures suggested moderate effects of intervention ($g = 0.35$), but these effects were markedly diminished for standardized measures ($g = 0.01$) as compared with experimental measures ($g = 0.43$).

If the field is to move forward with respect to literacy interventions, we need to find ways to design instruction that impacts those hard-to-move standardized reading and language outcomes that have proved stubbornly unresponsive in previous research. In service of this, Hall et al. (2017) suggested that future interventions may require emphasis on broader dimensions of language, beyond vocabulary and comprehension, to show effects on standardized literacy measures. The present intervention, called CLAVES (an acronym for comprehension, linguistic awareness, and vocabulary in English and Spanish, meaning *keys* or *clues* in Spanish) meets this need by focusing broadly on language instruction via vocabulary, morphology, and syntax, alongside guided reading, discussion, and writing activities. Academic language is something that all students, to varying degrees, are acquiring, but bilingual students, who navigate such acquisition across languages and instructional contexts, are arguably unique. Thus, in designing CLAVES, we aimed to develop an intervention that provided access to academic language and to promote

reading comprehension while simultaneously valuing students’ linguistic resources and recognizing the varied instructional contexts in which bilingual students learn. In the present study, CLAVES was implemented in small groups and was designed to affect standardized academic language and reading comprehension outcomes.

Theoretical Framework

Core Academic Language Skills

In recent years, the topic of academic language has received considerable attention as a mainstream instructional commodity. Academic language is important because students must be able to linguistically engage with content area disciplines in school. From understanding a lecture or video presentation, to participating in whole- or small-group discussions, to reading and writing in science, mathematics, social studies, and English language arts (ELA), academic language is a key ingredient to success in school and beyond (e.g., Bailey, 2007; Schleppegrell, 2004). The importance of academic language tends to increase in upper elementary school when expectations to learn through reading grow alongside the increasingly complex language of instructional texts (Kieffer, 2010; Mancilla-Martinez & Lesaux, 2010). This phenomenon creates a linguistic challenge, and many students have resulting language-related reading difficulties. For bilingual students, this linguistic challenge can be exacerbated as they navigate the complexity of proficiencies in two languages while typically being taught monolingually, via a language in which they are not always fully proficient (i.e., English; Goldenberg, 2013). Thus, interventions that aim to support academic language may be of particular value in the upper elementary grades for bilingual learners.

For the purposes of the current study, we started with Uccelli and colleagues’ (Uccelli, Barr, et al, 2015; Uccelli, Phillips Galloway, et al., 2015) recent theoretical conceptualization and empirical operationalization of academic language, which they referred to as *core academic language skills* (CALS). To the best of our knowledge, Uccelli and colleagues are the first researchers who have attempted to both theorize and operationalize academic language. The CALS construct derives from an interdisciplinary perspective on academic language that convenes second-language acquisition (Cummins, 1979, 1991), systemic functional linguistics (Brisk & Zhang-Wu, 2017; Halliday & Matthiessen, 2014), and theories of reading comprehension (Hoover & Gough, 1990; Perfetti & Stafura, 2014).

From a second-language perspective, Cummins (1979, 1991) hypothesized that cognitive academic language proficiency (CALP) is necessary for success in schools, with a particular focus on second-language learners. A justified criticism of this view of language arises when it is compared with basic interpersonal communication skills, a

comparison which sets up a binary that has been problematized for implying that CALP is somehow superior to basic interpersonal communication skills, and thus promotes unequal, diglossic contexts in school (e.g., Flores, 2015; Poza, 2015; Valdés, 2004). We agree with these critiques and make no assumptions here about the superiority of one linguistic register over another. We further contend that bringing all languages and linguistic registers to bear on teaching and learning allows students access to a full linguistic repertoire (García & Wei, 2014) that can enhance understanding and expression.

Yet, as far as theoretical definitions of academic language are concerned, Cummins's (1979) early description of CALP was noteworthy. The framework comprised three major components: vocabulary-concept knowledge, metalinguistic insights, and decontextualized language. The first, vocabulary-concept knowledge, requires a student's understanding of broad "concepts or meanings embodied in words" (Cummins, 1979, p. 287) and maps to contemporary notions of general academic vocabulary knowledge that middle-grade interventions have recently targeted (e.g., Carlo et al., 2004; Lesaux et al., 2010; Lesaux, Kieffer, Kelley, & Harris, 2014; Proctor et al., 2011; Snow, Lawrence, & White, 2009). The second component of the CALP framework, metalinguistic insights, revolves around becoming aware that written language functions differently from spoken language in school. Finally, the third component of CALP describes the highly decontextualized nature of academic language, which typically comprises abstract concepts and ideas.

Academic language, however, is more than just vocabulary, metalinguistics, and abstraction. It is a semi-otic system that changes as a function of communicative and literary contexts. Systemic functional linguistics (Bailey, 2007; Gibbons, 2009; Schleppegrell, 2004) thus provides an important expansion to the CALP framework for the current conceptualization of academic language. For example, the language of history is distinctive in many ways from the language of mathematics, which is also different from the language of science (see, e.g., Fang & Schleppegrell, 2008). Despite discipline-specific differences, however, there are also linguistic commonalities across content areas. Brisk and Zhang-Wu (2017) described a set of academic language characteristics that collocate across disciplines, including the presence of nominalizations, the passive voice, and multiclausal sentences that cohere around conjunctions and connectives. Like the CALP framework, vocabulary knowledge is also central to systemic functional linguistics orientations to academic language, as academic vocabulary is often packed with conceptual knowledge that goes beyond simple word definitions (e.g., *photosynthesis*, *democracy*) and word meanings that can vary across disciplinary contexts (e.g., the word *factor* in history vs. mathematics).

Bringing acquisition and functional use into a conceptualization of academic language has clear implications for literacy generally and reading comprehension specifically. That is, given that academic language is incorporated in the texts that students are expected to read in school, it is important to consider its role in reading comprehension. The simple view of reading broadly articulates that reading comprehension is the interaction between decoding and linguistic comprehension (Hoover & Gough, 1990). Hoover and Gough (1990) described linguistic comprehension as "the ability to take lexical information (i.e., semantic information at the word level) and derive sentence and discourse interpretations" (p. 131). Going further, Perfetti and colleagues (Perfetti & Hart, 2002; Perfetti & Stafura, 2014) argued that the quality of that lexical information is driven by how well a reader knows not only how words are pronounced and spelled but also how words are connected semantically (e.g., *power*, *strength*, *electricity*), derived morphologically (e.g., *power*, *powerless*, *powerlessness*), and used syntactically (e.g., *power* = noun or verb; *powerless* = adjective). Central to these theories of reading is the role of language. Without a broad set of linguistic skills (the simple view) and a deep understanding of words (lexical quality), comprehension will be compromised.

Situational and Critical Contexts

Uccelli and colleagues' (Uccelli, Barr, et al., 2015; Uccelli, Phillips Galloway, et al., 2015) interdisciplinary approach to theorizing a construct of academic language is unique and relevant for the current study, but we integrated two additional perspectives that allow for a broader view of language and literacy as contextually situated and critically framed. First, the RAND Reading Study Group's (2002) conception of reading brings the CALS theoretical frameworks together to situate the current study. In their heuristic, the reader, the text, and the activity interact within varying sociocultural settings. The reader brings a host of attributes to a text, including cognition, motivation, knowledge, and experience. The text will also vary in terms of content, language, and modality. The reading activity is critical for comprehension because students' purposes for reading (e.g., getting the gist vs. a close reading of a paragraph) interact with reader and text characteristics. Finally, these interactions take place across broad sociocultural contexts that extend beyond just the classroom walls and include race, income, language, and zip code. Indeed, as the RAND Reading Study Group pointed out, "if the education community is to ensure universal success in reading comprehension, those in the community must understand the full range of sociocultural differences in communicative practices" (p. xvi).

Relatedly, we recognize that the very idea of academic language has been hotly debated. Critical literacy scholars have argued that "uncritical framings of academic

language as an objective set of linguistic forms that are dichotomous with the playground language of Latinos and other language-minoritized students” (Flores, 2015, para. 7) are problematic in that they situate some students in schools as having academic language and others as lacking it. This is problematic because proficiency in academic language and literacy is privileged in U.S. schools, as well as in many professional occupations, and as a result, command over this unique register of English is important for students to establish. As such, some educators have called for teaching academic language in support of bilingual students in school (e.g., Baker et al., 2014; Brisk & Zhang-Wu, 2017; Gebhard, Chen, & Britton, 2014; Uccelli, Phillips Galloway, & Qin, in press), whereas others have cautioned about this approach, expressing concerns about how bilingual students are characterized simply as lacking language and in need of remediation (Flores & Rosa, 2015; Poza, 2015; Valdés, 2004). In this study, we engaged these critical perspectives and were mindful in recognizing that students bring a breadth of linguistic knowledge and lived experiences to bear on teaching and learning (e.g., Goodwin & Jiménez, 2016).

Taking all of the above into account, we designed CLAVES to leverage how the readers (Spanish– and Portuguese–English bilingual fourth and fifth graders), the texts (selected or written by the research team), and the activities (targeting text-based vocabulary, morphology, syntax, discussion, and writing) interacted with one another across a variety of schooling (monolingual and bilingual education) and linguistic (Spanish– and Portuguese–English bilingualism) contexts, in the service of critically promoting academic language and reading comprehension.

Building on Academic Vocabulary

The intervention reported on here is not the first to address academic language, but to our knowledge, it extends previous 21st-century intervention work that has more specifically targeted academic vocabulary. In a relatively early study, Carlo et al. (2004) tested a whole-class and small-group English vocabulary enrichment program that combined explicit vocabulary and word-learning instructional strategies, with fifth-grade bilingual and monolingual students. The intervention was text-based and sought to teach words by targeting which words, how to introduce them instructionally, how frequently to use them, what aspects of word knowledge to target, and instructional techniques. The researchers employed heterogeneous grouping and organized the intervention around the topic of immigration. Lesaux et al. (2010, 2014) expanded on academic vocabulary intervention work with sixth-grade multilingual students. Like Carlo et al., Lesaux et al. (2014) “included a variety of whole-group, small-group, and

independent activities designed to incrementally build word knowledge” (p. 1169) and argued for the central importance of high-utility academic words (e.g., *factor*, *analyze*) that commonly appear in upper elementary and secondary-level texts. Proctor et al. (2011) tested a digital literacy environment with fifth-grade Spanish-English bilingual students and their monolingual counterparts that sought to leverage depth of academic vocabulary by exposing students to a variety of text-based word work. Students worked individually at computers and engaged with semantically oriented approaches that asked students to make personal and conceptual connections between words, identify cognates, and write or record definitions and image captions for relevant words. Snow et al. (2009) tested a grades 6–8 intervention that incorporated research-based vocabulary instruction but also included targeted opportunities for students to use taught words in classroom discussion, debate, and writing about high-interest texts.

All of these programs showed effects on vocabulary and/or comprehension outcomes. Carlo et al. (2004) found no effect on a standardized measure of vocabulary knowledge, but effects on researcher-developed measures of vocabulary and other academic language skills included reading comprehension ($\eta^2 = .05$), word mastery ($\eta^2 = .34$), and polysemy production ($\eta^2 = .05$). Lesaux et al.’s (2014) intervention had significant, positive, statistically significant effects on researcher-developed measures of word mastery and associations (Cohen’s $d = 0.41$ and 0.22 respectively), morphological decomposition and derivations ($d = 0.40$ and 0.21 , respectively), written expression ($d = 0.19$), and researcher-developed reading comprehension ($d = 0.15$). In general, the effects were larger for bilinguals and for students who began the program with limited vocabulary knowledge. Proctor et al. (2011) found no effects of the intervention on a standardized measure of reading comprehension but found that the amount of activities completed and texts read was associated with significant improvements on a measure of standardized vocabulary knowledge relative to control students. For their intervention, Snow et al. (2009) found an overall effect size of $d = 0.49$ across participating schools on a researcher-developed measure of academic vocabulary, and bilingual students in the treatment group showed greater growth on this measure than did monolinguals.

Thus, although we have learned a great deal about academic vocabulary instruction over the years, the broad framework of academic language that we have described (Uccelli, Barr, et al., 2015; Uccelli, Phillips Galloway, et al., 2015) makes clear the need to move beyond vocabulary and more squarely into language. Reading research has converged on the critical role of language for comprehension (Hoover & Gough, 1990; Perfetti & Hart, 2002; Perfetti & Stafura, 2014), and conceptualizations of CALP (Cummins, 1979) and

systemic functional linguistics (Brisk & Zhang-Wu, 2017; Schleppegrell, 2004) provide further nuance and contour to the nature of the types of language that are expected to predict reading outcomes for upper elementary school students.

What Works and for Whom

The CLAVES intervention drew on seminal theories, as well as recent instructional reports of what works for bilingual learners in K–8 contexts. Specifically, a recent report published by the Institute of Education Sciences synthesized research on teaching academic language and literacy to bilingual students in elementary and middle school (Baker et al., 2014). The report outlined four broad recommendations:

- **Recommendation 1:** Teach a set of academic vocabulary words intensively across several days using a variety of instructional activities.
- **Recommendation 2:** Integrate oral and written English language instruction into content-area teaching.
- **Recommendation 3:** Provide regular, structured opportunities to develop written language skills.
- **Recommendation 4:** Provide small-group instructional intervention to students struggling in areas of literacy and English language development. (p. 3)

In our view, the first three recommendations require an operating theory of language to implement them instructionally. Thus, the theoretical frameworks described above constituted a languages-in-literacy foundation for implementing the first three recommendations outlined in Baker et al.'s (2014) report. We also followed the fourth recommendation and developed CLAVES around small-group instruction, which aligns with previous intervention work on academic vocabulary.

In addition to following guidance about what works, we designed the CLAVES intervention to also consider for whom. Specifically, we designed the intervention, with feedback from teachers and through work with students, with a focus on Spanish– and Portuguese–English bilingual learners at English-language proficiency levels ranging from moderately to fully English proficient per standards set forth in the WIDA (2015) standards (i.e., WIDA levels 3–6). The students' bilingualism adds instructional depth when thinking about for whom we are designing instruction. CLAVES was developed to leverage a fuller range of linguistic assets to access academic language and literacy in school. Additionally, given the broad variability in language and literacy outcomes that presents for bilingual learners (Calderón, Slavin, & Sánchez, 2011), we wanted to know whether the effects of this approach differed for bilingual students with different levels of pre-intervention language and literacy. In

sum, to broaden the research base on what works and for whom, we specifically designed CLAVES for Spanish– and Portuguese–English bilingual students and then asked two research questions:

1. Do bilingual students assigned to CLAVES outperform their counterparts assigned to the business-as-usual control group on standardized measures of academic language and reading comprehension?
2. Do pre-intervention language proficiency and reading comprehension covariates moderate main effects of CLAVES on standardized measures of academic language and reading comprehension?

To answer these research questions, we implemented a quasi-experimental field trial of CLAVES in eight schools in two regions of the United States: the Northeast and the Mid-Atlantic. We trained teachers and specialists to implement the intervention with small groups of fourth- and fifth-grade bilingual students. To answer research question 1, we controlled for scores on pre-intervention measures and compared students' post-intervention performance on standardized measures of academic language and reading comprehension as a function of assignment to condition (CLAVES or business-as-usual control). To answer research question 2, we tested interactions between pre-intervention measures and condition to determine whether individual differences in starting points were associated with effects of intervention on standardized academic language and reading comprehension.

Method

Participants

Schools

The intervention was implemented in the Northeast and Mid-Atlantic regions of the United States at four schools in each region. Table 1 shows the size, program type, instructional languages, grade levels, and race, language background, and income statistics for each of the eight participating schools. As the table shows, Latinx students composed the majority population in all schools, save for school 4 in the Northeast, where white students were the majority population. This particular school educated large numbers of children with Brazilian heritage who were classified as white in school records data. Across all schools, the percentage of students who were bilingual (59.5–81.4% in the Northeast region; the Mid-Atlantic region did not report this statistic) and who were labeled by their districts as English learners (ELs; 43.1–58.1%) were well above national averages. More than half of the students were considered economically disadvantaged, as

TABLE 1
Characteristics of the Participating Schools

	School 1	School 2	School 3	School 4	School 5	School 6	School 7	School 8
Region	Northeast	Northeast	Northeast	Northeast	Mid-Atlantic	Mid-Atlantic	Mid-Atlantic	Mid-Atlantic
Size	681	476	1,004	566	527	845	795	808
Program	Dual language	Transitional bilingual education	Monolingual	Transitional bilingual education	Monolingual	Monolingual	Monolingual	Monolingual
Languages	Spanish–English	Spanish–English	English	Portuguese–English	English	English	English	English
Grades	K–5	K–5	Pre-K–8	K–5	K–5	K–5	K–5	K–5
Multirace (percentage)	3.2	1.9	0.6	2.1	a	a	a	a
Latinx (percentage)	69.9	54	87.3	22.1	78.7	90.0	84.0	85.5
Black (percentage)	3.7	6.1	3.2	9.5	12.3	5.3	8.7	11.1
Asian (percentage)	0.6	3.8	2.3	0.7	4.4	a	a	1.9
Native American (percentage)	a	a	a	0.2	a	a	a	a
Hawaiian/Pacific Islander (percentage)	a	a	a	0.0	a	a	a	a
White (percentage)	3.2	34.2	6.6	65.4	3.5	3.1	5.0	a
Bilingual (percentage)	59.5	60.7	81.4	80.1	Not available	Not available	Not available	Not available
English learner (percentage)	43.8	43.1	59.4	58.5	54.6	58.2	60.2	59.9
Economically disadvantaged (percentage)	47.1	49.8	64.4	55.7	83.6	>95.0	88.5	93.2

*Fewer than 10 students.

indicated by qualifying for the National School Lunch Program.

The Northeast schools represented programmatic diversity as well. Of these four schools, one was a Spanish–English dual–language program, in which instruction was delivered 50% of the time in Spanish and 50% of the time in English (the current intervention was delivered during English time). Two other schools housed transitional bilingual education programs (one in Spanish, the other in Portuguese); however, none of the participating students in these two schools were bilingually instructed at the time of the study. Finally, one of the Northeast schools was entirely monolingual, as were all four of the Mid-Atlantic schools.

Teachers

The participating teachers ($n = 22$) delivered the intervention. These same teachers were also responsible for control group instruction. They held one of two roles at their schools: classroom teacher or specialist teacher. Among the implementing teachers, there were 12 classroom teachers (seven in Northeast schools, five in Mid-Atlantic schools) and 10 specialist teachers (five in Northeast schools, five in Mid-Atlantic schools). Of the 12 participating teachers in the Northeast schools, all were fully certified, seven had master’s degrees, nine were white, and three were Latinx. Seven of these teachers spoke either Spanish ($n = 5$) or Portuguese ($n = 2$) in addition to English. Participating teachers had taught in K–6 for an

average of 12 years, with a range of 4–15+ years. Of the 10 participating teachers in the Mid-Atlantic schools, all were fully certified, four had master’s degrees, seven were white, one was Asian/Pacific Islander, and two were African American. Two spoke a language in addition to English and had taught in K–6 for an average of 7 years, with a range of 2–13 years.

Students

All students ($n = 239$) were bilingual (Spanish– or Portuguese–English) and currently or formerly designated by their schools as ELs. These designations were used across all eight schools to identify students who spoke a language other than English at home and who were not fully proficient across English-language domains according to the WIDA (2015) ACCESS for ELLs assessment. WIDA categorizes six overall classes of language proficiency: level 1, *entering*; level 2, *beginning*; level 3, *developing*; level 4, *expanding*; level 5, *bridging*; and level 6, *reaching*. The districts in the current study reclassified students from EL status when they reached level 5 or 6. Thus, a student with a level 5 or 6 WIDA score was considered a former EL (FEL). Over the course of our two-year development work, we included students from across all WIDA levels; however, in consultation with teachers during the development phase, it was concluded that EL students at WIDA levels 3 and 4 and FEL students at WIDA levels 5 and 6 were most likely to benefit from the CLAVES instructional approaches. Given the growing English proficiency of these students, they were typically provided with limited supplemental support. However, teachers felt that such supplemental support would be useful to these students who were still learning to use their linguistic resources to access English-language and literacy instruction in school.

With this stratification criterion in place, students were selected for participation in the intervention or control group. Table 2 displays the sample breakdown by region and EL/FEL status, cross-tabulated with condition and WIDA levels. At the Northeast site ($n = 111$), 79 students (58 females) were in grade 4, and 32 students were in grade 5. In three of the schools, 88 students spoke Spanish and English. The remaining 23 students in school 4 spoke Portuguese and English. Of the 111 students in the Northeast schools, 45 were designated as EL and 66 as FEL. A total of 60 students were in the intervention group (18 in grade 5), and 51 students were in the business-as-usual comparison group (14 in grade 5). In the Mid-Atlantic schools, all students received English-only instruction, and all 128 students were in fourth grade. Overall, 55 students were currently designated as EL and 73 as FEL. In the Mid-Atlantic schools, 60 students were in the intervention group, and 68 students were in the business-as-usual comparison group. All students spoke Spanish and English. Across the sample, of the 100

TABLE 2
Sample Breakdown by Region and EL Status, Cross-Tabulated With Condition and WIDA Level

Condition	Northeast ($n = 111$)		Mid-Atlantic ($n = 128$)		Total
	EL	FEL	EL	FEL	
Intervention	23	37	27	32	119
Control	22	29	28	41	120
<i>WIDA level</i>					
3	11	0	9	0	20
4	34	0	46	0	80
5	0	41	0	0	41
6	0	25	0	73	98
Total	45	66	55	73	239

Note. EL = English learner; FEL = former English learner.

students who were still considered ELs, 20 (20%) were at WIDA level 3, and 80 (80%) were at WIDA level 4. Of the remaining 139 FEL students, 41 (29%) were at WIDA level 5, and 98 (71%) were at WIDA level 6.

The CLAVES Intervention

We developed CLAVES over a two-year period of working with teachers and students to select texts, to experiment with lesson plan implementation and instructional approaches, and to build units. Text selection, and representation in those texts, served as an important starting point for unit construction. Over the two-year development period preceding the field trial, unit themes and texts were derived from direct consultation with bilingual fourth- and fifth-grade students. Indeed, per student insistence, books that involved animals and animal characters were included in the text selection, and although texts were chosen to align with the unit themes, they were also selected to privilege authors, illustrators, and/or characters who were, like the participating students, bilingual people of color, often with immigrant backgrounds and experiences.

CLAVES comprised three thematic units that integrated ELA and social studies content and included rich, complex texts, including print and supplemental videos that presented different perspectives on controversial issues (e.g., deforestation, workers’ rights, bilingual education). Each unit comprised three instructional cycles (see Table 3). The first two cycles were text-based and consisted of five days of language-based reading instruction. The third cycle in a unit was a three-day writing cycle. Table 3 shows that cycles 1 and 2 focused on academic vocabulary and reading comprehension (days 1 and 2), morphology and syntax (days 3 and 4), and a culminating small-group

TABLE 3
Unit Overview Consisting of Two 5-Day Text-Based Cycles and One 3-Day Writing Cycle

Cycle	Day	Focus
Cycle 1: Text-based language and comprehension	1	Comprehension and vocabulary
	2	Comprehension and vocabulary
	3	Morphology
	4	Syntax
	5	Discussion
Cycle 2: Text-based language and comprehension	1	Comprehension and vocabulary
	2	Comprehension and vocabulary
	3	Morphology
	4	Syntax
	5	Discussion
Cycle 3: Writing	1	Planning
	2	Drafting and revising
	3	Publishing

discussion on the fifth and final day of the cycle. Cycle 3 took the form of an abbreviated writer’s workshop. Attention to the students’ bilingualism (Spanish– or Portuguese–English) was part of the instructional design. Table 4 shows the focus of each unit (nature, rights, or immigration), the texts used for each cycle in that unit, and the discussion questions that framed each text. See the Appendix for a full overview of the intervention. Below, we provide a research-based rationale for our instructional approaches, describing the five-day text-based language and comprehension cycles and the three-day writing cycle (see Table 3). Throughout, we describe how bilingual approaches were integrated into the instructional design of this English-language intervention.

Days 1 and 2: Comprehension and Vocabulary

The first two days of a cycle included activating prior knowledge and introducing academic vocabulary using student-friendly definitions, examples, and multimedia supports, such as interactive PowerPoint presentations and short videos (see Figure 1 for an example), which have been found in prior research to benefit bilingual learners (Baker et al., 2014; Lesaux et al., 2010, 2014; Proctor, Dalton, & Grisham, 2007; Silverman & Hines, 2009). Additionally, students were encouraged to notice cognates and use translation to support their word learning, which research has shown to be

beneficial for bilingual students (Goodwin & Jiménez, 2016; Hernández, Montelongo, & Herter, 2016). Academic vocabulary instruction included the development of semantic maps of words and using contextual analysis to determine word meanings (e.g., Baumann, Edwards, Boland, Olejnik, & Kameñui, 2003; Lipson, 1995).

During these first two days, vocabulary instruction was followed by guided reading of text. Recent evidence indicates that using comprehension strategies is important for bilingual learners (e.g., Taboada, Bianco, & Bowerman, 2012; Taboada & Rutherford, 2011), and thus intervention teachers were prompted to have students summarize, question, and make inferences throughout guided reading to build comprehension. In particular, students focused on identifying author’s perspective and arguments on both sides of controversial issues (see the discussion questions in Table 4) using more than one source (e.g., text, video). Note that students had been taught comprehension strategies in their ELA instruction. The intervention provided additional opportunities and support for students to practice these strategies in a small-group context. Following other interventions with bilingual students in elementary and middle school, lesson plans guided teachers to encourage student-centered discussion of texts and to compare and contrast ideas across texts with their peers (Carlo et al., 2004; Lesaux et al., 2010, 2014).

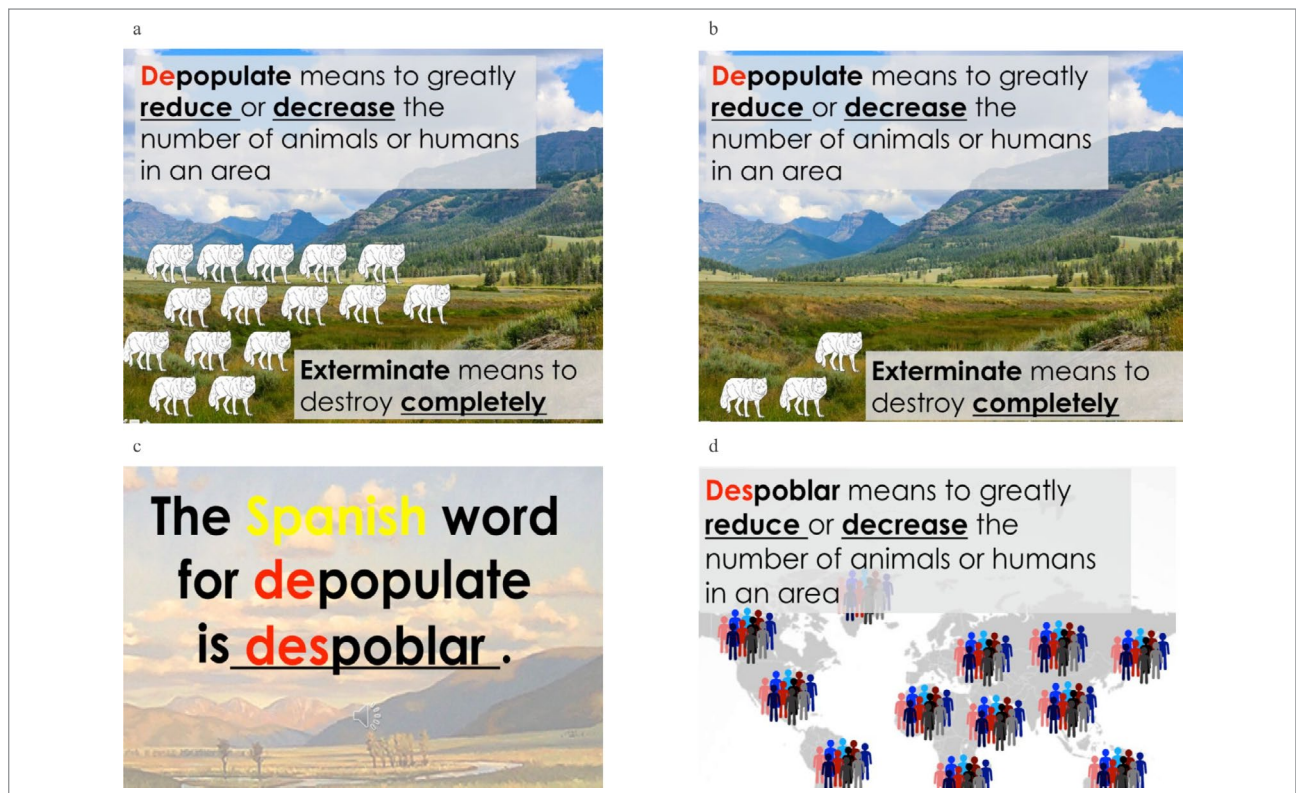
Days 3 and 4: Morphology and Syntax

Days 3 and 4 of the instructional cycle included explicit attention to morphology and syntax. Instruction focused on developing students’ understanding about the language components generally but specifically used morphological and syntactic examples from the intervention texts. We chose to focus on morphology and syntax because of their documented relations with both academic language (Brisk & Zhang-Wu, 2017; Uccelli et al., in press) and reading comprehension (Proctor, Silverman, Haring, & Montecillo, 2012; Silverman et al., 2015). In line with previous research, morphology instruction included a focus on common prefixes, suffixes (Baumann et al., 2003; Goodwin & Perkins, 2015; Kieffer & Lesaux, 2007), and root word meanings (Reed, 2008) via word study techniques (Bear, Templeton, & Invernizzi, 2008; Ganske, 1999). Building on work on supporting syntax, instruction also targeted tracking pronominal references, identifying tense, and connecting ideas in complex sentences using sentence mapping, combining, and segmenting (Stoddard, Valcante, Sindelar, O’Shea, & Algozzine, 1993; P.A. Wilkinson & Patty, 1993). Finally, across both morphology and syntax instruction, gamification techniques were employed to strengthen linguistic understanding. For example, students created new words via affix manipulation and discussed whether constructed words were real and

TABLE 4
Unit Content Overview: Core Discussion Questions and Texts for Cycles

Unit	Guiding questions	Text
<i>Unit 1: Nature</i>		
Cycle 1	How do plants and animals depend on one another?	<i>The Wolves Are Back</i> by Jean Craighead George (2008)
Cycle 2	What are the different ways that humans can impact nature?	<i>Species Revival</i> by the CLAVES Research Team (2015b)
<i>Unit 2: Rights</i>		
Cycle 1	How can people stand up for their rights?	<i>Ivan: The Remarkable True Story of the Shopping Mall Gorilla</i> by Katherine Applegate (2014)
Cycle 2	What can people do to change things that are unfair?	<i>¡Sí, Se Puede! / Yes, We Can! Janitor Strike in L.A.</i> by Diana Cohn (2002)
<i>Unit 3: Immigration</i>		
Cycle 1	How does immigration affect communication and culture?	<i>Home at Last</i> by Susan Middleton Elya (2002)
Cycle 2	Should schools make sure that students learn more languages than just English?	<i>Bilingual Education</i> by the CLAVES Research Team, 2015a)

FIGURE 1
Sample PowerPoint Presentation Highlighting Bridging Concepts for Exploring the Word *Populate*



Note. The color figure can be viewed in the online version of this article at <http://ila.onlinelibrary.wiley.com>.

what they meant (or might mean). Students also competed against one another to be the first to identify a correct irregular past tense of a word.

Day 5: Small-Group Discussions

At the end of each text cycle, teachers facilitated a discussion about the text that had been read. The discussion was

grounded in a big yes/no question that required the students to take a position on the question and argue in support of that position using the text (a critical-analytic perspective) and/or lived experiences (an aesthetic perspective) as evidence, while listening carefully to others' opinions (see Chinn, Anderson, & Waggoner, 2001). In these discussions, the teacher steps into and out of the conversation as needed to keep students focused on the big question and the text, prompts deeper thinking about a particular claim, offers alternative viewpoints to consider, and ensures that students follow the expectations of participation.

These discussions had several ground rules: encouragement to speak freely without teacher nomination, respect for whomever is speaking, participation from all group members (not a requirement), and consideration of all ideas. In this framework, students learn to take and yield the floor, speak clearly and listen carefully, express reasons and cite evidence to justify positions, issue challenges, and respond to the challenges of others (see Parra et al., 2016). During the introductory unit of the intervention, students were taught the ground rules and given an opportunity to practice the discussion style with an easy-to-access question (i.e., "Should students be able to use cell phones in class?"). As described later, monthly teacher working groups also focused on how to facilitate and improve these conversations over the course of the intervention. Discussions of this type, done well, have been shown to deepen text comprehension (Chinn et al., 2001; Reznitskaya et al., 2009; I.A.G. Wilkinson, Soter, & Murphy, 2010) and promote academic language, notably argumentation skills (August, Branum-Martin, Cardenas-Hagan, & Francis, 2009; Vaughn et al., 2009). Small-group discussions may be particularly beneficial to bilingual learners and other students who benefit from oral language practice (Baker et al., 2014; Zhang, Anderson, & Nguyen-Jahiel, 2013).

Writing

In a unit, the two text-based language and comprehension cycles were followed by a three-day writing cycle. This cycle took the form of an abbreviated writers' workshop in which students were prompted to write to an authority figure (e.g., congressional representative, school principal) or for a public audience (e.g., school newspaper) on a topic relevant to the big questions from the day 5 discussions. These prompts required the use of an argumentative register to take a stance on a position and to provide evidence for that position (for an overview of the systemic functionality of the argumentation genre, see Brisk, 2015). Given the intricate relations among academic language, reading, and writing (Fitzgerald & Shanahan, 2000; Newell, Beach, Smith, & VanDerHeide, 2011), we reasoned that having students write about the ideas they had encountered and

discussed in text would help reinforce, via written practice, the text-based linguistic instruction provided in cycles 1 and 2.

Leveraging Bilingualism

From a bilingual literacy perspective, lessons were designed to leverage instructional principles of bridging (Beeman & Urow, 2012). Students' bilingualism was tapped to support their understandings of academic vocabulary, morphology, and syntax (Goodwin & Jiménez, 2016; Hernández et al., 2016). With academic vocabulary instruction, this often took the form of using cognates to promote better understandings of word meanings (Hernández et al., 2016) or having students brainstorm translations of words and sentences from one language to another (Goodwin & Jiménez, 2016; Puzio, Keyes, Cole, & Jiménez, 2013). Morphologically, students were led to make insights into the morphological commonalities that exist between English and Portuguese or Spanish; for example, *-ity*, *-idade*, and *-idad* are synonymous bound morphemes, and when they are added to a root word that is also a cognate (e.g., *real*), the new words (e.g., *reality*, *realidade*, *realidad*) are translations of one another that all conform to the same morphological convention. Syntax instruction also leveraged bilingualism, such as when students worked through differential operations of syntax as a function of the language (e.g., "the excellent student" = "*el estudiante excelente*" in Spanish and "*o aluno excelente*" in Portuguese; Parra, 2018). Instruction in the current intervention took advantage of the fact that students were bilingual speakers of English and Spanish or Portuguese, and thus instruction was designed to draw students' attention to their two languages to elucidate more clearly the nature of the linguistic constructs.

Instructional Time

In all, lessons were designed to last for 30 minutes. However, the amount of instruction included in each lesson often resulted in lessons that lasted for 40 minutes. Thus, teachers regularly ended the 30-minute lessons before finishing and then completed those lessons at the subsequent small-group meeting. Thus, in some cases, it took teachers two sessions to complete one lesson, making the number of sessions greater than the number of lessons in some cases.

Quasi-Experimental Design

Parents of all bilingual fourth- and fifth-grade students at participating schools, whose children were at WIDA levels 3–6, were contacted. Parental consent and student assent for participation were obtained for 111 students across 19 homerooms in the Northeast schools and 128 students across 13 homerooms in the Mid-Atlantic schools. In six

of the eight participating schools, students were assigned to condition (CLAVES or business-as-usual control) individually within homeroom, stratified by WIDA level. Due to departmentalization and scheduling constraints, students in two of the four Mid-Atlantic schools were block-assigned to condition by homeroom. In three of the four Northeast schools and three of the four Mid-Atlantic schools, students who received instruction from specialist teachers were pulled from across classrooms to form intact intervention and control groups. For example, at school 1 in the Northeast, one classroom had seven students, and another classroom had four. Students were assigned to condition within classroom so two groups, consisting of five intervention and six business-as-usual control students, were formed and then pulled out for instruction at different times by the same specialist.

At the Northeast schools, there were 13 intervention groups and 13 control groups, and 12 teachers delivered CLAVES and business-as-usual instruction to the groups. One of the specialist teachers taught two groups. In the Mid-Atlantic schools, there were 10 intervention groups and 11 control groups, and 10 teachers delivered CLAVES and business-as-usual instruction to the groups. One of the specialist teachers taught two groups. Within and across the Northeast and Mid-Atlantic sites, ELA instruction was consistent across classrooms even though the specific curricula implemented across sites differed. At both sites, ELA instruction included reading and writers' workshop with explicit comprehension instruction and close reading, as well as guided and independent reading.

As previously noted, all participating teachers were responsible for delivering both intervention and control group instruction. Participating teachers were either classroom teachers or specialists. When the participating teachers were classroom teachers, they delivered the intervention with a small group of students (between four and six per group) in their classrooms during ELA time. During intervention instruction, students in the control group engaged in independent work (see the Business-as-Usual Control subsection for more detail). Before or after intervention instruction, teachers met with control group students and conducted typical guided reading instruction instead of the intervention, during which time intervention students engaged in independent work. Thus, for classroom teacher participants, the intervention took the place of guided reading instruction for roughly three days a week for intervention students, and business as usual for control students consisted of typical guided reading instruction.

When participating teachers were specialists, instruction was pull-out, meaning that teachers met with small groups of four to six students outside of their classrooms. Specialists met with the intervention group and implemented the intervention. When they met with the students in the control group, they used their

typical instructional practices (see the Business-as-Usual Control subsection). Thus, when the participating teachers were specialists, the intervention took the place of the typical pull-out services that students would normally receive from specialists, and business as usual for control students included typical pull-out support described next. All students who were pulled out (i.e., intervention and control) also engaged in ELA instruction with their regular classroom teacher as usual during non-pull-out time.

Business-as-Usual Control

As previously stated, all participating teachers (classroom teachers and specialists) were responsible for both intervention and control group instruction. Here we describe the nature of the instruction that control students received from the participating teachers. As mentioned, the CLAVES intervention was implemented instead of typical small-group ELA instruction or supplemental instruction for ELs. We conducted interviews with, and observations of, all participating teachers to describe the nature of instruction delivered in the control conditions.

Classroom Teacher Control Instruction

Observations and interviews revealed that teachers used typical guided reading practices with students in the control group and other students in the class. Guided reading included meeting with small groups of students working on the same levelled text. Teachers typically previewed the text with students, which included using chapter titles or headings to think about what the text is about. Generally, teachers asked students some guiding questions and had students read silently or aloud. Next, teachers had students discuss the answers to the guiding questions using comprehension strategies that had been introduced in class (e.g., asking and answering questions, monitoring, summarizing). This instruction was similar to the guided reading instruction in the intervention. Across teachers, interviews and observations revealed limited evidence of vocabulary instruction beyond definitions. Instruction that focused on syntax, morphology, and student-to-student discussion of text was rarer still, and our observations revealed no text-based writing activities. Cross-linguistic connections were occasionally observed among classroom teachers in the Northeast schools but not in the Mid-Atlantic schools.

Specialist Teacher Control Instruction

Control students, like the intervention students, were pulled from class during independent reading time, and students in the control condition received typical support instruction from these specialists. Observations and interviews revealed that specialists also used the guided reading approach discussed previously, with more focused

attention to vocabulary and comprehension strategies. Typically, specialists dedicated more time to focusing on vocabulary than in regular guided reading instruction, but as with regular classroom instruction, vocabulary activities focused mainly on defining words, not on comparing and contrasting words or thinking about how words are related or used across multiple contexts. Specialists spent more time on supporting comprehension and comprehension strategy use than in typical classroom instruction, but this was mainly in a traditional initiate (i.e., the teacher asks a question)–response (i.e., a student responds to the question)–evaluate (i.e., the teacher evaluates the student’s response) format rather than student-to-student interactions. As with the classroom teachers, some cross-linguistic connections were observed among specialists in the Northeast schools but not in the Mid-Atlantic schools.

Preparing and Supporting Teachers

We prepared implementing teachers prior to, and supported them during, their implementation of the intervention. Implementing teachers attended a full-day summer workshop before the beginning of the 2016–2017 school year. The workshop included information about the underlying theories of CLAVES, an overview of the intervention, a review of materials and resources, detailed analyses of the structure of the lesson plans, supported with videos of implementation taken from the two-year development phase, and live demonstrations of lesson implementation. We provided teachers with online and hard copy access to all program materials.

Throughout the duration of the intervention, we ran teacher working groups designed to support participating teachers while they were delivering the intervention. Each of the eight schools had one or two doctoral student research assistants (RAs) who were assigned as liaisons and charged with running and leading the teacher working groups, during which the RAs and principal investigators met with teachers to troubleshoot challenges to implementation and to receive feedback from teachers to inform future revisions of the intervention. Each working group met once per month for 60–90 minutes during the implementation period (November 2016–May 2017) and was grounded in challenges specific to a given set of teachers in a given school. These challenges were identified based on ongoing fidelity of implementation video recordings (see the Fidelity of Implementation subsection), collected by the RAs, and thus tailored specifically to the needs of the teachers at participating schools. Across teacher working groups, RAs kept detailed field notes of areas of implementation challenges, suggestions offered and attempted by teachers, and teacher recommendations for future modifications to the intervention. Although not the focus of the current study, some common topics that arose across working groups were teaching about morphology and

syntax while promoting student talk, questions regarding rules of grammar and syntax, and challenges in facilitating small-group discussions.

As we have noted, CLAVES was partly designed to encourage teachers to have bilingual students draw on their own funds of linguistic knowledge throughout instruction. Rather than asking students to check their language skills at the classroom door to create “subjects who have mastered the empirical linguistic practices deemed appropriate for a school context” (Flores & Rosa, 2015, p. 157), we endeavored, through the summer training, the lesson plan development, and the monthly teacher working groups, to promote instructional approaches that allowed students flexible opportunities to use a more complete linguistic repertoire (García & Wei, 2014) in exploring understandings and awareness of language as it occurs in school texts.

Finally, because implementing teachers also worked with the control group students, there may have been contamination in the control group. However, such contamination did not negate the added infusion of the intervention’s instruction among the treatment group. Further, observations suggested that tenets of instruction that were addressed in CLAVES and the teacher working groups did not appear to bleed into instruction outside the intervention.

Fidelity of Implementation

To measure fidelity of implementation, we developed a low-inference observation tool that captured whether implementing teachers addressed various steps in the lesson plans. RAs viewed video-recorded lessons and documented whether teachers implemented each component in a given lesson plan. Because the intervention was not fully scripted, however, full credit for implementing a particular lesson plan component did not require teachers to complete every single step of the component. For example, on the first day in each lesson cycle, teachers were asked to “preview vocabulary concepts.” In the lesson plan, teachers were guided to use a PowerPoint presentation to introduce specific target words. Using the PowerPoint presentation, if teachers previewed the target word using at least two of the suggested methods for introducing words (e.g., “1. say the word; 2. discuss the word parts; 3. review the Spanish/Portuguese cognate if applicable; 4. preview the word in the text and ask students to explain what it means”), then the teacher received full credit for implementing the component. For another example, when teachers guided students to find morphological or syntactical features in text, they could ask students to do this on their own or work in pairs. In training, we had worked with teachers to implement each component by choosing from the steps provided to ensure that teachers had some choices over how to implement instruction to fit with their teaching style

and to meet their students' needs. For every lesson plan, each implementation step was scored on a Likert-type scale, where 0 = *not implemented* (no lesson components observed), 1 = *partially implemented* (one lesson component observed), and 2 = *fully implemented* (two or more lesson components observed). To calculate fidelity, we divided the total score earned by the total possible score to represent the fidelity percentage obtained for each observed lesson. We averaged these scores across all observed lessons for each teacher.

At each site, two RAs were trained to evaluate fidelity. Training included reviewing lesson plans and videos with the principal investigators and coming to consensus on what constituted completion of each lesson component. Then, the two RAs separately coded 10% of fidelity observations to establish inter-rater reliability. Cohen's kappa was .90 across the two raters at each site. Once reliability was established, RAs split the coding responsibilities of the remaining videos. In the Mid-Atlantic schools, there were at least two videos for each teacher, with an average of 4.5 across teachers. Fidelity scores ranged from 75% to 100%. The mean across teachers was 87%. In the Northeast schools, there were two videos for one teacher, three videos for 10 teachers, and four videos for one other teacher, with an average of 3.0 across teachers. Fidelity scores ranged from 75.3% to 100%. The mean across teachers for the Northeast schools was 88%. Differences in fidelity between the classroom teachers ($M = 85.72$, $SD = 7.4$) and the specialists ($M = 89.64$, $SD = 7.9$) were not significant, $F(1, 44) = 1.44$, $p = .24$.

Measures

We administered parallel measures of reading comprehension in both the fall and spring of the intervention year, where the fall measurement occasion served as an autoregressor in assessing the effect of condition on spring reading. In assessing the effect of condition on academic language, we were unable to use a fall autoregressor due to the fact that the measure of academic language we used had only one form at the time of the study. Thus, we administered measures of expressive vocabulary, morphology, and syntax to serve as covariates in assessing the effect of condition on spring academic language.

Fall and spring measures were administered by undergraduate and graduate RAs. All RAs were trained before administering and scoring assessments. Raw scores were used for analyses unless otherwise indicated.

Expressive Vocabulary: Fall Only

The Woodcock–Muñoz Language Survey–Revised (Woodcock, Muñoz-Sandoval, Ruef, & Alvarado, 2005) Picture Vocabulary subtest is designed to assess aspects of oral language, including expressive vocabulary knowledge or semantic breadth. In this individually

administered assessment, a student is shown pictured items ordered by increasing difficulty (e.g., balloons, blimp, weather vane) and prompted to verbally identify each picture. The test authors reported Cronbach's alphas for students 8 and 11 years old as .90 and .92, respectively (Woodcock et al., 2005).

Morphology: Fall Only

The Extract the Base measure (Anglin, 1993; August, Kenyon, Malabonga, Louguit, & Caglarcan, 2001; Carlisle, 1988) assesses derivational morphology. Students are first read a target word (e.g., *happiness*) and a corresponding sentence (e.g., "My pet dog makes me very ____"). Then, they are asked to extract the base from a derived word (e.g., *happy* from *happiness*) and write their responses on the assessment form. Each item was rated on a scale of 0–2. Students received 0 for an incorrect response, 1 for a misspelled but phonologically correct response (e.g., *happi* instead of *happy*), and 2 for a correct response. August and colleagues (2001) reported Rasch-based reliability as .98.

Syntax: Fall Only

The Comprehensive Assessment of Spoken Language (Carrow-Woolfolk, 1999) Grammaticality Judgment subtest is individually administered and measures students' ability to recognize and correct grammatical errors (e.g., noun–verb agreement, irregular forms, pronouns). First, test administrators orally present sentences with or without grammatical errors (e.g., "The baby is crying"; "Her goes into the house") to students. Then, students are asked to say if the sentence is or is not grammatically correct. If it is not correct, students are asked to correct the sentence by changing only one word without changing the meaning of the sentence. Each item was scored on a scale of 0–2. For grammatically correct sentences, students received a maximum of 1 point for identifying the sentences as correct. For grammatically incorrect sentences, students received 1 point for identifying them as incorrect and another point for fixing them correctly. According to the author, the test–retest coefficient is .91 for ages 8–11. The correlation between the Grammaticality Judgment subtest and other related assessments (i.e., Oral and Written Language Scales second edition, Peabody Picture Vocabulary Test third edition, Kaufman Brief Intelligence Test) range from .52 to .74 for ages 7–11.

Reading Comprehension: Fall and Spring

The Gates–MacGinitie Reading Tests fourth edition (MacGinitie, MacGinitie, Maria, & Dreyer, 2002) is a norm-referenced group assessment that includes a subtest designed to evaluate reading comprehension. This untimed comprehension subtest requires students to read a passage independently and respond to multiple-choice questions about the text. Most students did not take longer

than 45 minutes to complete this assessment. We administered form S in the fall and form T in the spring. The publisher reported Kuder–Richardson formula 20 reliability coefficients of .94–.96, as well as robust validity evidence. The test–retest reliability coefficient was reported as .81 for fourth grade (MacGinitie et al., 2002).

Academic Language: Spring Only

We used the CALS assessment (Uccelli, Barr, et al., 2015; Uccelli, Phillips Galloway, et al., 2015) to measure academic language. The CALS is a 45-minute standardized measure designed to evaluate a specific set of students’ language skills in grades 4–8. This group-administered assessment is aligned with the theoretical framework for academic language described previously and entails eight tasks: connecting ideas, tracking themes, organizing texts, breaking words, comprehending sentences, identifying definitions, epistemic stance, and metalanguage. These tasks involve multiple-choice, sorting, matching, and short written response items. Although the intervention did not teach to the test, the components of academic language measured in the CALS were addressed either directly or indirectly in the intervention. This is illustrated in Table 5, which serves two purposes. First, it provides information about the skills assessed on the CALS measure, as well as directions and examples of the items that reflect the skill. Second, we link these discrete skills to examples from instruction in the intervention, which aligns the CALS both theoretically, as described in the introduction, and empirically, relative to how the CALS operationalizes academic language. According to the test authors, reliability evidence is robust (.93 as indexed by coefficient alpha and .90 by split-half reliability).

Analytic Plan

The nature of the nested data structure (i.e., students nested within teachers) arguably necessitated using a multilevel modeling framework (also known as hierarchical linear modeling; HLM) as the analytic approach to the analyses. A two-level model, for example, allows for the grouping of student outcomes within teachers and includes residuals at both student and teacher levels. Consequently, the residual variance can be partitioned into a between-teacher component (the variance of the teacher-level residuals) and a within-teacher component (the variance of the student-level residuals; Hox, Moerbeek, & van de Schoot, 2017). The teacher residuals, often referred to as teacher effects, represent latent teacher characteristics that impact student outcomes. These unobserved variables lead to correlation between outcomes for students from the same teacher (Raudenbush & Bryk, 2002).

For the HLM analyses, we used a two-level multilevel model for outcome y (e.g., raw scores on the Gates–

MacGinitie Reading Tests fourth edition) for student i ($i = 1, \dots, n_j$) within teacher j ($j = 1, \dots, J$), which can be expressed as

$$\text{Level 1: } y_{ij} = \beta_{0j} + \beta_{1j}x_{1ij} + \dots + \beta_{pj}x_{pij} + r_{ij} \quad (1)$$

where for the sample of students within each teacher, β_{0j} represents the intercepts and $\beta_{1j}, \dots, \beta_{pj}$ are the coefficients (slopes) when predicting y_{ij} from the collection of p student-level predictors, x_{1ij}, \dots, x_{pij} , which in the current context are pretest scores and a dichotomized variable representing treatment group affiliation ($0 = \text{control}, 1 = \text{treatment}$). The residuals, r_{ij} , are assumed to be normally distributed with mean zero and variance, σ^2 .

At level 2, the collection of intercepts has a single covariate, w_{1j} , which in the following analyses will be a dichotomous variable denoting the region ($0 = \text{Northeast}, 1 = \text{Mid-Atlantic}$). This region variable was added as a blocking variable to account for variability in region-to-region differences. Note that the remaining slopes at level 1 are fixed at level 2. The level 2 model is specified as

$$\begin{aligned} \text{Level 2: } \beta_{0j} &= \gamma_{00} + \gamma_{01}w_{1j} + u_{0j} \\ \beta_{1j} &= \gamma_{10} \\ &\vdots \\ \beta_{pj} &= \gamma_{p0} \end{aligned} \quad (2)$$

The coefficients at level 2 are the model’s fixed effects, $\gamma_{00}, \gamma_{01}, \gamma_{10}$, and so forth. These parameters form the structural part of the multilevel model and are used to describe the general patterns in the data based on teacher-specific intercepts and slopes. The level 2 random effect for the intercepts, u_{0j} , represent teacher-specific deviations from these patterns and are assumed to be normally distributed with mean zero and variance, ϕ .

The analytic plan consisted of three primary activities. First, we computed intraclass correlation coefficients (ICCs) for each of four outcomes using an unconditional multilevel model (i.e., one with no predictors at level 1 or level 2) to ascertain whether there was reliable between-teacher variation. For the unconditional two-level model, the ICC could be computed as the ratio of teacher-level variance to total variance:

$$\text{ICC} = \frac{\phi}{\phi + \sigma^2}$$

Second, two-level multilevel models were fit to the academic language and reading comprehension outcome variables and aligned with the multilevel model presented in equations 1 and 2. We checked distributional assumptions of the model, as well as a series of diagnostics, to identify potential outliers and/or influential cases.

Although the nested data structure traditionally calls for a multilevel modeling framework, the small number of clustering units in the study made HLM difficult to implement because the assumptions of the models are tenuous and difficult to check with so few clusters (e.g.,

TABLE 5
Description of the Discrete Core Academic Language Skills (CALs) Tasks, Directions and Examples From the Assessment, and Alignment Between CALs Tasks and Intervention Activities

CALS skill	Directions and example	Example intervention activity
Connecting ideas: This task assesses skill in understanding school-relevant words that connect ideas.	<p><i>Directions:</i> “First, read the sentences to yourself. Then, circle the answer that best completes the sentence. Some questions have pictures, others do not.”</p> <p><i>Example:</i> “Kate wears sneakers, ___ Jim wears sandals.</p> <p>a.but b.then c.so d.also”</p>	Students were provided with explicit instruction on coordinating and subordinating conjunctions and guided to use these in mapping, combining, and segmenting sentences.
Tracking themes: This task assesses skill in tracking referents through a text.	<p><i>Directions:</i> “First, read the sentences. Then, pay attention to the underlined words in the sentence. Finally, circle the option that refers to the same person, thing or event as the underlined words in the sentence.”</p> <p><i>Example:</i> “Read the passage, ‘Water heats up more slowly than land. It also cools down more slowly than land.’ In the sentence above, ‘It’ refers to:</p> <p>a.water b.heat c.land”</p>	Students were provided with explicit instruction in subject–verb–object agreement and in pronominal reference. Students engaged in sentence-mapping and constructing activities to reinforce these skills.
Organizing texts: This task assesses skill in argumentative text organization.	<p><i>Directions:</i> “Can you help Jim fix his essays? Jim needs your help! After his computer crashed, all of the sentences in three of his essays were out of order. Can you reorganize his essays?”</p> <p><i>Example:</i> “Write a number from 1 to 4 next to each sentence.</p> <ul style="list-style-type: none"> • One reason is that recess is healthy for kids. • That’s why I think it is important to try to save recess. • Some schools are getting rid of recess, but I think recess is good for students. • For example, kids can do a lot of exercise during recess.” 	Students were encouraged to debate big questions using evidence from the texts that they read or viewed. Then, students were asked to write an argumentative essay in response to a prompt related to these big questions. As students were preparing to engage in discussion and write their essays, they were taught to use sentence starters and graphic organizers that guided their argumentative text organization.
Breaking words: This task assesses skill in breaking down complex words.	<p><i>Directions:</i> “Maria is having trouble remembering words. Can you help her to think of the correct word form? Change the word into a different form to complete the sentence.”</p> <p><i>Example:</i> “(driver) Children are too young to ___.”</p>	Students were explicitly taught prefixes and suffixes and guided to break apart and put together meaningful words using different word parts. Students were also taught to derive words from root words by changing the prefixes and suffixes during lessons on morphology.
Comprehending sentences: This task assesses skill in understanding complex sentence structures.	<p><i>Directions:</i> “I am going to say something and I want you to select the picture that goes with what I say.”</p>	During reading of text, students were guided to infer meaning from complex sentences. During lessons on syntax, students were taught to map and develop complex sentences of their own.

(continued)

McNeish & Stapleton, 2016a, 2016b). As an alternative to HLM that also accounts for the clustering in the data, we also implemented a design-based method, cluster-robust

standard errors (CRSEs; McNeish, 2014), which does not require the specification or estimation of random effects or their covariance structure and thus does not require

TABLE 5
Description of the Discrete Core Academic Language Skills (CALs) Tasks, Directions and Examples From the Assessment, and Alignment Between CALs Tasks and Intervention Activities (continued)

CALS skill	Directions and example	Example intervention activity
Identifying definitions: This task assesses skill in identifying a more academic versus a more colloquial register.	<p><i>Directions:</i> “Maria is writing a dictionary for adults. She has written many definitions, but she does not know which definition to include in a dictionary for adults. She needs your help! First, listen to each definition that Maria has written. Then, for each definition circle ‘children’ if it seems that it was written for children, or ‘adults’ if it seems that it was written for adults. Finally, from the three definitions choose the option that is best for a dictionary for adults.”</p> <p><i>Example:</i> “An <i>umbrella</i> is what you use so you do not get wet in the rain; it has cloth over a frame that you can fold. This was written for: children or adults”</p>	During the focus on academic vocabulary, students were explicitly taught definitions of words and encouraged to define vocabulary in their own words. In this way, students had exposure to both dictionary (adult-like) definitions of words and colloquial (child-friendly) definitions of words that they and their peers developed.
Sure or unsure?: This task assesses skill in interpreting markers that signal a writer’s level of certainty about a claim.	<p><i>Directions:</i> “First, read the situation below. Then, read what each person says. Finally, decide how sure each person is and check the appropriate box.”</p> <p><i>Example:</i> “A group of friends is trying to figure out if the teacher is going to give a quiz. Some friends are very sure there will be a quiz and some friends have doubts. There will be a quiz tomorrow. How sure is this friend? Yes, Maybe Yes, Maybe No, No”</p>	During guided reading, students were guided to infer meaning from text using context clues. During lessons on syntax, students were taught how to use adverbs (e.g., <i>however, although</i>) in developing complex sentences.
Understanding responses: This task assesses skill in understanding words that label or qualify language or thinking moves.	<p><i>Directions:</i> “First, listen to the newspaper’s idea. Then, listen to each student’s sentence. Finally, choose the option that best describes what each student is saying or writing.”</p> <p><i>Example:</i> “The newspaper says, ‘Students need recess to play and relax at school.’ Peter says, ‘Yes, at recess we play games like tag, soccer or basketball.’ In the sentence above, Peter is giving...</p> <p>a. excuses b. examples c. definitions d. exaggerations”</p>	Students were guided to use inferencing skills to determine the meaning of text. Additionally, students were taught to use accountable talk during discussions of text, as well as sentence frames and graphic organizers to organize their arguments in writing.

the same assumptions of HLM that can be questionable with smaller samples. A CRSE approach estimates a traditional single-level regression model but then applies a statistical correction to the standard errors to ensure that they mirror the hierarchical nature of the data (Silverman et al., 2014). As with HLM, CRSEs are known to be biased with a small number of clusters (Lu et al., 2007); however, there are a handful of small sample corrections that can be successfully applied with as few as 10 clusters. Lu et al. (2007) suggested using the Kauermann–Carroll correction (Kauermann & Carroll, 2001), which was found to

perform well for cluster sizes less than 10 (for our data, the average cluster size was approximately seven). In the current study, we implemented this correction method using the SAS (version 9.4) GLIMMIX procedure (SAS Institute, 2015), and we present results from both HLM and CRSE approaches in answering our research questions.

We included either an autoregressor (for reading) or, in the case of academic language, a set of covariates (expressive vocabulary, morphology, and syntax) given the lack of multiple forms of the CALs. To these

covariates, along with blocking for region, we added the treatment variable. For each analytic method (HLM and CRSE), we fitted inferential models to the data for the two outcome variables (academic language and reading comprehension). In each model, we used appropriate pre-intervention measures as covariates, and a dichotomously scored variable indicating treatment group membership was added (at level 1 for HLM). All pre-intervention measures were grand mean centered. We added a dichotomously scored region (0 = *Northeast*; 1 = *Mid-Atlantic*) variable (level 2 for HLM) as a blocking variable to account for any site-to-site differences.

We calculated effect sizes of treatment on our outcomes using Hedges’s *g*, which is a traditional means by which to evaluate intervention effects. We calculated this via the adjusted group mean difference divided by the unadjusted pooled within-group standard deviation. As articulated by the What Works Clearinghouse (2017), the calculation was derived from the following equation:

$$g = \frac{\omega\gamma}{\sqrt{\frac{(n_i-1)s_i^2 + (n_c-1)s_c^2}{n_i+n_c-2}}}$$

The second index quantifying the size of effect in the multilevel modeling settings is what Raudenbush and Bryk (2002) and Rights and Sterba (2018) referred to as variance accounted (i.e., *R*²) measures. LaHuis, Hartman, Hakoyama, and Clark (2014) emphasized that “explained variance measures provide a useful summary of the magnitude of effects and may be particularly useful in multilevel studies where unstandardized coefficients are reported often” (p. 446).

Power

An a priori power analysis was conducted as part of the grant application that funded the current work. A multi-level analysis of covariance model used in the current analysis was the archetypal analytic model used for the power analysis. We used Optimal Design software (Raudenbush et al., 2011) to perform the sample size/power calculations. We considered a multisite trial, which we treated as a two-level design with students within blocks—in this case, classrooms. Holding power at a constant level of .80, we determined that the number of classrooms necessary to realize small to large standardized group mean differences (.20–.80) ranged from 97 classrooms for a small effect to nine classrooms necessary for a large effect. This trend—the ability to detect an increasingly small mean difference as clustering units increase—follows general rules found in power computations across a variety of designs and analytic models.

Results

Descriptive statistics are presented in Table 6 for the entire sample and disaggregated by region. One-way analyses of variance indicated no differences between control and intervention at pre-intervention and differences approaching significance for spring academic language, *F*(1, 235) = 3.18, *p* = .076. Additionally, there were significant differences across covariates and outcome measures as a function of region (in favor of Northeast) for morphology, syntax, spring reading, and academic language (all *F*s ≥ 12.87, all *p*s < .001), which necessitated including region as a blocking variable in our models.

The ICC values for teachers within school and for school level are presented in Table 7. Note that the values

TABLE 6
Means (and Standard Deviations) for Fall Covariates and Fall and Spring Outcomes, Disaggregated by Intervention and by Site

	Fall covariates			Spring outcomes		
	Vocabulary	Morphology	Syntax	Reading comprehension	Reading comprehension	Academic language
Overall	30.09 (4.15)	25.93 (9.89)	35.78 (10.11)	19.01 (7.96)	23.16 (8.12)	481.71 (20.60)
<i>Condition</i>						
Control	30.24 (3.97)	25.40 (10.00)	35.96 (10.31)	19.05 (8.50)	22.55 (8.47)	479.31 (21.39)
Intervention	29.95 (4.34)	26.46 (9.79)	35.62 (9.96)	18.97 (7.41)	23.76 (7.77)	481.71 [†] (20.60)
<i>Region</i>						
Northeast	30.00 (4.22)	33.10*** (10.12)	38.71*** (11.14)	19.69 (8.35)	25.39*** (8.02)	488.56** (19.89)
Mid-Atlantic	30.16 (4.11)	20.13 (4.36)	33.22 (8.34)	18.43 (7.59)	21.30 (7.76)	475.83 (19.42)

p* < .01. *p* < .001. [†]*p* < .10.

TABLE 7
Intraclass Correlation Coefficients for Teachers Within Schools and for School Level

Outcome	Intraclass correlation coefficients	
	Teacher (school)	School ^a
Reading comprehension	.231	.026
Academic language	.164	.112

Note. We computed the values using a three-level unconditional model in which students were nested within teachers, and teachers were nested within schools.

indicate a wide range of clustering at the teacher level and less clustering overall at the school level. Because the school-level variance was small overall, we consider the school level as nuisance variability and decided not to model it as part of the multilevel model (i.e., a three-level model with school as the top level of clustering). To guard against potential Type I errors in the two-level analysis for not accounting for school-level variability, robust standard errors were computed (see, e.g., McNeish, 2014) using a sandwich estimator for the standard errors by adding the empirical argument to the PROC MIXED statement. Overall, the ICCs indicate that a two-level analysis is necessary, as values above .05 involve sufficient clustering to warrant the use of multilevel modeling (Hedges & Hedberg, 2007; McCoach & Adelson, 2010).

The results of the analyses for research question 1 (Do bilingual students assigned to CLAVES outperform their counterparts assigned to the business-as-usual control group on standardized measures of academic language and reading comprehension?) are presented in Table 8. After controlling for pre-intervention differences and blocking on site, the unstandardized coefficients for spring reading on treatment for the HLM and CRSE methods were 1.35 and 1.14, respectively ($p = .072$ and $.118$, respectively). This equates to a Hedges's g effect size of 0.166, with treatment explaining an additional 2% of the variance in reading comprehension above and beyond that explained by the pre-intervention autoregressor and region.

The unstandardized coefficients for spring academic language on treatment for the HLM and CRSE methods were 5.07 for both approaches ($p = .023$ and $.002$, respectively). This equates to a Hedges's g effect size of 0.248, with treatment accounting for an additional 3% in the variance of academic language above and beyond the variance explained by pre-intervention covariates and region.

The results of the analyses for research question 2 (Do pre-intervention language proficiency and reading comprehension covariates moderate main effects of CLAVES on standardized measures of academic language and reading comprehension?) are presented in Table 9. To answer the question, we added an interaction term to the main effects multilevel analysis of covariance model and

TABLE 8
Hierarchical Linear Modeling (HLM) and Cluster-Robust Standard Error (CRSE) Results for Assessing Treatment Effects on Spring Reading and Academic Language

Fixed effect	Spring reading comprehension				Spring academic language			
	HLM		CRSE		HLM		CRSE	
	Estimate (SE)	p	Estimate (SE)	p	Estimate (SE)	p	Estimate (SE)	p
Intercept	27.07 (1.67)	<.001	27.44 (1.89)	<.001	479.39 (4.95)	<.001	479.39 (3.80)	<.001
<i>Controls</i>								
Region	-3.00 (1.03)	.007	-3.11 (1.14)	.007	0.09 (2.98)	.976	0.09 (2.25)	.968
Fall reading comprehension	0.67 (0.05)	<.001	0.69 (0.06)	<.001				
Fall syntax					0.38 (0.14)	.005	0.38 (0.14)	.007
Fall vocabulary					1.12 (0.31)	<.001	1.12 (0.32)	<.001
Fall morphology					0.84 (0.16)	<.001	0.84 (0.11)	<.001
Treatment	1.35 ^a (0.75)	.072	1.14 (0.90)	.118	5.07 ^b (2.21)	.023	5.07 (1.58)	.002
<i>Variance components</i>								
Level 1 variance	27.74 (2.75)		31.77 (2.97)		261.72 (25.42)		261.72 (25.42)	
Level 2 variance	4.14 (2.05)				0.00			

^aHedges's g is 0.166 (the within-student variance in Gates–MacGinitie Reading Tests fourth edition raw posttest scores was 60.44). ^bHedges's g is 0.248 (the within-student variance in Core Academic Language Skills posttest scores was 433.38).

TABLE 9
Hierarchical Linear Modeling Results for Assessing Treatment and Treatment-by-Pretest Interaction Effects and on Spring Reading Comprehension and Academic Language

	Spring reading comprehension		Spring academic language	
	Estimate (SE)	<i>p</i>	Estimate (SE)	<i>p</i>
Intercept	27.07 (1.68)	<.001	480.20 (4.99)	<.001
<i>Controls</i>				
Region	-3.00 (1.04)	.007	-0.35 (3.01)	.908
Fall reading comprehension	0.71 (0.06)	<.001		
Fall syntax			0.34 (0.21)	.101
Fall vocabulary			1.42 (0.47)	.003
Fall morphology			1.03 (0.22)	<.001
Treatment	1.34 (0.75)	.074	5.00 (2.20)	.024
<i>Interaction effects</i>				
Fall Reading Comprehension × Treatment	-0.09 (0.09)	.336		
Fall Syntax × Treatment			-0.01 (0.28)	.975
Fall Vocabulary × Treatment			-0.37 (0.61)	.542
Fall Morphology × Treatment			-0.37 (0.26)	.161
<i>Variance components</i>				
Level 1 variance	27.72 (2.76)		260.85 (25.52)	
Level 2 variance	4.21 (2.07)		=0	

analyses previously reported. We constructed the interaction terms as products of the dummy-coded intervention variable and appropriate reading or language pretest measures. Table 9 shows that interaction effects were nonsignificant for the Fall Reading Comprehension × Condition variable and also for the Fall Academic Language (Vocabulary, Morphology, and Syntax) × Condition variables, suggesting that pre-intervention performance was not associated with main effects of condition.

Discussion

Literacy instruction is a complex matter. Teachers in the early grades need to promote the development of

code-based skills (phonological and phonemic awareness, word recognition, and fluency) and language (Kieffer & Vukovic, 2013) across myriad instructional contexts and multiple languages. In the upper elementary and secondary grades, decoding typically becomes more automatic, texts become more linguistically complex, and teachers need to focus more of their instruction on language to keep pace with the comprehension demands of text. This has been shown to be consistent for monolingual and bilingual learners alike (Proctor et al., 2012; Silverman et al., 2015). However, meta-analytic work has shown that bilingual children, relative to monolinguals, are advantaged in thinking about language (i.e., in metalinguistic awareness) because “the experience of acquiring and maintaining two different languages—with different forms and structures—allows bilingual speakers to develop an explicit and articulated understanding of how language works” (Adesope, Lavin, Thompson, & Ungerleider, 2010, p. 209). Teaching bilingual learners, therefore, presents unique instructional opportunities for teachers to center language in their literacy instruction. This was the basic theoretical premise of CLAVES: If language predicts reading comprehension, then instruction about language that is text-based ought to promote both language and reading comprehension. In addition, for bilingual learners, research has suggested that their facilities with two or more languages allow us to broaden our instructional options and integrate students’ full linguistic repertoires (García & Wei, 2014). Given this, the findings presented here have implications for theory and practice and also make us think about what academic language is and how to address it in schools.

The instructional approaches leveraged in CLAVES included a focus on language via explicit instruction in vocabulary, syntax, and morphology, as well as guided reading of text, text-based collaborative discussions, and argumentative writing activities. Instruction also leveraged bilingualism via bridging techniques (Beeman & Urow, 2012), and findings showed effect sizes larger than typically detected in the literacy intervention research (e.g., Hall et al., 2017). The CLAVES intervention showed an effect size of $g = 0.248$ effect on academic language (HLM $p = .023$; CRSE $p = .002$) and an effect size of $g = 0.166$ on reading comprehension (HLM $p = .072$; CRSE $p = .118$). From these effect sizes, it can be understood that participation in CLAVES was associated with a difference of 24.8% and 16.6% of a standard deviation for academic language and reading comprehension, respectively, relative to the control condition, net pre-intervention covariates and blocking on region. Although explained variation was relatively small (i.e., 3% for academic language, 2% for reading comprehension), this was an effect above and beyond controls for pre-intervention covariates and regional differences. Given that the average effect size

on standardized measures from Hall et al.'s (2017) meta-analysis of instruction in the middle grades was $g = 0.01$, and previous research on literacy interventions for upper elementary students suggested that it is difficult to produce gains on standardized assessments (Wanzek et al., 2017), the findings here are noteworthy and have implications for theory and instruction.

Theoretically, this study builds on previous intervention work among bilingual learners that focused more directly on academic vocabulary and reading comprehension (e.g., Carlo et al., 2004; Lesaux et al., 2010, 2014; Proctor et al., 2011; Snow et al., 2009). In designing CLAVES, we realized in developing our lesson plans that we needed an operating theory of language that included vocabulary but also attended more broadly to other important aspects of language, including syntax and morphology. Grounded in the CALS framework (Uccelli, Barr, et al., 2015; Uccelli, Phillips Galloway, et al., 2015), which brings together second-language acquisition and systemic functional linguistics, the CLAVES intervention merged a discrete focus on language components (vocabulary, morphology, and syntax) with holistic expectations of language use (i.e., talking about how language works, small-group discussions, authentic writing tasks). CALS is also grounded in the simple view of reading (Hoover & Gough, 1990) and the lexical quality hypothesis (Perfetti & Stafura, 2014). To this, we added the RAND Reading Study Group (2002) model of reading comprehension, into which instruction can be readily applied. Pulling these together, in the CLAVES instructional context, we focused on academic language and reading comprehension among particular readers (i.e., upper elementary bilingual students), activities (e.g., discussions, writing), and texts (e.g., print, digital) in a particular sociocultural context (i.e., monolingual and bilingual programs). The findings from this study provide empirical evidence for the theoretical orientation of the intervention.

Instructionally, there were some key takeaways that emerged from this study. Of course, the specific texts, languages, and discussion and writing prompts in this study may not be appropriate in all instructional contexts in which bilingual learners and their teachers find themselves. However, we suggest five broad takeaways that undergird the current findings that have bearing on language-based literacy instruction with bilingual learners in the upper elementary grades, irrespective of context:

1. *Build on what we know about effective literacy instruction.* Research has suggested that effective literacy instruction includes selecting meaningful texts for instruction, teaching reading comprehension strategies, and guiding students through text reading (e.g., Kamil et al., 2008). We designed CLAVES to build on this well-established research base by choosing texts that maximize engagement,

encouraging students to use the comprehension strategies they were learning in their ELA instruction, and scaffolding students in reading challenging texts. Instruction for bilingual students should include these features.

2. *Center literacy instruction around big questions.* In CLAVES, thematic units were designed around big questions that were relevant to the texts being read. For example, the Rights unit was structured around two overarching questions: How can people stand up for their rights? and What can people do to change things that are unfair? Texts, discussions, and writing tasks were grounded in contemplating these questions, undergirded by evidence that these types of broad questions can serve to structure students' engagement in ways that promote understanding and development of academic vocabulary (Ma et al., 2017).
3. *Focus broadly on language in literacy instruction.* The constructs that we focused on in CLAVES were academic vocabulary, morphology, and syntax. We started by establishing a foundation of academic vocabulary before moving into morphology and syntax. Importantly, these constructs are not orthogonal; that is, words and grammatical structures are related to each other. Attending to each of these dimensions of language promotes a deeper treatment of vocabulary and may improve lexical representations (Perfetti & Stafura, 2014). A focus on these constructs also requires students and teachers to think about how language works, and thus serves the development of metalinguistic awareness, which has been shown to be an asset for bilingual learners (Adesope et al., 2010).
4. *Encourage student talk.* CLAVES lesson plans were designed to promote student talk, including talk about texts, talk about the targeted language components (vocabulary, morphology, and syntax), and talk about big questions related to the texts. Further, we worked with teachers to underscore that student talk need not conform to rigid expectations of academic language. The goal was to use a fuller linguistic repertoire (García & Wei, 2014) to grapple with big concepts and ideas. Anecdotally, teachers expressed that this was one of the most challenging aspects of teaching CLAVES. However, we contend that releasing students to use all of their languages serves both linguistic insights and language development.
5. *Write to extend talk.* We did not center writing in the design of CLAVES; however, the writing requirements were argumentative and aligned with the small-group discussions that ended each

instructional cycle. Some recent evidence shows that student talk in small groups increases ideation and length of argumentative writing (Wagner, Parra, & Proctor, 2017). Importantly, talk and writing are both expressive language skills and are far less privileged in typical classrooms than their receptive counterparts (i.e., reading, listening; Silverman et al., 2014). Given the focus on language in CLAVES, it is also worthwhile to consider research that has explored how reading and writing are related to each other via language (e.g., Fitzgerald & Shanahan, 2000).

Critical Perspectives on Academic Language

Academic language has been a point of clear contention in literacy studies. Although the CALS perspective on academic language merges second-language acquisition, systemic functional linguistics, and reading comprehension, it does not integrate more recent theoretical perspectives from the field of applied linguistics. Poza (2015), for example, cautioned against positioning academic language as a means unto itself. Educators and researchers alike should take care to not cast academic language “en clara distinción e implícita elevación sobre los vernáculos cotidianos, requiriendo enseñanza explícita” (in clear distinction and implicit elevation over daily vernaculars, requiring explicit instruction; p. 232). Flores (2015) further argued that teaching and researching academic language should have as its “ultimate goal...to develop a new conceptualization of language that is situated within a larger critique of racial inequalities that current conceptualizations of academic language normalize” (para. 7). Ultimately, being mindful about academic language as key to both success and oppression in schools is critical. Educators who work with multilingual populations can and should recognize the nuanced and powerful ways in which languages can be used to marginalize, racialize, and silence students. In our work, we wrestled with these issues in designing units and themes that focused on important topics related to social justice (e.g., rights, immigration, language, the environment), asking questions that were meant to elicit feelings and discussions about these topics. Although not a tested outcome of our work, throughout development and implementation, our research teams endeavored to be critically aware of how issues of power manifest throughout literacy instruction both in terms of the content we teach and the ways in which we teach.

Finally, as a matter of positionality, the two lead authors of this study, who were the principal investigators on this project, are both white, native English-speaking literacy researchers working with predominantly white ($n = 16$; 72%) and variably bilingual teachers (58% in Northeast schools, 20% in Mid-Atlantic schools) to implement an intervention

intended for primarily black and brown bilingual speakers of English and Spanish or Portuguese. The issues of power inherent in these relationships are worthy of continued acknowledgment and scrutiny. Taking issues of language, race, and power into consideration, schools must strive to become places where students and teachers speak multiple languages and operate in various registers. Our work as educators and researchers can recognize and make audible the remarkable heteroglossia in our schools in service of a more humanizing pedagogy (Bartolomé, 1994; Ladson-Billings & Tate, 1995; Paris, 2012) in which teachers and researchers work across paradigms to “critically evaluate their own beliefs and engage students in critical dialogue that problematizes reality (Bartolomé, 1994; Huerta, 2011; Schugurensky, 2011; Strobel & Tillberg-Webb, 2006)” (Salazar, 2013, pp. 138–139).

Future Research

There are different research directions that could be taken with this intervention. First, we might seek to test which of the multiple components (language skills, discussion, and writing) were more or less effective in promoting outcomes. Counterfactuals that include morphology only, discussion only, or other systematic and theoretically driven combinations of the various skills involved in the intervention could be manipulated and experimentally tested to yield important insights into high-leverage instructional approaches that support bilingual learners’ academic language development. Second, the findings here are noteworthy in that effect sizes are larger than typically reported for standardized measures, but quasi-experimental and randomized control studies can only tell us so much about what works, and typically such studies struggle to explain or explore for whom and why. Explanatory research with data from the current study, or from other instructional contexts, might serve to contextualize the findings here. Specifically, discourse analyses of fidelity data from this study might explore the nature of teacher and student talk in these small groups and the degrees to which teachers were able to leverage interactions to promote student talk that suggests metalinguistic awareness. Data from this study might also be triangulated with other qualitative or mixed-methods literacy research exploring the nature of small-group discussions and the roles that students play in them (see Farnsworth, 2012; Proctor & Bacon, in press). Merging different empirical perspectives on common topics will push the literacy field forward in synthesizing research findings with practical application. If we are going to be asking hard questions not only about what works but also for whom and why, the field will need to consider “critical changes in the types of research paradigms that are seen as worthy of funding and as yielding significant contributions to the knowledge base about reading and literacy development” (Alexander & Fox, 2019, p. 54).

Limitations

There are some clear limitations in the present study. First, the sample was relatively small, and the design was not optimal. This may partly explain why interaction effects between pre-intervention performance and condition were null. Future research should be conducted on a larger scale so there is greater power to detect effects with randomized control designs and to explore how effects might differ as a function of different reader characteristics. Second, the intervention was not fully implemented in that many teachers were unable to complete the intervention due to scheduling conflicts. Working with schools and administrators to determine how to effectively embed intervention into the regular fabric of the school day and manage competing priorities is important to ensure that students who could benefit from such intervention are able to participate in it consistently. Notably, we implemented this intervention in typical school contexts, which are complicated. Although this led to a more convoluted design than we would have liked, it also adds to the ecological validity of the results. Conducting research in such complex contexts is important to try to bridge research and practice and leads to other questions that are important to consider, such as these: (a) Who are the optimal implementers of such an intervention (i.e., classroom teachers or specialists)? (b) How does alignment between the ELA curriculum and the intervention impact implementation and effectiveness? (c) What contextual factors need to be in place to ensure the feasibility of the intervention? It is a limitation of the current study that these questions cannot be presently answered. However, given that CLAVES showed promising effects of a language-based reading intervention on standardized indicators of academic language and reading comprehension, further research to replicate and extend these findings is warranted.

NOTES

The first two authors contributed equally to this manuscript and share first authorship. The research reported here was supported by funding from the Institute of Education Sciences, U.S. Department of Education, through a grant (R305A140114) awarded to C. Patrick Proctor (principal investigator), Rebecca D. Silverman, and Jeffrey R. Harring (co-investigators). The opinions expressed are those of the authors and do not represent views of the Institute of Education Sciences or the U.S. Department of Education. The authors acknowledge the contributions of the teachers and students at the participating schools, as well as Christina Budde, Natalia Guzmán, Yalda Kaveh, Yewon Lee, Meredith Moore, Marcela Ossa Parra, Joelle Pedersen, and Shawn Savage.

REFERENCES

Adesope, O.O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Review of Educational Research, 80*(2), 207–245. <https://doi.org/10.3102/0034654310368803>

Alexander, P.A., & Fox, E. (2019). Reading research and practice over the decades: A historical analysis. In D.E. Alvermann, N.J. Unrau, M. Sailors, & R.B. Ruddell (Eds.), *Theoretical models and processes of literacy* (7th ed., pp. 35–64). New York, NY: Routledge.

Anglin, J.M. (1993). Vocabulary development: A morphological analysis. *Monographs of the Society for Research in Child Development, 58*(10). <https://doi.org/10.2307/1166112>

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. <https://doi.org/10.1080/19345740903217623>

August, D., Kenyon, D., Malabonga, V., Louguit, M., & Caglarcan, S. (2001). *Extract the Base Test*. Washington, DC: Center for Applied Linguistics.

Bailey, A.L. (2007). *The language demands of school: Putting academic English to the test*. New Haven, CT: Yale University Press.

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, Institute of Education Sciences, U.S. Department of Education.

Bartolomé, L. (1994). Beyond the methods fetish: Toward a humanizing pedagogy. *Harvard Educational Review, 64*(2), 173–195. <https://doi.org/10.17763/haer.64.2.58q5m5744t325730>

Baumann, J.F., Edwards, E., Boland, E.M., Olejnik, S., & Kameenui, E.J. (2003). Vocabulary tricks: Effects of instruction in morphology and context on fifth-grade students' ability to derive and infer word meanings. *American Educational Research Journal, 40*(2), 447–494. <https://doi.org/10.3102/00028312040002447>

Bear, D.R., Templeton, S., & Invernizzi, M. (2008). *Words their way: Word study for phonics, vocabulary, and spelling instruction*. Englewood Cliffs, NJ: Prentice Hall.

Beeman, K., & Urow, C. (2012). *Teaching for biliteracy: Strengthening bridges between languages*. Philadelphia, PA: Caslon.

Brisk, M.E. (2015). *Engaging students in academic literacies: Genre-based pedagogies for K–5 classrooms*. New York, NY: Routledge.

Brisk, M.E., & Zhang-Wu, Q. (2017). Academic language in K–12 contexts. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (Vol. 3, pp. 82–100). New York, NY: Routledge.

Calderón, M., Slavin, R., & Sánchez, M. (2011). Effective instruction for English learners. *The Future of Children, 21*(1), 103–127. <https://doi.org/10.1353/foc.2011.0007>

Carlisle, J.F. (1988). Knowledge of derivational morphology and spelling ability in fourth, sixth, and eighth graders. *Applied Psycholinguistics, 9*(3), 247–266. <https://doi.org/10.1017/S0142716400007839>

Carlo, M.S., August, D., McLaughlin, B., Snow, C.E., Dressler, C., Lippman, D.N., ... White, C.E. (2004). Closing the gap: Addressing the vocabulary needs of English-language learners in bilingual and mainstream classrooms. *Reading Research Quarterly, 39*(2), 188–215. <https://doi.org/10.1598/rrq.39.2.3>

Carrow-Woolfolk, E. (1999). *The Comprehensive Assessment of Spoken Language*. San Antonio, TX: Pearson.

Chinn, C., Anderson, R.C., & Waggoner, M.A. (2001). Patterns of discourse in two kinds of literature discussion. *Reading Research Quarterly, 36*(4), 378–411. <https://doi.org/10.1598/RRQ.36.4.3>

Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research, 49*(2), 222–251. <https://doi.org/10.3102/00346543049002222>

- Cummins, J. (1991). Interdependence of first- and second-language proficiency in bilingual children. In E. Bialystok (Ed.), *Language processing in bilingual children* (pp. 70–89). Cambridge, UK: Cambridge University Press.
- Fang, Z., & Schleppegrell, M.J. (2008). *Reading in secondary content areas: A language-based pedagogy*. Ann Arbor: University of Michigan Press.
- Farnsworth, M. (2012). Who's coming to my party? Peer talk as a bridge to oral language proficiency. *Anthropology & Education Quarterly*, 43(3), 253–270. <https://doi.org/10.1111/j.1548-1492.2012.01178.x>
- Fitzgerald, J., & Shanahan, T. (2000). Reading and writing relations and their development. *Educational Psychologist*, 35(1), 39–50. https://doi.org/10.1207/S15326985EP3501_5
- Flores, N. (2015, October 1). Is it time for a moratorium on academic language? [Web log post]. Retrieved from <https://educationallinguistics.wordpress.com/2015/10/01/is-it-time-for-a-moratorium-on-academic-language/>
- Flores, N., & Rosa, J. (2015). Undoing appropriateness: Raciolinguistic ideologies and language diversity in education. *Harvard Educational Review*, 85(2), 149–171. <https://doi.org/10.17763/0017-8055.85.2.149>
- Gándara, P., & Santibañez, L. (2016). The teachers our English language learners need. *Educational Leadership*, 73(5), 32–37.
- Ganske, K. (1999). The developmental spelling analysis: A measure of orthographic knowledge. *Educational Assessment*, 6(1), 41–70. https://doi.org/10.1207/S15326977EA0601_4
- García, O., & Wei, L. (2014). *Translanguaging: Language, bilingualism and education*. New York, NY: Palgrave MacMillan.
- Gebhard, M., Chen, I.A., & Britton, L. (2014). “Miss, nominalization is a nominalization”: English language learners’ use of SFL meta-language and their literacy practices. *Linguistics and Education*, 26, 106–125. <https://doi.org/10.1016/j.linged.2014.01.003>
- Gibbons, P. (2009). *English learners’ academic literacy and thinking*. Portsmouth, NH: Heinemann.
- Goldenberg, C. (2013). Unlocking the research on English learners: What we know—and don’t yet know—about effective instruction. *American Educator*, 37(2): 4–11, 38.
- Goodwin, A.P., & Jiménez, R. (2016). TRANSLATE: New strategic approaches for English learners. *The Reading Teacher*, 69(6), 621–625. <https://doi.org/10.1002/trtr.1400>
- Goodwin, A.P., & Perkins, J. (2015). Word detectives: Morphological instruction that supports academic language. *The Reading Teacher*, 68(7), 510–523. <https://doi.org/10.1002/trtr.1342>
- Grosjean, J. (2010). *Bilingual: Life and reality*. Cambridge, MA: Harvard University Press.
- Hall, C., Roberts, G.J., Cho, E., McCulley, L.V., Carroll, M., & Vaughn, S. (2017). Reading instruction for English learners in the middle grades: A meta-analysis. *Educational Psychology Review*, 29(4), 763–794. <https://doi.org/10.1007/s10648-016-9372-4>
- Halliday, M.A.K., & Matthiessen, C.M.I.M. (2014). *Halliday’s introduction to functional grammar* (4th ed.). New York, NY: Routledge.
- Hedges, L.V., & Hedberg, E.C. (2007). Intraclass correlation values for planning group randomized trials in education. *Educational Evaluation and Policy Analysis*, 29(1), 60–87. <https://doi.org/10.3102/0162373707299706>
- Hernández, A.C., Montelongo, J.A., & Herter, R.J. (2016). Using Spanish-English cognates in Children’s Choices picture books to develop Latino English learners’ linguistic knowledge. *The Reading Teacher*, 70(2), 233–239. <https://doi.org/10.1002/trtr.1511>
- Hoover, W.A., & Gough, P.B. (1990). The simple view of reading. *Reading and Writing*, 2(2), 127–160. <https://doi.org/10.1007/BF00401799>
- Hox, J.J., Moerbeek, H., & van de Schoot, R. (2017). *Multilevel analysis: Techniques and applications* (3rd ed.). New York, NY: Routledge.
- Kamil, M.L., Borman, G.D., Dole, J., Kral, C.C., Salinger, T., & Torgesen, J. (2008). *Improving adolescent literacy: Effective classroom and intervention practices: A practice guide* (NCEE 2008-4027). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- Kauermann, G., & Carroll, R.J. (2001). A note on the efficiency of sandwich covariance matrix estimation. *Journal of the American Statistical Association*, 96(456), 1387–1396. <https://doi.org/10.1198/016214501753382309>
- Kieffer, M.J. (2010). Socioeconomic status, English proficiency, and late-emerging reading difficulties. *Educational Researcher*, 39(6), 484–486. <https://doi.org/10.3102/0013189X10378400>
- Kieffer, M.J., & Lesaux, N.K. (2007). Breaking down words to build meaning: Morphology, vocabulary, and reading comprehension in the urban classroom. *The Reading Teacher*, 61(2), 134–144. <https://doi.org/10.1598/RT.61.2.3>
- Kieffer, M.J., & Vukovic, R.K. (2013). Growth in reading-related skills of language minority learners and their classmates: More evidence for early identification and intervention. *Reading and Writing*, 26(7), 1159–1194. <https://doi.org/10.1007/s11145-012-9410-7>
- Ladson-Billings, G., & Tate, W.F. (1995). Toward a critical race theory of education. *Teachers College Record*, 97(1), 47–68.
- LaHuis, D.M., Hartman, M.J., Hakoyama, S., & Clark, P.C. (2014). Explained variance measures for multilevel models. *Organizational Research Methods*, 17(4), 433–451. <https://doi.org/10.1177/1094428114541701>
- Lesaux, N.K., Crosson, A., Kieffer, M.J., & Pierce, M. (2010). Uneven profiles: Language minority learners’ word reading, vocabulary, and reading comprehension skills. *Journal of Applied Developmental Psychology*, 31(6), 475–483. <https://doi.org/10.1016/j.appdev.2010.09.004>
- 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. <https://doi.org/10.3102/0002831214532165>
- Lipson, M. (1995). The effect of semantic mapping instruction on prose comprehension of below-level college readers. *Reading Research and Instruction*, 34(4), 367–378. <https://doi.org/10.1080/19388079509558192>
- Lu, B., Preisser, J.S., Qaqish, B.F., Suchindran, C., Bangdiwala, S.I., & Wolfson, M. (2007). A comparison of two bias-corrected covariance estimators for generalized estimating equations. *Biometrics*, 63(3), 935–941. <https://doi.org/10.1111/j.1541-0420.2007.00764.x>
- Ma, S., Zhang, J., Anderson, R.C., Morris, J., Nguyen-Jahiel, K.T., & Miller, N., ... Grabow, K. (2017). Children’s productive use of academic vocabulary. *Discourse Processes*, 54(1), 40–61. <https://doi.org/10.1080/0163853x.2016.1166889>
- MacGinitie, W., MacGinitie, R., Maria, K., & Dreyer, L. (2002). *Gates-MacGinitie Reading Tests* (4th ed.). Itasca, IL: Riverside.
- Mancilla-Martinez, J., & Lesaux, N.K. (2010). Predictors of reading comprehension for struggling readers: The case of Spanish-speaking language minority learners. *Journal of Educational Psychology*, 102(3), 701–711. <https://doi.org/10.1037/a0019135>
- McCoach, D.B., & Adelson, J.L. (2010). Dealing with dependence (part I): Understanding the effects of clustered data. *Gifted Child Quarterly*, 54(2), 152–155. <https://doi.org/10.1177/0016986210363076>
- McNeish, D.M. (2014). Modeling sparsely clustered data: Design-based, model-based, and single-level methods. *Psychological Methods*, 19(4), 552–563. <https://doi.org/10.1037/met0000024>
- McNeish, D., & Stapleton, L.M. (2016a). Modeling clustered data with very few clusters. *Multivariate Behavioral Research*, 51(4), 495–518. <https://doi.org/10.1080/00273171.2016.1167008>

- McNeish, D.M., & Stapleton, L.M. (2016b). The effect of small sample size on two-level model estimates: A review and illustration. *Educational Psychology Review*, 28(2), 295–314. <https://doi.org/10.1007/s10648-014-9287-x>
- National Center for Education Statistics. (2017). *Indicator 7: Racial/ethnic concentration in public schools*. Retrieved from https://nces.ed.gov/programs/raceindicators/indicator_rbe.asp
- Newell, G.E., Beach, R., Smith, J., & VanDerHeide, J. (2011). Teaching and learning argumentative reading and writing: A review of research. *Reading Research Quarterly*, 46(3), 273–304. <https://doi.org/10.1598/RRQ.46.3.4>
- Paris, D. (2012). Culturally sustaining pedagogy: A needed change in stance, terminology, and practice. *Educational Researcher*, 41(3), 93–97. <https://doi.org/10.3102/0013189X12441244>
- Parra, M.O. (2018). *“En español es distinto”: Translanguaging for linguistic awareness and meaningful engagement with texts* (Unpublished doctoral dissertation). Boston College, Chestnut Hill, MA.
- Parra, M.O., Wagner, C.J., Proctor, C.P., Leighton, C., Robertson, D.A., Paratore, J., & Ford-Connors, E. (2016). Dialogic reasoning: Supporting emergent bilingual students’ language and literacy development. In C.P. Proctor, A. Boardman, & E.H. Hiebert (Eds.), *Teaching emergent bilingual students: Flexible approaches in an era of new standards* (pp. 119–137). New York, NY: Guilford.
- Perfetti, C.A., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189–213). Philadelphia, PA: John Benjamins.
- Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading*, 18(1), 22–37. <https://doi.org/10.1080/10888438.2013.827687>
- Poza, L.E. (2015). Mecanismos de exclusión: Ideologías lingüísticas y ambigüedades en la conceptualización de lenguaje académico [Mechanisms of exclusion: Linguistic ideologies and ambiguities in the conceptualization of academic language]. *Miríada Hispánica*, 10, 225–243.
- Proctor, C.P., & Bacon, C. (in press). Demographic realities and methodological flexibility in literacy teaching and research. In E.B. Moje, P.P. Afflerbach, P. Enciso, & N.K. Lesaux (Eds.), *Handbook of reading research* (Vol. 5). New York, NY: Longman.
- Proctor, C.P., Dalton, B., & Grisham, D. (2007). Scaffolding English language learners and struggling readers in a multimedia hypertext environment with embedded strategy instruction and vocabulary support. *Journal of Literacy Research*, 39(1), 71–93. <https://doi.org/10.1080/10862960709336758>
- Proctor, C.P., Dalton, D., Uccelli, P., Biancarosa, G., Mo, E., Snow, C., & Neugebauer, S. (2011). Improving comprehension online: Effects of deep vocabulary instruction with bilingual and monolingual fifth graders. *Reading and Writing*, 24(5), 517–544. <https://doi.org/10.1007/s11145-009-9218-2>
- Proctor, C.P., Silverman, R.D., Harring, J.R., & Montecillo, C. (2012). The role of vocabulary depth in predicting reading comprehension among English monolingual and Spanish-English bilingual children in elementary school. *Reading and Writing*, 25(7), 1635–1664. <https://doi.org/10.1007/s11145-011-9336-5>
- Puzio, K., Keyes, C.S., Cole, M.W., & Jiménez, R.T. (2013). Language differentiation: Collaborative translation to support bilingual reading. *Bilingual Research Journal*, 36(3), 329–349. <https://doi.org/10.1080/15235882.2013.845118>
- RAND Reading Study Group. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Santa Monica, CA: RAND.
- Raudenbush, S.W., & Bryk, A.S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.
- Raudenbush, S.W., Spybrook, J., Congdon, R., Liu, X.F., Martinez, A., & Bloom, H. (2011). Optimal Design Software for Multi-Level and Longitudinal Research (Version 3.01) [Computer software]. New York, NY: William T. Grant Foundation.
- Reed, D.K. (2008). A synthesis of morphology interventions and effects on reading outcomes for students in grades K–12. *Learning Disabilities Research & Practice*, 23(1), 36–49. <https://doi.org/10.1111/j.1540-5826.2007.00261.x>
- Reznitskaya, A., Kuo, L., Clark, A., Miller, B., Jadallah, M., Anderson, R.C., & Nguyen-Jahiel, K. (2009). Collaborative reasoning: A dialogic approach to group discussions. *Cambridge Journal of Education*, 39(1), 29–48. <https://doi.org/10.1080/03057640802701952>
- Rights, J.D., & Sterba, S.K. (2018). Quantifying explained variance in multilevel models: An integrative framework for defining R-squared measures. *Psychological Methods*. Advance online publication. <https://doi.org/10.1037/met0000184>
- Ryan, C. (2013). *Language use in the United States: 2011*. Washington, DC: U.S. Census Bureau.
- Salazar, M.C. (2013). A humanizing pedagogy: Reinventing the principles and practice of education as a journey toward liberation. *Review of Research in Education*, 37(1), 121–148. <https://doi.org/10.3102/0091732X12464032>
- Samson, J.F., & Collins, B.A. (2012). *Preparing all teachers to meet the needs of English language learners: Applying research to policy and practice for teacher effectiveness*. Washington, DC: Center for American Progress.
- SAS Institute. (2015). *SAS/STAT 9.4 user’s guide*. Cary, NC: Author.
- Schleppegrell, M.J. (2004). *The language of schooling: A functional linguistics perspective*. Mahwah, NJ: Erlbaum.
- Shin, S.J. (2013). *Bilingualism in schools and society: Language, identity, and policy*. New York, NY: Routledge.
- Silverman, R., & Hines, S. (2009). The effects of multimedia-enhanced instruction on the vocabulary of English-language learners and non-English-language learners in pre-kindergarten through second grade. *Journal of Educational Psychology*, 101(2), 305–314. <https://doi.org/10.1037/a0014217>
- Silverman, R.D., Proctor, C.P., Harring, J.R., Doyle, B., Mitchell, M.A., & Meyer, A.G. (2014). Teachers’ instruction and students’ vocabulary and comprehension: An exploratory study with English monolingual and Spanish-English bilingual students in grades 3–5. *Reading Research Quarterly*, 49(1), 31–60. <https://doi.org/10.1002/rrq.63>
- Silverman, R.D., Proctor, C.P., Harring, J.R., Hartranft, A.M., Doyle, B., & Zelinke, S.B. (2015). Language skills and reading comprehension in English monolingual and Spanish-English bilingual children in grades 2–5. *Reading and Writing*, 28(9), 1381–1405. <https://doi.org/10.1007/s11145-015-9575-y>
- Snow, C.E., Lawrence, J.F., & White, C. (2009). Generating knowledge of academic language among urban middle school students. *Journal of Research on Educational Effectiveness*, 2(4), 325–344. <https://doi.org/10.1080/19345740903167042>
- Stoddard, K., Valcante, G., Sindelar, P., O’Shea, L., & Algozzine, B. (1993). Increasing reading rate and comprehension: The effects of repeated readings, sentence segmentation, and intonation training. *Reading Research and Instruction*, 32(4), 53–65. <https://doi.org/10.1080/19388079309558133>
- Taboada, A., Bianco, S., & Bowerman, V. (2012). Text-based questioning: A comprehension strategy to build English language learners’ content knowledge. *Literacy Research and Instruction*, 51(2), 87–109. <https://doi.org/10.1080/19388071.2010.522884>
- Taboada, A., & Rutherford, V. (2011). Developing reading comprehension and academic vocabulary for English language learners through science content: A formative experiment. *Reading Psychology*, 32(2), 113–157. <https://doi.org/10.1080/02702711003604468>
- Uccelli, P., Barr, C.D., Dobbs, C.L., Phillips Galloway, E., Meneses, A., & Sánchez, E. (2015). Core academic language skills: An expanded operational construct and a novel instrument to chart school-

- relevant language proficiency in preadolescent and adolescent learners. *Applied Psycholinguistics*, 36(5), 1077–1109. <https://doi.org/10.1017/S014271641400006X>
- Uccelli, P., Phillips Galloway, E., Barr, C.D., Meneses, A., & Dobbs, C.L. (2015). Beyond vocabulary: Exploring cross-disciplinary academic-language proficiency and its association with reading comprehension. *Reading Research Quarterly*, 50(3), 337–356. <https://doi.org/10.1002/rrq.104>
- Uccelli, P., Phillips Galloway, E., & Qin, W. (in press). The language for school literacy: Widening the lens on language and reading relations. In E.B. Moje, P.P. Afflerbach, P. Enciso, & N.K. Lesaux (Eds.), *Handbook of reading research* (Vol. 5). New York, NY: Longman.
- Valdés, G. (2004). Between support and marginalisation: The development of academic language in linguistic minority children. *International Journal of Bilingual Education and Bilingualism*, 7(2/3), 102–132. <https://doi.org/10.1080/13670050408667804>
- 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(4), 297–324. <https://doi.org/10.1080/19345740903167018>
- Wagner, C.J., Parra, M.O., & Proctor, C.P. (2017). The interplay between student-led discussions and argumentative writing. *TESOL Quarterly*, 51(2), 438–449. <https://doi.org/10.1002/tesq.340>
- Wanzek, J., Petscher, Y., Al Otaiba, S., Rivas, B.K., Jones, F.G., Kent, S.C., ... Mehta, P. (2017). Effects of a year long supplemental reading intervention for students with reading difficulties in fourth grade. *Journal of Educational Psychology*, 109(8), 1103–1119. <https://doi.org/10.1037/edu0000184>
- What Works Clearinghouse. (2017). *What Works Clearinghouse standards handbook version 4.0*. Washington, DC: Institute of Education Sciences, U.S. Department of Education.
- WIDA. (2015). *ACCESS for ELLs 2.0*. Madison, WI: Author.
- Wilkinson, I.A.G., Soter, A.O., & Murphy, P.K. (2010). Developing a model of quality talk about literary text. In M.G. McKeown & L. Kucan (Eds.), *Bringing reading research to life* (pp. 142–169). New York, NY: Guilford.
- Wilkinson, P.A., & Patty, D. (1993). The effects of sentence combining on the reading comprehension of fourth grade students. *Research in the Teaching of English*, 27(1), 104–125.
- Woodcock, R.W., Muñoz-Sandoval, A.F., Rief, M.L., & Alvarado, C.G. (2005). *Woodcock-Muñoz Language Survey-revised*. Itasca, IL: Riverside.
- Zhang, J., Anderson, R.C., & Nguyen-Jahiel, K. (2013). Language-rich discussions for English language learners. *International Journal of Educational Research*, 58, 44–60. <https://doi.org/10.1016/j.ijer.2012.12.003>

LITERATURE CITED

- Applegate, K. (2014). *Ivan: The remarkable true story of the shopping mall gorilla*. New York, NY: Clarion.
- CLAVES Research Team (2015a). *Bilingual education*. Chestnut Hill, MA: Boston College.
- CLAVES Research Team (2015b). *Species revival*. Chestnut Hill, MA: Boston College.
- Cohn, D. (2002). *¡Sí, se puede!/Yes, we can! Janitor strike in L.A.* El Paso, TX: Cinco Puntos.
- Elya, S.M. (2002). *Home at last*. New York, NY: Lee & Low.
- George, J.C. (2008). *The wolves are back*. New York, NY: Dutton.

Submitted June 5, 2018

Final revision received April 3, 2019

Accepted April 3, 2019

C. PATRICK PROCTOR (corresponding author) is a professor in the Department of Teacher Education, Special Education, Curriculum and Instruction at Boston College, Chestnut Hill, Massachusetts, USA; email proctoch@bc.edu. He is a former bilingual teacher whose research focuses on bilingualism and literacy across instructional languages and contexts.

REBECCA D. SILVERMAN is an associate professor in the Department of Curriculum Studies and Teacher Education at Stanford University, California, USA; email rdsilver@stanford.edu.

JEFFREY R. HARRING is a professor in the Department of Human Development and Quantitative Methodology at the University of Maryland, College Park, USA; email harring@umd.edu.

RENATA LOVE JONES is a doctoral candidate specializing in language, literacy, and culture and critical perspectives in curriculum and instruction at Boston College, Chestnut Hill, Massachusetts, USA; email jonesqc@bc.edu.

ANNA M. HARTRANFT was a faculty research associate at the University of Maryland, College Park, at the time of this study and is now an academic specialist at Partners in Learning, Washington, DC, USA; email agenevievemeyer@gmail.com.

APPENDIX

Intervention Scope and Sequence

Unit 1: Nature

Introduction: Focus on interdependence within an ecosystem.

Texts	Day 1: Comprehension and vocabulary	Day 2: Comprehension and vocabulary	Day 3: Language skills (syntax and morphology)	Day 4: Language skills (syntax and morphology)	Day 5: Small-group discussion	Writing cycle
<ul style="list-style-type: none"> Text: <i>The Wolves Are Back</i> (Lexile level 630) Video: “Wolves and People” 	<ul style="list-style-type: none"> Introduce background knowledge on Yellowstone National Park. Use the PowerPoint presentation to introduce the key words (definitions, examples, word parts, cognates, and visuals/interactives): <i>depopulate</i>, <i>exterminate</i>, and <i>reintroduce</i>. Facilitate reading of <i>The Wolves Are Back</i> with questioning, inferencing, and summarizing. 	<ul style="list-style-type: none"> Review the text and vocabulary and introduce new words: <i>balance</i> and <i>restore</i>. Complete facilitated reading. Facilitate viewing of the “Wolves and People” video with questioning, inferencing, and summarizing. Guide the cause-and-effect interdependence activity with picture cards. 	<ul style="list-style-type: none"> Guide the word web activity focused on <i>depopulation</i>, <i>extermination</i>, and <i>reintroduction</i>. Introduce the word parts <i>re-</i> and <i>de-</i> and connect to the days 1 and 2 texts. Guide the morphology activity in which students identify, analyze, and construct words with <i>re-</i> and <i>de-</i>. Encourage students to create sentences using their constructed words. 	<ul style="list-style-type: none"> Review parts of speech and teach subject, verb, and object in the context of simple sentences. Encourage students to identify simple sentences in <i>The Wolves Are Back</i>. Guide students to generate simple sentences and identify parts of speech in a cut-up sentences word card game. 	<ul style="list-style-type: none"> Facilitate student discussion of the following prompt: “Should animals, like wolves, who eat other animals, be reintroduced into areas where they will encounter humans and livestock?” Model the prompt, clarify, summarize, repeat, and intervene as needed. Take a poll and discuss students’ stances. 	<p>Guide students to write a letter to a national park director or a member of the U.S. Congress.</p>
<ul style="list-style-type: none"> Text: <i>Species Revival</i> (Lexile level 890) Video: “Revive and Restore” 	<ul style="list-style-type: none"> Introduce background knowledge on extinction. Use the PowerPoint presentation to introduce the key words (definitions, examples, word parts, cognates, and visuals/interactives): <i>species</i>, <i>revive</i>, and <i>extinct</i>. Facilitate reading of <i>Species Revival</i> with questioning, inferencing, and summarizing. 	<ul style="list-style-type: none"> Review the text and vocabulary and introduce new words: <i>endangered</i> and <i>illegal</i>. Facilitate viewing of the “Revive and Restore” video with a focus on questioning, inferencing, and summarizing. Guide the negotiation of meaning activity using picture cards to discuss endangered versus extinct animals. 	<ul style="list-style-type: none"> Guide the word web activity focused on <i>endangered</i> and <i>illegal</i> and how these words are related. Introduce the word parts <i>en-/em-</i> and <i>il-/ir-/im-/in-</i> and connect them to the days 1 and 2 texts. Guide the morphology activity in which students identify, analyze, and construct words with the prefixes above. Encourage students to create sentences using their constructed words. 	<ul style="list-style-type: none"> Use <i>Species Revival</i> to introduce object and subject pronouns. Encourage students to find object and subject pronouns in <i>Species Revival</i>. Guide students to work with partners to determine which object or subject pronoun fits best in provided sentences and discuss why. 	<ul style="list-style-type: none"> Facilitate student discussion of the following prompt: “Should scientists revive extinct animals?” Model the prompt, clarify, summarize, repeat, and intervene as needed. Take a poll and discuss students’ stances. 	

Unit 2: Rights

Introduction: Focus on rights, fair, freedom, and advocate.						
Texts	Day 1: Comprehension and vocabulary	Day 2: Comprehension and vocabulary	Day 3: Language skills (syntax and morphology)	Day 4: Language skills (syntax and morphology)	Day 5: Small-group discussion	Writing cycle
<ul style="list-style-type: none"> Text: <i>Ivan: The Remarkable True Story of the Shopping Mall Gorilla</i> (Lexile level 620) Video: “Gorillas Reintroduced” 	<ul style="list-style-type: none"> Activate background knowledge on zoos. Use the PowerPoint presentation to introduce the key words (definitions, examples, word parts, cognates, and visuals/interactives): <i>captivity</i>, <i>wild</i>, and <i>treatment</i>. Facilitate reading of <i>Ivan</i> with questioning, inferencing, and summarizing. Guide students to compare and contrast Ivan’s treatment in different settings. 	<ul style="list-style-type: none"> Review the text and vocabulary and introduce new words: <i>petition</i> and <i>protest</i>. Complete the facilitated reading. Facilitate viewing of the “Gorillas Reintroduced” video with questioning, inferencing, and summarizing. Guide the cause-and-effect activity connecting Ivan’s treatment to people’s response (petitions, protests, etc.). 	<ul style="list-style-type: none"> Guide the word web activity focused on <i>wild</i>, <i>captivity</i>, and <i>treatment</i>. Introduce the word parts <i>-ity</i> and <i>-ment</i> and connect them to the days 1 and 2 texts. Guide the morphology activity in which students identify, analyze, and construct words with <i>-ity</i> and <i>-ment</i>. Encourage students to create sentences using their constructed words. 	<ul style="list-style-type: none"> Teach or review past, present, and future tenses. Encourage students to identify past, present, and future tenses in <i>Ivan</i>. Guide students to identify sentences using past, present, and future tenses via a verb tense Bingo game. 	<ul style="list-style-type: none"> Facilitate student discussion of the following prompt: “Should people protest if they believe it is against animal rights to hold animals in captivity?” Model the prompt, clarify, summarize, repeat, and intervene as needed. Take a poll and discuss students’ stances. 	<p>Guide students to write a speech for school administrators, fellow students, or parents.</p>
<ul style="list-style-type: none"> Text: <i>¡Si, Se Puede! Yes, We Can! Janitor Strike in L.A.</i> (Lexile level 800) Video: “Chicago Teacher Strike” 	<ul style="list-style-type: none"> Activate background knowledge on <i>petition</i> and <i>protest</i>. Use the PowerPoint presentation to introduce the key words (definitions, examples, word parts, cognates, and visuals/interactives): <i>strike</i> and <i>march</i>. Facilitate reading of <i>¡Si, Se Puede! Yes, We Can!</i> with a focus on inferencing, questioning, and summarizing. 	<ul style="list-style-type: none"> Review the text and vocabulary and introduce new words: <i>union</i> and <i>rally</i>. Facilitate viewing of the “Chicago Teacher Strike” video with a focus on inferencing, questioning, and summarizing. Guide the negotiation of meaning activity using picture cards to discuss connections among <i>rally</i>, <i>strike</i>, <i>march</i>, and <i>unite</i>. 	<ul style="list-style-type: none"> Guide the word web activity focused on words related to <i>rally</i>, <i>strike</i>, <i>march</i>, and <i>unite</i> and how these words are related. Introduce the word parts <i>er</i> (as in <i>worker</i>) and <i>or</i> (as in <i>janitor</i>) and connect to the days 1 and 2 texts. Guide the morphology activity in which students identify, analyze, and construct words with <i>er</i> and <i>or</i>. Encourage students to create sentences using their constructed words. 	<ul style="list-style-type: none"> Use <i>¡Si, Se Puede! Yes, We Can!</i> to introduce or review simple past (<i>-ed/irregular</i>), present (including <i>-s/-es</i>), future (<i>will/going to</i>) tenses. Encourage students to find different tenses in text. Guide the Slap It! game in which students match present-tense verbs with their past and future forms. 	<ul style="list-style-type: none"> Facilitate student discussion of the following prompt: “Should people/workers go on strike to protest working conditions?” Model the prompt, clarify, summarize, repeat, and intervene as needed. Take a poll and discuss students’ stances. 	

Unit 3: Immigration

Introduction: Focus on immigration, culture, and diversity.

Texts	Day 1: Comprehension and vocabulary	Day 2: Comprehension and vocabulary	Day 3: Language skills (syntax and morphology)	Day 4: Language skills (syntax and morphology)	Day 5: Small-group discussion	Writing cycle
<ul style="list-style-type: none"> Text: <i>Home at Last</i> (Lexile level 620) Video: “Immersion” 	<ul style="list-style-type: none"> Activate background knowledge on immigration. Use the PowerPoint presentation to introduce the key words (definitions, examples, word parts, cognates, and visuals/interactives): <i>adapt/adaptation</i> and <i>communicate/communication</i>. Facilitate reading of <i>Home at Last</i> with questioning, inferencing, and summarizing. Guide students to compare and contrast how characters in the text adapt to the United States. 	<ul style="list-style-type: none"> Review the text and vocabulary and introduce new words: <i>assimilate</i> and <i>immersion</i>. Complete the facilitated reading. Facilitate viewing of the beginning of the “Immersion” video with questioning, inferencing, and summarizing. Guide the T-chart activity. On one side, define <i>immigration</i>, <i>adaptation</i>, and <i>assimilation</i>, and on the other side, find evidence in the text and video. 	<ul style="list-style-type: none"> Guide the word web activity focused on <i>immigration</i>, <i>communication</i>, <i>adaptation</i>, <i>assimilation</i>, and <i>immersion</i>. Introduce the word parts <i>-tion/-ation</i> and <i>-sion</i> and connect them to the days 1 and 2 texts. Guide the morphology activity in which students identify, analyze, and construct words with <i>-tion/-ation</i> and <i>-sion</i>. Encourage students to create sentences using their constructed words. 	<ul style="list-style-type: none"> Teach or review coordinating conjunctions and compound sentences. Encourage students to identify coordinating conjunctions and compound sentences in <i>Home at Last</i>. Guide students to play a sentence-combining game in which they practice generating compound sentences given two clause cards and one conjunction card. 	<ul style="list-style-type: none"> Facilitate student discussion of the following prompt: “Should immigrants change their language and culture when they move to a new country?” Model the prompt, clarify, summarize, repeat, and intervene as needed. Take a poll and discuss students’ stances. 	<p>Guide students to write an article for a local or student newspaper.</p>
<ul style="list-style-type: none"> Text: <i>Bilingual Education</i> (Lexile level 1020) Video: “Immersion” 	<ul style="list-style-type: none"> Activate background knowledge on bilingual education. Use the PowerPoint presentation to introduce the key words (definitions, examples, word parts, cognates, and visuals/interactives): <i>proponent/opponent</i> and <i>advantage/disadvantage</i>. Facilitate reading of the text with inferencing, questioning, and summarizing. 	<ul style="list-style-type: none"> Review the text and vocabulary and introduce new words: <i>policy</i> and <i>proficient</i>. Facilitate viewing of the end of the “Immersion” video with a focus on inferencing, questioning, and summarizing. Guide the negotiation of meaning activity using a cause-and-effect chart to discuss policies and their consequences. 	<ul style="list-style-type: none"> Guide the word web activity focused on words related to <i>rally</i>, <i>policy</i>, <i>proponent</i>, and <i>opponent</i> and how these words are related. Introduce the word part <i>-ly</i>, as in <i>proficiently</i> in the days 1 and 2 texts. Guide the morphology activity in which students identify, analyze, and construct words with <i>-ly</i>. Encourage students to create sentences using their constructed words. 	<ul style="list-style-type: none"> Use <i>Bilingual Education</i> to introduce or review subordinating conjunctions and complex sentences. Encourage students to find subordinating conjunctions and complex sentences in the text. Guide students to play a sentence-combining game in which they practice generating compound sentences given clause and conjunction cards. 	<ul style="list-style-type: none"> Facilitate student discussion of the following prompt: “Should schools teach in English only or offer bilingual education?” Encourage students to use evidence from the days 1 and 2 texts. Model the prompt, clarify, summarize, repeat, and intervene as needed. Take a poll and discuss students’ stances. 	