Teachers' Instruction and Students' Vocabulary and Comprehension: An Exploratory Study With English Monolingual and Spanish–English Bilingual Students in Grades 3–5

Rebecca D. Silverman
University of Maryland, College Park, USA

C. Patrick Proctor
Boston College, Chestnut Hill, Massachusetts, USA

Jeffrey R. Harring
Brie Doyle
Marisa A. Mitchell
Anna G. Meyer
University of Maryland, College Park, USA

ABSTRACT

The primary aim of this study was to explore the relationship between teachers' instruction and students' vocabulary and comprehension in grades 3–5. The secondary aim of this study was to investigate whether this relationship differed for English monolingual and Spanish–English bilingual students. To meet these aims, we observed and recorded reading/language arts instruction in 33 classrooms at three points during an academic year, and we assessed 274 students on vocabulary and comprehension at the beginning and end of the year. Using field notes and student utterances to understand the context, we coded teacher utterances (e.g., questions, comments, prompts) as vocabulary instruction, comprehension instruction, other instruction, or noninstruction. We then identified five types of vocabulary-related instruction and five types of comprehension-related instruction. Using latent difference modeling, we investigated how the frequency of different types of instruction was associated with change in students' vocabulary and comprehension across the school year. Teachers' instruction related to definitions, word relations, and morphosyntax was positively associated with change in vocabulary; teachers' instruction related to application across contexts and literal comprehension was negatively associated with change in vocabulary; and teachers' instruction related to inferential comprehension was positively associated with change in comprehension. The findings also revealed an interaction between language status and teachers' instruction, such that instruction that attended to comprehension strategies was associated with greater positive change in comprehension for bilingual (but not for monolingual) students.

Given that vocabulary and comprehension are necessary for college and career readiness, it is no wonder that there is substantial focus on these skills in the Common Core State Standards (CCSS; National Governors Association Center for Best Practices & Council of Chief State School Officers [NGACBP & CCSSO], 2010). These widely adopted standards set high expectations for all students, including linguistically diverse learners, in reading, writing, speaking and listening, and language. As educators consider how to support students in meeting these standards, information on the relationship between teachers' instruction and students' vocabulary and comprehension is essential. In particular, educators need to understand whether and how the relationship between instruction and vocabulary and comprehension differs for students from diverse language backgrounds.

Reading Research Quarterly, 49(1)
pp. 31–60 | doi:10.1002/rrq.63

Experimental research on reading interventions has proliferated over the past 30 years, and results from this line of inquiry provide robust evidence that effective instruction under controlled conditions leads to improved literacy outcomes for monolingual and bilingual students alike (see, e.g., Gersten et al., 2007; Hairrell, Rupley, & Simmons, 2011; Kamil et al., 2008; National Institute of Child Health and Human Development [NICHD], 2000; Shanahan et al., 2010). Research on teaching and learning in natural classroom settings, however, has not kept pace. Hoffman, Maloch, and Sailors (2011) noted that although there has been a recent uptick in observational research, "the number of observational studies is still pitifully few in relation to the number of studies conducted annually and painfully small in relation to the formulation of policies that shape practice" (p. 26). In short, there is a paucity of observational research conducted in linguistically diverse classrooms.

Considering that many students falter in reading in upper elementary school as they encounter more challenging texts and complex content than ever before (Catts, Compton, Tomblin, & Bridges, 2012; Chall & Jacobs, 2003; National Center for Education Statistics [NCES], 2011), continued research on the relationship between instruction and vocabulary at this critical juncture is needed. Additionally, students who speak a language other than English at home are at particular risk for experiencing difficulty in the upper elementary grades (Lesaux, 2006; Mancilla-Martinez & Lesaux, 2010; NCES, 2011). Considering the continued growth of Latino students from Spanish-dominant homes in U.S. schools, educators are especially concerned about addressing the needs of this group of students (Calderón, Slavin, & Sanchez, 2011). Many of these students speak both Spanish and English socially and in familial contexts; thus, they are, to varying degrees, bilingual (Grosjean, 2010). However, many of these students have had limited exposure to academic English, which is needed for success in school (Lesaux & Geva, 2006). Consequently, there is an urgent need to identify instruction that is positively related to English academic outcomes for this vulnerable group of learners.

Thus, the purpose of our study was to explore the relationship between teachers' instruction and monolingual and bilingual students' vocabulary and comprehension in linguistically diverse upper elementary school classrooms. Considering the important relationship between teacher talk and student learning (e.g., Applebee, Langer, Nystrand, & Gamoran, 2003; Dickinson & Porche, 2011; Duffy, Roehler, & Rackliffe, 1986), we used teacher utterances as our unit of analysis. We referred to field notes and student utterances to understand the context of instruction, and we coded teachers' instructional talk (e.g., questions, comments, prompts) as vocabulary instruction, comprehension instruction, other instruction, or noninstruction.

We identified five types of vocabulary-related instruction and five types of comprehension-related instruction. Using latent difference modeling, we investigated how the frequency of different types of instructional talk was associated with change in third-, fourth-, and fifth-grade students' vocabulary and comprehension performance across the school year. In addition, we examined whether associations between types of instructional talk and students' vocabulary and comprehension differed for monolinguals and bilinguals.

Note that we refer to students whose parents reported that Spanish was spoken in the home as bilingual. In other research, these students may be called, for example, English learners, English as a second language students, limited English-proficient students, or language-minority learners. We use these terms interchangeably with the term *bilingual* in the following review of the research that informed this study.

The Landscape of Observational Research

Researchers have taken a variety of approaches to studying classroom instruction. Some researchers have investigated broad dimensions of teacher instruction, others have explored specific instructional practices, and still others have examined instructional discourse in classrooms. The following is a brief review of the research from these different traditions and an explanation of how they informed our approach to studying instruction. In this review, we pay particular attention to studies that focused on vocabulary and comprehension in upper elementary school (i.e., including students in grades 3–5) and the extent to which these studies included linguistically diverse students.

Studies on Broad Dimensions of Teacher Instruction

In the early 1970s, Brophy and colleagues (Brophy, 1973, Brophy & Good, 1970; Veldman & Brophy, 1974) studied teacher effectiveness in natural classroom settings, primarily in second and third grades, and correlated teacher behaviors (e.g., questions and feedback) with student achievement. These researchers noted that what teachers do and say affects student learning. Subsequently, research by Greenwood and colleagues (e.g., Greenwood, Abbott, & Tapia, 2003; Greenwood, Arreaga-Mayer, & Carta, 1994; Greenwood, Carta, Arreaga-Mayer, & Rager, 1991), conducted mostly in fourth- and fifth-grade classrooms, showed that students tended to have greater academic achievement

in classrooms where there was more fast-paced, direct instruction, including teacher feedback.

Recently, studies by Connor, Morrison, and colleagues (e.g., Connor, Morrison, & Katch 2004; Connor, Morrison, & Petrella, 2004; Connor, Morrison, & Slominski, 2006) have added to the body of classroom observation research on broad dimensions of instruction. Specifically, these researchers have been interested in identifying the effects of teacher-managed versus student-managed instruction, meaning-related versus code-related instruction, and explicit versus implicit instruction. Connor's group also is concerned with student × instruction interactions in which instruction is differentially effective for students of higher and lower ability levels.

For example, in a study of instruction in third-grade classrooms, Connor, Morrison, and Petrella (2004) found that students with average or low initial comprehension demonstrated more comprehension growth in classrooms in which there was more teacher-managed rather than student-managed instruction. In classrooms in which there was more student-managed instruction, however, students with above-average initial comprehension exhibited more comprehension growth.

Finally, an emerging set of studies, recently published by Carlisle and colleagues (e.g., Carlisle, Kelcey, Berebitsky, & Phelps, 2011; Kelcey & Carlisle, 2013), has similarly focused on multiple dimensions of instruction, including pedagogical structure, teacher-directed instruction, and support for student learning. Pedagogical structure focuses on how teachers draw students' attention to the "purpose and structure of a given lesson" (Carlisle et al., 2011, p. 412). Examples include giving directions and explaining lesson objectives. Teacher-directed instruction relates to the way in which teachers promote literacy skills through explicit instruction. Examples include providing explanations, modeling comprehension strategies, and guiding practice. Support for student learning corresponds to ways in which teachers "engage students in the lessons, assess their response to the content and activity of a lesson, and make use of students' skills, strategies, and knowledge" (p. 413). Results from the Carlisle et al. study, which was conducted in third-grade classrooms, indicated that teacher-directed instruction and support for student learning were positively related to students' reading comprehension.

This body of observation work, in which researchers investigated broad dimensions of reading instruction, provides an indication of how instruction is related to student outcomes. Nevertheless, this literature fails to provide information about specific types of vocabulary and comprehension instruction in natural classroom settings. As a result, conclusions about the relationships between regular classroom instruction of vocabulary and comprehension, established as effective through intervention research, cannot be made.

For example, intervention research suggests that instruction that includes definitional and contextual information is supportive of students' vocabulary, and instruction focused on comprehension strategies and text structures is supportive of students' comprehension (e.g., Gersten et al., 2007; Hairrell et al., 2011; Kamil, et al., 2008; NICHNational Institute of Child Health and Human Development, 2000; Shanahan et al., 2010). Yet, research on broad dimensions of instruction does not investigate instruction at this level of specificity. Further, although some studies on broad dimensions of instruction have included bilingual learners (e.g., 18% of students in the Carlisle et al., 2011, study were labeled as limited English proficient), these studies do not disaggregate findings for monolingual and bilingual students. More research is thus needed that connects specific types of vocabulary and comprehension instruction, provided in a regular classroom setting, and outcomes for students from different language backgrounds.

Research on Vocabulary and Comprehension Instruction

Another body of research that looks more specifically at comprehension and vocabulary instruction practices in schools dates back to the late 1970s, when Durkin (1978) published her groundbreaking observation study of reading instruction in third- through sixth-grade classrooms in which "practically no comprehension instruction was seen" (p. 520) and less than 3% of instructional time was paid to instruction of word meanings. Twenty years after Durkin's study, Pressley, Wharton-McDonald, Mistretta-Hampston, and Echevarria (1998) found that despite compelling evidence of the positive effects of comprehension strategy instruction, this form of instruction was still noticeably absent from the educational landscape.

Similarly, in an investigation of effective schools and accomplished teachers in first through third grades, Taylor, Pearson, Clark, and Walpole (2000) found that fewer than 10 of the 70 teachers participating in their study focused on comprehension strategy instruction. When comprehension instruction was included, however, asking and answering text-based and high-level thinking questions in response to reading were found to be dominant instructional practices across the classrooms in the study.

A more recent study by Ness (2011), who coded for teacher behaviors related to explicit reading comprehension instruction, showed a substantial increase in comprehension instruction overall, and strategy instruction in particular, compared with previous studies. Specifically, observations of first- through fifth-grade

classrooms showed relatively frequent instruction focused on question answering, predicting/prior knowledge, summarization, and text structure. Less focus on the use of visual representations, monitoring, and question generation was observed.

Regarding vocabulary, Blachowicz (1987), in her study of fourth-grade classrooms, found that approximately 15% of instructional time in reading groups was spent on vocabulary. The most prominent instructional practices were "determining the meanings of words in context" (p. 134) and "examining words as discrete items, either pronouncing them or dealing with definitions or synonyms" (p. 135). Much less time was spent on "relating the words to other words" and "generalized strategies for figuring out words" (p. 135).

Nearly a decade later, Watts (1995) arrived at similar findings in her observation study that explored vocabulary instruction in fifth- and sixth-grade classrooms. She investigated five specific types of vocabulary instruction: definitional (e.g., definition provided by the teacher), contextual (e.g., target word used in one or two sentences), organizational (i.e., based on a semantic framework of relations among words), mnemonic (e.g., paired association, keyword methods), and structural (i.e., morphological and syntactical features of words). Watts found that instruction was mostly definitional and contextual in nature, which, she suggested, was more likely to foster surface-level rather than deep-level processing of words.

Following this study, Scott, Jamieson-Noel, and Asselin (2003) used similar methods in their observation study in fourth- through seventh-grade classrooms in Canada. Similar to Blachowicz and to Watts, these researchers found that vocabulary instruction focused mainly on definitional and contextual information about words and that instruction included semantic analysis of words as well. As in Watts's study, mnemonic and structural instruction were rarely seen.

This body of observation research provides useful snapshots of specific types of vocabulary and comprehension instruction in everyday classroom contexts and allows investigators to determine the extent to which classroom instruction is aligned with evidence-based practice (e.g., Gersten et al., 2007; Hairrell et al., 2011; Kamil et al., 2008; NICHNational Institute of Child Health and Human Development, 2000; Shanahan et al., 2010). However, this research base is limited because the unit of analysis (i.e., teacher practice or behavior) is underspecified in many of the studies, which makes it difficult to compare findings across studies (Hoffman et al., 2011). Considering that teacher practices and behaviors blend together and that delineation of teacher practices and behaviors can be subjective, more well-defined units of analysis are needed for research on teacher instruction. Further, the existing research base is limited because it fails to address the relationship between teachers' instruction and students' vocabulary and comprehension outcomes and to attend to both monolingual and bilingual learners as potentially unique responders to instruction.

The Role of Discourse in Classroom Instruction

Another tradition of investigating classroom instruction has its roots in the classroom discourse research of Bellack, Kleibard, Hyman, and Smith (1966), Flanders (1970), and Sinclair and Coulthard (1975). These researchers and others have investigated classroom talk to identify patterns of discourse in educational settings. Given that a substantial amount of classroom speech can be attributed to teachers (Boyd & Rubin, 2002; Chaudron, 1988; Nystrand, 2006; Sinclair & Coulthard, 1975), it makes sense for researchers to examine what teachers are saying and how they are saying it. Indeed, research on classroom talk has been conducted in monolingual and bilingual contexts (e.g., Cazden, 1998; Chaudron, 1988; Long, 1985). Specifically, studies have explored how teacher talk (e.g., asking questions, providing explanations, offering feedback) guides classroom discourse in classrooms with monolingual and bilingual students (Chaudron, 1988; Lindsay, 1990; Nystrand, Wu, Gamoran, Zeiser, & Long, 2003).

This body of research has developed well-specified units of analysis with which to focus investigations, making it relatively easy to compare findings across studies (Chaudron, 1988; Crookes, 1990; Lindsay, 1990). In fact, much of the research in this genre suggests that typical instruction can often be characterized by an Initiate-Response-Evaluate structure, in which teachers initiate closed-ended questions, students respond to these questions, and teachers evaluate students' responses (e.g., Cazden, 1998; Nystrand, 2006; Sinclair & Coulthard, 1975). There is little research, however, on classroom discourse in natural classroom settings that has delved into the specific vocabulary- and comprehension-related practices (e.g., defining and contextualizing words, focusing on comprehension strategies and text structure) that are inherent in teacher talk and how these practices are related to student outcomes.

Research that has investigated teacher talk focused on vocabulary and comprehension has been conducted mainly in the context of studies in which investigators have trained teachers on a specific instructional approach. For example, Duffy et al. (1986) trained teachers to incorporate explicit explanations into their reading instruction and studied how teachers' instructional talk, specifically teacher explanations of reading comprehension strategies, influenced students' understanding of lesson content in

fifth-grade classrooms. A major finding of the study was that "differences in what teachers say may create differences in student understanding" (p. 12). In fact, promising research on how to improve classroom talk comes from studies of specific discussion-based instructional approaches with monolingual and bilingual students alike (e.g., Beck, McKeown, Sandora, Kucan, & Worthy, 1996; Chinn, Anderson, & Waggoner, 2001; Goldenberg & Patthey-Chavez, 1995; Michaels, O'Connor, & Resnick, 2008).

In a recent meta-analysis of studies on various discussion approaches, many of which were conducted in fourth- through sixth-grade classrooms, Murphy, Wilkinson, Soter, Hennessey, and Alexander (2009) found that most discussion approaches led to reductions in teacher talk and increases in student talk and that many were effective at promoting students' literal and inferential comprehension. However, these researchers caution that an increase in student discussion was not necessarily related to an increase in student comprehension. In a related analysis, Soter et al. (2008) determined that discussion approaches that were more effective included a greater number of authentic questions, reasoning words, and elaborated explanations.

Although this research offers promise for the differential effect of varying types of teacher talk under managed conditions, it does not forward our understanding of teachers' vocabulary and comprehension instruction and its relationship to student outcomes in regular classroom settings. Work with younger students suggests the important role of teacher talk in supporting students' language and literacy development (e.g., Aukrust, 2007; Bowers & Vasilyeva, 2011; Dickinson & Porche, 2011), and research with older students indicates that teacher questions and scaffolding support student performance in middle and high school (e.g., Applebee et al., 2003). However, there is little research on the relationship between teachers' instructional talk and students' vocabulary and comprehension in upper elementary schools and even less research on these relationships with both monolingual and bilingual students.

The Present Study

We position this study at the intersection of these reviewed approaches to exploring teachers' instruction. Specifically, similar to research on broad dimensions of instruction, we are interested in the relationship between teacher instruction and student outcomes. Certainly, to inform efforts to improve instruction, connecting instruction to student learning is essential. We also align ourselves with the research base that has investigated instructional practice more discretely because this research has more concrete implications for the development of interventions focused on improving student vocabulary and comprehension. To investigate instructional practice, however, we borrow from research on classroom discourse to examine instruction through the lens of teacher talk. Although research on instructional practice has included data on teacher talk, this research has not used teacher talk as the specific unit of analysis.

Although we used field notes to contextualize teacher talk, and we referred to student talk in transcripts of lessons to understand the nature of teacher talk, we focused on teacher utterances as a vehicle of instruction that provides an indication of the extent to which teachers focus on specific vocabulary- and comprehension-related practices. Thus, it is the relationship between instruction, as manifested in teacher talk, and students' gains in vocabulary and comprehension that we investigated. Accordingly, the following research questions guided this study:

1. What are the relationships among specific types of instruction as seen in teacher talk and change in the vocabulary and comprehension of English monolingual and Spanish–English bilingual students in grades 3–5 over an academic year?
2. To what extent do these relationships depend on students' language status (i.e., monolingual or bilingual)?

Methods

Setting

We conducted our research in two semi-urban school districts, one in the Mid-Atlantic and the other in the Northeastern United States. Both districts have experienced substantial recent immigration from Central America and Mexico. We worked with three schools at each site (see Appendix A for a Table of the district and school demographics). In total, we observed 33 classrooms overall: 21 Mid-Atlantic classrooms and 12 Northeastern classrooms.

Student Participants

Across the 33 classrooms observed, we assessed 274 student participants (204 from the Mid-Atlantic site and 70 from the Northeastern site). All students in the focal classrooms were eligible to enroll in the study, and approximately 70% of parents allowed their children to participate. From those students whose parents provided consent for their child's participation, we included approximately 50% in our sample. In some classrooms, where overall consent was low, we included all students whose parents provided consent. In other classrooms, we used a stratified random selection criteria to identify

TABLE 1
Student Participants' Demographics (*N* = 274)

	Bilingual		Monolingual		Total sample	
Demographic category	***n***	**Percentage**	***n***	**Percentage**	***n***	**Percentage**
Gender						
Female	69	25.2	75	27.4	144	52.6
Grade						
3	46	16.8	60	21.9	106	38.7
4	50	18.2	44	16.1	94	34.3
5	27	9.9	47	17.2	74	27.0
Ethnicity						
Latino	122	44.5	6	2.2	128	46.7
White	1	0.4	45	16.4	46	16.8
Black	0	0	95	34.7	95	34.7
Other	0	0	5	1.8	5	1.8
Additional services						
Individualized Education Plan	13	4.7	32	11.7	45	16.4
Federal school lunch program	110	40.1	92	33.6	202	73.7

a sample with comparable numbers of monolingual and bilingual students, targeting eight students per classroom (see Table 1 for sample demographics).

Note that we identified students as bilingual based on data from parent surveys. Specifically, we classified students whose parents reported Spanish spoken in the home as bilingual. Students who spoke other non-Spanish languages in the home were excluded from the study.

Teacher Participants

Third- through fifth-grade classroom teachers at the six schools across the two research sites were invited to participate in the study. Of the 33 teachers in the target classrooms who volunteered to participate, 12% were male, 82% Caucasian, 3% African American, and 3% other, and 12% did not report their race/ethnicity. Teachers had a mean of 8.43 years of teaching experience (standard deviation [*SD*] = 8.25 years). Two-thirds of the teachers (67%) indicated that they held a master's degree, although two did not report their highest level of education.

In the classrooms at both the Mid-Atlantic and Northeastern sites, adults other than the primary classroom teacher were sometimes present during reading/language arts instructional time. These other adults (referred to as "second teacher" in coding) included reading specialists, special education teachers, and teachers of English for speakers of other languages. When these second teachers took on instructional responsibilities (e.g., coteaching) in the target classrooms, their instruction was captured alongside the instruction of the primary classroom teacher and was included in analyses. There were no difficulties in identifying the first and second teacher in each observation, as field notes were used to identify teachers in transcripts.

Classroom Context

In two schools in the Mid-Atlantic site and in all schools in the Northeastern site, students received instruction in self-contained classrooms in which the primary classroom teacher delivered all instructional content throughout the day. These sites thus yielded 24 teachers and 24 classrooms. One school in the Mid-Atlantic site was departmentalized such that reading/language arts instruction was delivered by one teacher for an entire grade, although, in this school, homeroom teachers held a workshop time that included additional reading/language arts instruction. Therefore, nine teachers were responsible for reading/language arts instruction across three grades in this school.

Because students across grade levels worked with unique combinations of reading/language arts and

homeroom teachers, we identified students' instructional group as the combination of their regular reading/language arts period (i.e., first, second, or third period) plus their workshop period (i.e., homeroom). There were six or seven different instructional groups per grade at this school. Consequently, the present sample included 43 instructional groups (i.e., groups of students who received the same instruction by the same teachers across the school day). As subsequently discussed, analyses accounted for the clustering of students in these groups (see Appendix B for additional information).

Classroom Observations

To capture teacher talk focused on vocabulary and comprehension during reading/language arts periods across the classrooms in the study, we collected audio recordings of instruction and used transcriptions from them as the primary data source. During observations, we simultaneously collected field notes as a secondary data source to provide context on classroom groupings, instructional materials, and nonverbal information (e.g., a picture or graphic organizer presented to the class).

We conducted observations of reading/language arts instruction on three separate days throughout the year (i.e., early winter, late winter/early spring, late spring). At each observation point, researchers collected data during all reading/language arts periods throughout the day. In other observational studies, researchers have analyzed data from two or three timepoints and have found this number to be adequate for capturing typical classroom instruction (e.g., Connor et al., 2011; Connor, Morrison, & Petrella, 2004; Wasik & Bond, 2001). In fact, recent research suggests that teacher instruction is fairly stable across time (Al Otaiba et al., 2008; Smolkowski & Gunn, 2012). Of the scheduled observations, 3% were missed due to scheduling conflicts (e.g., maternity leave) and technical issues (e.g., the audio recorder stopped working).

Collecting the Data

Prior to conducting observations, the first and second authors trained research assistants (RAs) on setting up digital recorders for capturing audio data and taking field notes. The RAs achieved consistency in taking field notes before live classroom observations began. During live observations, the RAs situated digital recorders near classroom teachers to best capture teacher talk and moved the recorders as needed when teachers moved around the classroom. Most instruction was delivered to the whole class, and teachers often stood at the front of the classroom during observations. Thus, there was little movement of teachers around classrooms, and it was not difficult to capture teacher talk on the digital recorders.

When small-group or individual work was observed, the RAs moved digital recorders to follow the teacher and focused field notes on what the teacher was doing. During small-group and individual work, teachers typically walked around the room and provided feedback to students. During guided reading lessons, teachers often taught the same lesson to different small groups of students. Because we observed for a limited amount of time, given that we saw very little differentiated instruction when we observed and considering that we conceptualized teacher talk as representing teachers' instructional focus in general, we decided to apply teachers' instructional talk during small-group or individual work to the whole class with the premise that all students in a given class would be subject to the same kinds and amounts of instructional talk by teachers over time.

Once observations were completed, we had the audio recordings transcribed and merged the field notes' data into the transcriptions for additional context.

Determining the Unit of Analysis

At first, we considered using the teacher turn (i.e., a segment of teacher speech bounded on each side by student speech) as the unit of analysis. However, we discovered that several different types of instruction often occurred within one teacher turn and that coding at the level of the turn appeared to mask the full extent to which teachers provided specific types of instruction. Even applying multiple codes to one turn did not seem ideal in that this approach concealed teachers' relative focus on one type of instruction versus another during the turn. We reasoned that enumerating at a more fine-grain level the extent to which teachers employed specific types of instruction would reveal a more comprehensive and cumulative picture of the teachers' instructional focus.

Practically, regardless of whether we assigned one code or multiple codes to one turn, coding at the level of the teacher turn, when teachers implemented multiple types of instruction in one turn, also made it difficult to establish reliability. Thus, to improve reliability in coding and obtain a fuller picture of teachers' instructional focus, we moved to the level of utterance as the unit of analysis. According to Crookes (1990), an utterance is a unit of speech under a single "breath group" or intonation contour that is bounded by pauses on either side (p. 194). Using this unit of analysis and coding for every teacher utterance, we found that teachers typically exhibited only one type of instruction per utterance, 75% of all utterances across all lessons were teacher utterances, and teacher utterances

were, on average, more than 4 times as long as student utterances.

Developing the Codebook

We employed an iterative process of content analysis to develop the codebook (Krippendorf, 2012; Neuendorf, 2002; Schreier, 2012). We began with a starting list of codes based on research on effective practice. We specifically identified types of instruction from the extant literature base that were associated with student gains in vocabulary and comprehension. Specifically, we derived our starting list of codes for vocabulary from a recent review by Hairrell et al. (2011) on vocabulary intervention, which identified contextual analysis, semantics analysis, explicit instruction, including definitions and examples, and attention to morphology as among the most researched types of instruction for vocabulary.

Similarly, we based our starting list of codes for comprehension on recommendations outlined by Shanahan et al. (2010) and Kamil et al. (2008) that comprehension instruction should include attention to comprehension strategies, text structure, and asking questions to facilitate discussion about text. Finally, given recommendations by Gersten et al. (2007), we included instruction on decoding and fluency as necessary for struggling students to access and comprehend texts.

Using field notes and student utterances for context, we applied these codes to every teacher utterance in every transcript gathered during a pilot year of the observation study. During this phase, the research team met regularly to discuss modifying the codebook to accurately represent the instruction we observed. After the coding of pilot data was complete, and we had a draft of the codebook with descriptions and examples of each code, we applied the codes to data from the present study. Again, the research team met regularly to further refine the codes. Codes were applied, discussed, and reassigned in a repeated iterative process until saturation of coding was reached.

Once this process was complete, the final codebook included four categories of codes: vocabulary instruction, comprehension instruction, other instruction, and noninstruction. These codes represented 12%, 19%, 13%, and 53%, respectively, of teacher utterances overall. Note that only 3% of teacher utterances were unintelligible, and all others received a code. Vocabulary instruction was further coded into one of the following types of instruction: definitions, application across contexts, word relations, morphosyntax, or context clues. Comprehension instruction was further coded into one of these types of instruction: literal comprehension, inferential comprehension, comprehension strategies, text elements, or decoding and fluency.

Note that attention to literal and inferential comprehension, especially through questioning, could be considered instruction related to comprehension strategies. In our exploration of the data, however, we noticed that teachers used attention to literal and inferential comprehension to facilitate discussion of the text at hand, whereas instruction that was focused on previewing, activating prior knowledge, monitoring, visualizing, and summarizing tended to focus more on how to comprehend text in general. Thus, for the purposes of this investigation, we separated attention to literal and inferential comprehension and comprehension strategies instruction.

Similarly, attention to text elements is related to literal and inferential comprehension and comprehension strategies. In our data, however, we found that instruction on text elements appeared distinct from literal and inferential comprehension and comprehension strategies instruction. Therefore, we decided to investigate attention to text elements as a separate focus of instruction. See Appendixes C and D for descriptions and examples of codes and Appendixes E and F for samples of coded transcripts.

Establishing Reliability and Coding the Data

Two doctoral-level RAs with teaching and research experience were trained by the investigators to code the data. After several rounds of practice in applying codes to data from the pilot study, the two RAs independently coded each intelligible teacher utterance in each of three transcripts to establish reliability. Inter-rater reliability (i.e., Cohen's Kappa) was above .80. After reliability was established, the RAs met to discuss and resolve any discrepancies in their coding.

Coding was conducted in three waves. In the first wave, we coded data collected in early winter; in the second, we coded data collected in late winter/early spring; and in the third, we coded data collected in late spring. At the beginning of each wave of coding, the RAs coded every teacher utterance in three identical transcripts from that wave, checked reliability, and discussed discrepancies to mitigate drift in coding. Across all waves, inter-rater reliability was consistently above .80.

Transforming the Codes Into Frequencies

Once the data were coded, we calculated the frequency for each code in each lesson. Although the average length of reading/language arts observations across a day was 60 minutes, some observations were shorter and some longer due to scheduling (e.g., the day of the week on which the observation occurred). To allow for equivalent comparison across classrooms, we prorated all codes to their relative frequency within a standard 60-minute lesson. To do so, we calculated the total

number of codes per classroom for each code, multiplied each code by 60 (i.e., the average number of minutes across lessons), and divided by the number of minutes of the observation. Then, we calculated the average frequency for each code across the two or three observations per classroom. Thus, the instructional data used in analyses represented the average frequency of each instructional code per 60 minutes.

Student Assessments

Observational studies that investigate the connection between teacher instruction and student vocabulary and comprehension typically include only single norm-referenced measures (e.g., Connor, Morrison, & Petrella, 2004). However, research has established that specific measures of vocabulary and comprehension tap different facets of these skills (e.g., Cutting & Scarborough, 2006; Pearson, Hiebert, & Kamil, 2007). The use of multiple norm-referenced measures in latent variables of vocabulary and comprehension arguably provides a fuller representation of the underlying constructs under investigation. Therefore, we included a wide range of norm-referenced vocabulary and comprehension measures and used the data to create latent variables for these constructs.

Trained RAs in both sites administered seven English-language measures in the fall and spring of an academic year. RA administration fidelity was established above .90 on all measures. Unless otherwise noted, all tests were individually administered, and raw scores were used in analyses.

Vocabulary Measures

Research on vocabulary often includes measures of receptive vocabulary and vocabulary breadth (i.e., at least surface-level knowledge of a wide range of words). However, vocabulary knowledge is multidimensional and includes other facets that are often not measured in research (Nagy & Scott, 2000; Pearson et al., 2007). For example, measures of expressive vocabulary and vocabulary depth (i.e., knowledge of relationships among words and morphological variations of words as well as knowledge of how words are used across various syntactical constructions) may be particularly important for comprehension (see, e.g., Proctor, Silverman, Harring, & Montecillo, 2012). Thus, in this study, we used a broad definition of vocabulary and included receptive and expressive measures of breadth and depth of vocabulary in our latent construct. Specifically, we collected data using four measures that tapped different aspects of vocabulary knowledge (i.e., semantic, syntactic, and morphological awareness) to have a maximally inclusive representation of vocabulary in our latent construct.

To assess students' expressive vocabulary breadth, we administered the Woodcock–Muñoz Language Survey–Revised (WMLS–R; Woodcock, Muñoz-Sandoval, Ruef, & Alvarado, 2005) picture vocabulary subtest. We used form A in the fall and form B in the spring. On the picture vocabulary subtest, the RAs prompted students to verbally identify names of pictured objects that increased in difficulty. The internal reliability for students ages 7–13 years on this subtest is .88–.92 (Woodcock et al., 2005). *W*-scores from the subtest were used to anchor the vocabulary latent variable subsequently discussed.

Additionally, based on research that shows the importance of semantic and syntactic aspects of word knowledge (e.g., Proctor et al., 2012), we administered the Clinical Evaluation of Language Fundamentals–Fourth Edition (CELF–4; Semel, Wiig, & Secord, 2003) word classes 2 and formulated sentences subtests. For the word classes 2 subtest, the RAs read aloud a set of four words, two of which were semantically related (e.g., *fence, window, glass, rug*). The RAs prompted students to identify the two semantically related words from each set of four. On the formulated sentences subtest, students were asked to create a sentence to describe a picture prompt using a target word. For example, students were given the target word *forgot* accompanied by a picture of three children dressed for cold weather who are missing some warm articles of clothing. One common response to this prompt is the sentence, "The girl forgot her boots." Responses were scored on a scale of 0–2. A score of 2 represents a semantically and syntactically correct sentence.

Because there is only one available form of the CELF–4, the same form was administered in the fall and spring. Stability coefficients for the word classes 2 receptive subtest range from .65 to .91, and internal consistency (coefficient α) is .73–.84 for students ages 7.0–13.11 years. Stability coefficients for the formulated sentences subtest range from .62 to .77, and internal consistency (coefficient α) for students ages 7.0–13.11 years is .75–.82 (Semel et al., 2003).

Finally, because research has established the importance of knowledge of morphological derivations in reading development (Kuo & Anderson, 2006), we administered the Extract the Base test (Anglin, 1993; Carlisle, 1988; Goodwin et al., 2012). In this assessment, students separate the base of a word (e.g., *elect* from *election*) to complete a provided sentence (e.g., "How many women did they ___?"). Although the examiner read the target word and corresponding sentence aloud, students independently wrote their responses in the blank space. The Extract the Base test was scored using a 0–2 coding scheme, in which 0 indicates an incorrect response, 1 indicates a misspelled but phonologically possible correct response (e.g., *empti* instead of *empty*),

and 2 indicates a correctly spelled and correct response. Like the CELF–4 measures, because the Extract the Base test has only one form, the same form was given in the fall and spring. The most recent validation of the Extract the Base test had a Cronbach's α of .92 and a Rasch person reliability of .86 (Goodwin et al., 2012).

Comprehension Measures

In many studies of reading comprehension, outcomes focus on either the sentence or passage level, and assessments are usually either timed or untimed. To understand content across contexts and under various conditions, students need to be able to comprehend at the sentence and passage levels and on timed and untimed tasks. Because research shows that measures of different aspects of reading comprehension tap different skills (e.g., Cutting & Scarborough, 2006; Keenan, Betjemann, & Olson, 2008; Leider, Proctor, Silverman, & Harring, 2013), we included three different measures of reading comprehension that captured different facets of this construct (i.e., comprehension at the sentence and passage levels and comprehension under timed and untimed conditions using different types of tasks) to more fully represent the range of activities in which students are expected to exhibit comprehension skills in school.

To assess students' sentence- and passage-level comprehension individually and under untimed conditions, we administered the WMLS–R passage comprehension subtest. This subtest consists of increasingly difficult cloze passages. Students read and orally produced the missing word for each passage. Each student response was scored as correct or incorrect, based on whether the response was appropriate in the context of the text. The internal reliability of the passage comprehension assessment for students ages 7–13 years is .80–.94 (Woodcock et al., 2005). *W*-scores from the subtest were used to anchor the comprehension latent variable subsequently discussed.

To capture students' comprehension skills under other conditions (e.g., timed assessments with different task types), we also used two group-administered reading comprehension measures: the Gates–MacGinitie Reading Test–Fourth Edition (GMRT–4; MacGinitie, MacGinitie, Maria, & Dreyer, 2000) reading comprehension subtest and the Test of Silent Reading Efficiency and Comprehension (TOSREC; Wagner, Torgesen, Rashotte, & Pearson, 2010). For the GMRT–4 (form S in the fall and form T in the spring), students read a series of grade-level-appropriate passages and responded to corresponding multiple-choice questions (including explicit and implicit questions), which varied in difficulty. The testing period was 35 minutes. Kuder–Richardson formula 20 reliability coefficients of the GMRT–4 are .92–.93 for grades 3–5, and alternate form reliabilities for the GMRT–4 are .86–.87 for the same grades (MacGinitie et al., 2000).

During the three-minute TOSREC administration (form A in the fall and form C in the spring), students were asked to read and determine whether a series of single sentence items were true or false (e.g., "A doughnut is made of very hard steel."). This measure evaluates students' text comprehension within a timed condition. The TOSREC manual (Wagner et al., 2010) reports high alternate-form reliability for grades 3–5 (r = .82–.96).

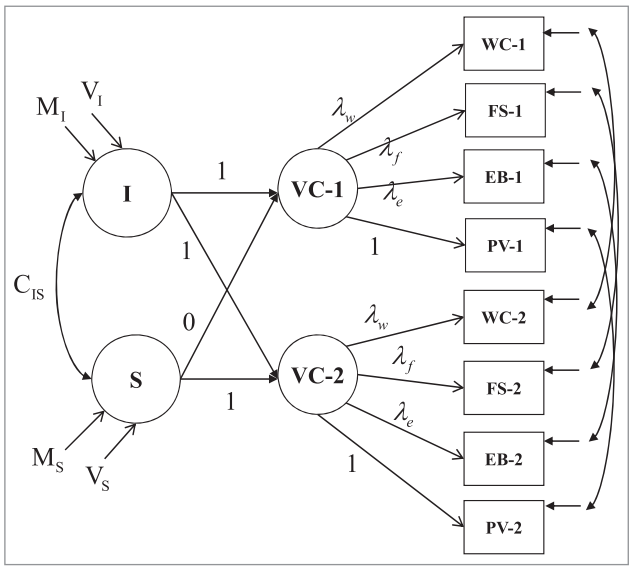
Analysis

To answer our research questions, we conducted separate analyses that investigated the relationship between instruction and fall–spring change in vocabulary and comprehension as separate outcomes. Each analysis was conducted in three steps. First, as an initial exploratory analysis, descriptive statistics for the teacher-level predictors and student-level vocabulary and comprehension outcomes under investigation were computed and examined. Second, latent constructs for vocabulary and comprehension were created for both the fall and spring timepoints, and a sequence of confirmatory factor analyses were conducted to test longitudinal measurement invariance across the two timepoints for each latent variable. Finally, we used an unconditional latent difference score model (McArdle, 2001; Proctor et al., 2012) to predict fall–spring change in student-level vocabulary and comprehension, and we predicted that change using the average observed frequencies of different types of vocabulary and comprehension instruction.

Clustering at the instructional group level (n = 43) was accounted for in the analysis through the use of robust standard errors, which are robust to heteroskedasticity or unequal variances (Muthén & Muthén, 1998–2010). This strategy was employed instead of multilevel or hierarchical linear modeling because there was an insufficient number of classes (level 2 units) to justify its use in providing reliable estimates of classroom variability and the corresponding standard errors of these variance components (Maas & Hox, 2005). Figures 1 and 2 show the schematics of the prototypical latent difference score models for each latent construct—vocabulary comprehension (VC) and reading comprehension (RC)—used in subsequent analyses. The latent difference score models for vocabulary and reading comprehension differ only in the observed vocabulary or comprehension indicators used to characterize the relevant latent construct.

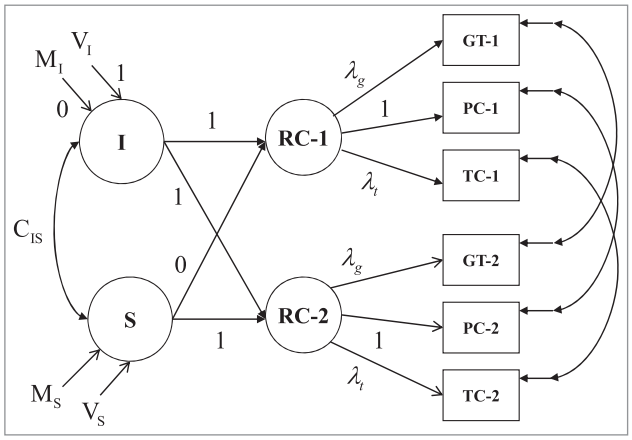
The circles in both figures, labeled VC and RC, represent latent vocabulary and reading comprehension at time 1 and time 2, respectively. These latent variables

FIGURE 1
Path Diagram of the Prototypical Latent Difference Score Model for Assessing Latent Vocabulary



Note. C = covariance. EB = Extract the Base. FS = Clinical Evaluation of Language Fundamentals–Fourth Edition (CELF–4) formulated sentences subtest. I = intercept of the change process. M = mean. PV = Woodcock–Muñoz Language Survey–Revised picture vocabulary subtest. S = slope of the change process. V = variance. VC = latent vocabulary comprehension. WC = CELF–4 word classes 2 subtest. Rectangles denote observed variables, and circles denote latent variables. The circles labeled VC-1 and VC-2 are the latent construct at time 1 and time 2, respectively.

FIGURE 2
Path Diagram of the Prototypical Latent Difference Score Model for Assessing Latent Reading Comprehension



Note. C = covariance. GT = Gates–MacGinitie Reading Test–Fourth Edition comprehension subtest. I = intercept of the change process. M = mean. PC = Woodcock–Muñoz Language Survey–Revised passage comprehension subtest. RC = latent reading comprehension. S = slope of the change process. TC = Test of Silent Reading Efficiency and Comprehension. V = variance. Rectangles denote observed variables, and circles denote latent variables. The circles labeled RC-1 and RC-2 are the latent construct at time 1 and time 2, respectively.

are operationalized vis-à-vis a measurement model, which connects the observed indicators (denoted in rectangles) with the latent variables. The latent variables themselves are indicators of a latent growth process represented by intercept (I) and slope (S). The intercept represents information concerning the mean (M_I) and variance (V_I) of the collection of individual intercepts that characterize each individual's growth curve. The slope growth factor represents the difference in each latent construct from time 1 to time 2. The slope factor has a mean (M_s) and variance (V_s) across the whole sample. The basis terms (factor loadings) of the intercept are fixed at 1. The basis terms of the slope are fixed at 0 and 1 to represent half-year change in each construct. The slope and intercept were allowed to covary (C_{IS}), shown by the double-headed arrow between the two factors. Not shown in the figures are the observed variable intercepts.

The static covariate, language status (e.g., monolingual vs. bilingual), was then added to the unconditional model so we could document any between-group language status differences in intercept and slope and test for interactions between instructional variables and language status on slope. Note that we chose not to include other variables, such as grade level or free or reduced-price lunch, in the model because these variables were not of interest, and we wanted to maintain parsimony, given the complexity of the model.

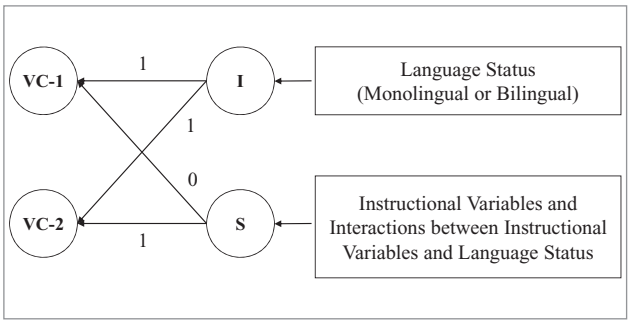
Finally, we executed a series of analyses to evaluate the relationship between instructional variables and change in latent vocabulary and comprehension (see Figures 3 and 4). Specifically, for the latent vocabulary outcome, we examined vocabulary instruction variables, interactions between language status and vocabulary instruction variables, comprehension instruction variables, and interactions between language status and comprehension variables. For the latent comprehension outcome, we investigated comprehension instruction variables, interactions between language status and comprehension instruction variables, vocabulary instruction variables, and interactions between language status and vocabulary instruction variables. As we moved from step to step in the analyses, we carried forward only significant variables. Here we report the final and most parsimonious model for each outcome.

Results

Preliminary Analyses

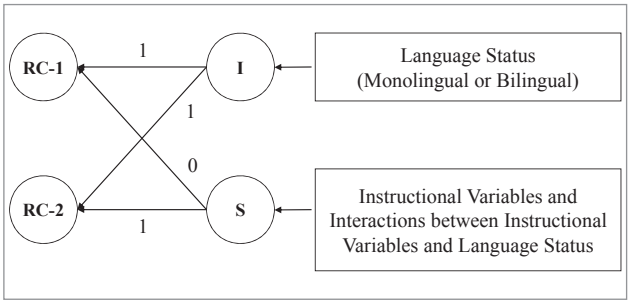
Tables 2 and 3 show descriptive statistics and correlations for the different types of vocabulary and comprehension instruction identified in the study. Of the five types of vocabulary instruction for which we coded, application across contexts was seen most often, and

FIGURE 3
Structural Model of the Change in the Vocabulary Factor



Note. VC = latent vocabulary comprehension. Language status is an explanatory variable of intercept (I), and instructional codes and interactions are explanatory variables of the slope (S).

FIGURE 4
Structural Model of the Change in the Reading Comprehension Factor



Note. RC = latent reading comprehension. Language status is an explanatory variable of intercept (I), and instructional codes and interactions are explanatory variables of the slope (S).

there was more variation across instructional groups in teachers' use of application across contexts. Definitions comprised the secondmost prominent type of instruction observed for vocabulary, and word relations followed definitions in frequency of occurrence. Morphosyntax and context clues were seen least often in the observations we conducted.

Of the comprehension instruction we observed, attention to literal comprehension was seen most often. The next most prominent types of instruction observed were attention to inferential comprehension and comprehension strategies. Attention to text elements also was seen relatively often, but there was little attention to decoding/fluency in the observations that we conducted. In general, comprehension instruction was seen more frequently than was vocabulary instruction. Overall, correlations revealed moderate relationships among most types of vocabulary instruction and most types of comprehension instruction, with some exceptions.

Tables 4 and 5 show the descriptive statistics and correlations of the variables for vocabulary and comprehension. Note that *W*-scores are reported for the picture vocabulary and passage comprehension subtests of the WMLS–R. *W*-scores, developed by Woodcock and Dahl (1971), are derived from an equal interval scale that takes into account item difficulty and ability level using Rasch modeling techniques. In the WMLS–R, raw scores are converted to *W*-scores, from which age and grade equivalents are derived. *W*-scores of about 500 for picture vocabulary and passage comprehension are roughly equivalent to the expected level at the middle of fourth grade, and a 6–8 point difference in *W*-scores is equal to about a year of growth on these measures.

TABLE 2
Means and Standard Deviations of Instructional Codes Across Classes (*N* = 43)

Instructional code	*M*	*SD*
Vocabulary instruction		
Definitions	19.92	14.51
Word relations	8.81	6.95
Application across contexts	30.32	21.72
Morphosyntax	4.86	6.21
Context clues	0.70	1.39
Comprehension instruction		
Literal comprehension	34.39	20.06
Inferential comprehension	26.06	21.52
Comprehension strategies	24.99	19.04
Features of text	22.42	15.12
Decoding and fluency	5.16	8.52

In general, there was improvement across measures from the beginning to the end of the year. Across measures, bilinguals started and ended the year lower, but growth was consistent across both groups of students. Correlations showed positive and moderate relationships among vocabulary variables and comprehension variables, which suggests that the variables are likely related and represent similar underlying constructs but assess different facets of that construct.

Latent Variables and Longitudinal Invariance

In creating latent variables and using them to estimate change across time, it is first necessary to investigate longitudinal measurement invariance, which ensures an equal definition of a latent construct (e.g., vocabulary, reading comprehension) over time. In other words, it is important to confirm that each indicator with the same surface characteristics (i.e., identical scaling and

TABLE 3
Correlations Among Instructional Variables

Instructional code	1	2	3	4	5	6	7	8	9	10
1. Definitions	1									
2. Word relations	.30*	1								
3. Application across contexts	.71*	.48*	1							
4. Morphosyntax	.43*	.01	.33*	1						
5. Context clues	.03	.21*	.03	−.07	1					
6. Literal comprehension	−.09	.16*	.02	.34*	−.13*	1				
7. Inferential comprehension	.52*	.26*	.44*	<.01	.01	.42*	1			
8. Comprehension strategies	.22*	.28*	.20*	−.11~	−.29*	.36*	.44*	1		
9. Text elements	−.15*	.22*	−.19*	−.19*	.08	.46*	.15*	.11~	1	
10. Decoding/fluency	−.11*	−.33*	−.16*	−.14*	−.08	.24*	.08	−.01	.24*	1

*≤.05. ~≤.10.

TABLE 4
Means (and standard deviations [SD]) Across Measures

	Time 1 (beginning of the year)						Time 2 (end of the year)					
	Total sample		Bilingual		Monolingual		Total sample		Bilingual		Monolingual	
Measure	*n*	Mean (*SD*)	*n*	Mean (*SD*)	*n*	Mean (*SD*)	*n*	Mean (*SD*)	*n*	Mean (*SD*)	*n*	Mean (*SD*)
CELF−4 word classes 2 subtest	266	8.99 (3.34)	119	8.39 (3.07)	147	9.48 (3.47)	267	10.01 (3.51)	122	9.18 (3.34)	145	10.72 (3.51)
CELF−4 formulated sentences subtest	266	35.95 (10.81)	119	31.71 (10.63)	147	39.38 (9.72)	266	37.86 (10.29)	122	34.20 (10.59)	144	40.96 (8.95)
Extract the Base	264	36.42 (9.89)	119	34.62 (10.14)	145	37.90 (9.46)	267	39.74 (9.15)	122	37.28 (9.69)	145	41.81 (8.14)
WMLS−R picture vocabulary subtest[a]	265	495.29 (16.67)	119	487.07 (18.99)	146	501.99 (10.59)	267	497.42 (18.09)	122	488.41 (19.35)	145	505.01 (12.77)
TOSREC	270	19.17 (8.68)	123	17.02 (8.76)	147	20.98 (8.20)	263	24.23 (9.98)	122	22.48 (10.68)	141	25.74 (9.11)
GMRT−4 comprehension subtest	267	22.01 (8.96)	123	19.91 (8.32)	144	23.86 (9.11)	254	25.65 (9.94)	118	23.67 (9.50)	136	27.37 (10.03)
WMLS−R passage comprehension subtest[a]	265	487.59 (16.50)	119	482.60 (19.09)	146	491.66 (12.73)	267	490.47 (15.34)	122	487.51 (16.86)	145	492.96 (13.49)

Note. CELF−4 = Clinical Evaluation of Language Fundamentals−Fourth Edition. GMRT−4 = Gates−MacGinitie Reading Test−Fourth Edition. TOSREC = Test of Silent Reading Efficiency and Comprehension. WMLS−R = Woodcock−Muñoz Language Survey−Revised.
[a]*W*-scores.

similar wording) relates to the underlying construct in the same fashion over time (Hancock, Kuo, & Lawrence, 2001). In this study, longitudinal measurement invariance was examined through fitting a series of confirmatory factor analysis models to the data. In fitting the latent difference score models, unique covariances for each variable over time were allowed to covary.

TABLE 5
Correlations Across Measures

Timepoint and measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Time 1 CELF–4 word classes 2 subtest	1													
2. Time 2 CELF–4 word classes 2 subtest	.71*	1												
3. Time 1 CELF–4 formulated sentences subtest	.61*	.63*	1											
4. Time 2 CELF–4 formulated sentences subtest	.62*	.66*	.86*	1										
5. Time 1 Extract the Base	.56*	.60*	.58*	.58*	1									
6. Time 2 Extract the Base	.57*	.62*	.63*	.66*	.85*	1								
7. Time 1 WMLS–R picture vocabulary subtest[a]	.57*	.60*	.67*	.71*	.58*	.64*	1							
8. Time 2 WMLS–R picture vocabulary subtest[a]	.59*	.60*	.69*	.72*	.59*	.64*	.86*	1						
9. Time 1 TOSREC	.52*	.54*	.55*	.54*	.57*	.62*	.51*	.52*	1					
10. Time 2 TOSREC	.60*	.56*	.57*	.60*	.61*	.67*	.51*	.51*	.69*	1				
11. Time 1 GMRT–4 comprehension subtest	.45*	.50*	.50*	.47*	.51*	.54*	.40*	.41*	.61*	.53*	1			
12. Time 2 GMRT–4 comprehension subtest	.47*	.47*	.48*	.48*	.48*	.54*	.41*	.40*	.62*	.58*	.74*	1		
13. Time 1 WMLS–R passage comprehension subtest[a]	.58*	.58*	.67*	.67*	.69*	.73*	.70*	.66*	.59*	.62*	.52*	.55*	1	
14. Time 2 WMLS–R passage comprehension subtest[a]	.58*	.55*	.63*	.66*	.63*	.70*	.63*	.64*	.59*	.62*	.49*	.56*	.70*	1

Note. CELF–4 = Clinical Evaluation of Language Fundamentals–Fourth Edition. GMRT–4 = Gates–MacGinitie Reading Test–Fourth Edition.TOSREC = Test of Silent Reading Efficiency and Comprehension. WMLS–R = Woodcock–Muñoz Language Survey–Revised.
[a]*W*-scores.
*$p < .01$.

Figure 1 shows the model for vocabulary (VC), in which the picture vocabulary subtest was used as the scaling reference for the common factors (VC-1 and VC-2), and the loadings for Extract the Base and the word classes 2 and formulated sentences subtests were constrained to be equal across time. Figure 2 shows the model for comprehension (RC), in which the passage comprehension subtest was used as the scaling reference for the common factors (RC-1 and RC-2), and the loadings for the GMRT–4 and the TOSREC were constrained to be equal across time. Further, observed variable intercepts (not pictured in Figures 1 and 2) for the same indicator were constrained to be equal across time, in accordance with principles of strong metric measurement invariance (Ferrer, Balluerka, & Widaman, 2008).

Given the complex sampling design in the current study (students nested within clusters), models were fitted with the M*plus* estimator for maximum likelihood with robust standard error, in which a scaled version of the chi-square difference test is traditionally used to compare models with increasing levels of measurement invariance (Satorra, 2000). However, other conventional structural equation modeling fit indexes, such as root mean square error of approximation (RMSEA) and comparative fit index (CFI), also were computed to determine goodness of fit of intercepts and slopes of latent vocabulary and comprehension. Hu and Bentler (1999) recommended using a two-index strategy for structural models, where RMSEA values below .08 and CFI values above .95 would be indicative of adequate data–model fit. Table 6 presents model fit statistics and model comparisons using the chi-square difference test.

Based on the scaled chi-square difference test for the latent vocabulary ($\Delta\chi^2 = 18.18$, $df = 10$, $p = .052$) and comprehension ($\Delta\chi^2 = 5.53$, $df = 2$, $p = .062$) variables and structural equation modeling fit indexes for latent

TABLE 6
Tests of Model Comparison and Fit Indexes for Comparing Models Under Increasingly Stringent Measurement Invariance for Change in Latent Vocabulary and Comprehension

Model	Measurement invariance	χ^2	*df*	Model comparison	$\Delta\chi^2$	Δdf	*p*	RMSEA	CFI
Latent vocabulary									
1	Configural	16.3	15					.020	.999
2	Weak metric $\lambda_{q1} = L = \lambda_{qm}$	19.8	18	2 vs. 1	3.15[a]	3	.369	.051	.996
3	Strong metric $\lambda_{q1} = L = \lambda_{qm}$ $\tau_{q1} = L = \tau_{qm}$	40.2	28	3 vs. 2	18.18[a]	10	.052	.067	.988
Latent comprehension									
1	Configural	5.8	5					.031	.999
2	Weak metric	12.3	7	2 vs. 1	5.12	2	.077	.050	.992
3	Strong metric	18.1	9	3 vs. 2	5.53	2	.062	.068	.974

Note. CFI = comparative fit index. *df* = degrees of freedom. RMSEA = root mean square error of approximation. The baseline model on which RMSEA and CFI are based is a model for which the covariances between observed variables across the two timepoints were freely estimated.
[a]Estimation was carried out in *Mplus* 6.2 using the MLR (robust) estimator, and as a consequence, a scaled version of the chi-square difference test was performed.

vocabulary (RMSEA = .067, CFI = .988) and comprehension (RMSEA = .068, CFI = .974), strong metric measurement invariance (e.g., factor loadings and intercepts for the same indicator were constrained to be equal at the two timepoints) was established for both vocabulary and comprehension latent variables, respectively.

The variance in the observed variables accounted for by variability in the latent vocabulary variable (R^2) were moderate to strong, ranging from .52 (time 1 Extract the Base) to .72 (time 2 formulated sentences subtest), and the variance in the observed variables accounted for by variability in the latent comprehension variable (R^2) were moderate to strong, ranging from .49 (time 1 passage comprehension subtest) to .75 (time 2 TOSREC). Confirmatory factor analyses for the latent variables at times 1 and 2 showed distinct latent variables that were related but separable: time 1: $\chi^2(20) = 34.256$, $p = .024$, CFI = .984, RMSEA = .060 [.034, .086], standardized root mean square residual [SRMR] = .029; time 2: $\chi^2(20) = 41.221$, $p = .009$, CFI = .983, RMSEA = .063 [.035, .090], SRMR = .030. The factor correlations were .79 and .75 at times 1 and 2, respectively.

Investigating the Latent Difference Score Model for Vocabulary

Using the latent difference score model previously described, we first tested whether language status (i.e., monolingual or bilingual) was significantly associated with starting points of vocabulary (i.e., intercept), with the results indicating significant differences in favor of monolingual students. We also tested a model to determine whether language status was significantly associated with the time 1–time 2 change and further tested for interactions between instructional variables and language status on slope. Language status was not associated with time 1–time 2 change, nor were any language status × instructional variable interactions. Therefore, we did not include language status in the final model as a slope-related covariate. Vocabulary and comprehension instruction variables at the class level were added as explanatory variables for the slope. The results of the final model are presented in Table 7.

The fit of the final model was adequate (RMSEA = .071, CFI = .959). Based on the way that the language status variable was coded (0 = monolingual, 1 = bilingual), bilingual students displayed initial latent vocabulary skills that were 0.41 of a standard deviation lower than their monolingual counterparts. On slope, controlling for other instructional variables in the model, we found that definitions, word relations, application across contexts, morphosyntax, and literal comprehension were significantly associated with time 1– time 2 change in vocabulary outcomes.

The relationships among definitions, word relations, and morphosyntax with vocabulary change were positive. For definitions ($\hat{\gamma} = .77$, $p = .0010$), a standard deviation increase in the frequency of definitions is related to two-thirds of a standard deviation increase in

TABLE 7
Unstandardized and Standardized Coefficients of Explanatory Covariates From Fitting a Conditional Latent Difference Score Model for Latent Vocabulary Using Maximum Likelihood Estimation With Robust Standard Errors

Parameter	Estimate	Standard error	*p*
Conditional intercept model			
μ_I	499.52	1.42	<.001
γ_{LS}[a]	−.41	.06	<.001
Conditional slope model			
μ_S	3.85	.64	<.001
γ_{DF}[a]	.77	.23	.001
γ_{RL}[a]	.54	.21	.009
γ_{CT}[a]	−.78	.24	.001
γ_{MS}[a]	.33	.16	.037
γ_{LI}[a]	−.51	.16	.001

Note. CT = application across contexts. DF = definitions. LI = literal instruction. LS = language status (i.e., bilingual or monolingual). MS = morphosyntax. RL = word relations. RMSEA = .071, and CFI = .959.
[a]Standardized coefficients.

positive change on latent vocabulary. A standard deviation increase in word relations ($\hat{\gamma}$ = .54, p = .0090) and morphosyntax ($\hat{\gamma}$ = .33, p = .0370) is associated with roughly a half and a third, respectively, of a standard deviation of positive change in latent vocabulary. In contrast, associations between application ($\hat{\gamma}$ = −.78, p = .0010) and literal comprehension ($\hat{\gamma}$ = −.51, p = .0010) and vocabulary change were negative, related to two-thirds and a half of a standard deviation negative change in latent vocabulary, respectively.

Investigating the Latent Difference Score Model for Comprehension

In an analog analysis, a latent difference score model was fitted to the comprehension variables with effects of language status estimated for both the intercept and slope. As with vocabulary, language status was significantly associated with the intercept of reading comprehension, indicating that language status was associated with students' initial level of reading comprehension. Although there was not a significant main effect of language status on slope, which suggests that there was not a relationship between language status and growth in reading comprehension, we found an interaction effect with language status for comprehension strategies (subsequently detailed) that indicated that the time 1–time 2 change for monolingual and bilingual students differed. Thus, the final model for latent comprehension change includes language status on time 1–time 2 change as well as on the intercept. As in the previous analysis, we tested the associations between both vocabulary and comprehension instructional variables with time 1–time 2 change in latent reading comprehension. Table 8 shows both unstandardized coefficients of mean growth parameters and standardized coefficients of the explanatory predictors in the final model.

The fit of the final model was adequate (RMSEA = .075, CFI = .941). Bilingual students displayed initial latent comprehension skills that were 0.28 of a standard deviation lower than their monolingual peers. The final model for change in latent comprehension includes inferential comprehension, which was positive and statistically significant ($\hat{\gamma}$ = 0.47, p = .0450). Accordingly, 1 standard deviation increase in frequency of inferential comprehension instruction is associated with a fourth of a standard deviation increase in growth in latent reading comprehension over the course of the school year.

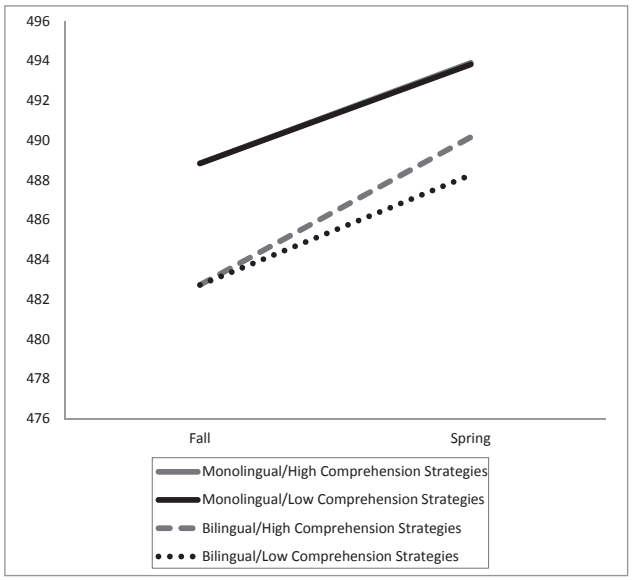
The final model also includes a significant interaction between language status and comprehension strategies ($\hat{\gamma}$ = .79, p = .0230). As can be seen in Figure 5, there is no effect of comprehension strategies for monolinguals, but there is a positive effect for bilinguals such that a higher frequency of comprehension strategies in instruction is related to greater growth in comprehension. For bilinguals, a difference from the first to the third quartile of instruction is roughly equivalent to an increase of 1.5 *W*-score points (i.e., roughly three months of growth, effect size = .07) in growth in comprehension.

TABLE 8
Unstandardized and Standardized Coefficients of Explanatory Covariates From Fitting a Conditional Latent Difference Score Model for Latent Reading Comprehension Using Maximum Likelihood Estimation With Robust Standard Errors

Parameter	Estimate	Standard error	*p*
Conditional intercept model			
μ_I	488.84	1.27	<.001
γ_{LS}[a]	−.29	.08	<.001
Conditional slope model			
μ_S	3.90	.77	<.001
γ_{LS}[a]	−.07	.36	.841
γ_{II}[a]	.47	.23	.045
γ_{CS}[a]	.03	.33	.920
γ_{CS*LS}[a]	.79	.35	.023

Note. . CS = comprehension strategies. II = inferential instruction. LS = language status (i.e., bilingual or monolingual). RMSEA = 0.075 and CFI = 0.941.
[a]Standardized coefficients.

FIGURE 5
The Interaction Effect of Language Status and Comprehension Strategies Instruction on Comprehension, Anchored by the WMLS–R Passage Comprehension W-Scores



Note. WMLS–R = Woodcock–Muñoz Language Survey—Revised. To understand this interaction, we examined prototypical data trajectories for a hypothetical (a) monolingual student with a high frequency of comprehension strategies (i.e., third quartile), (b) monolingual student with a low frequency of comprehension strategies (i.e., first quartile), (c) bilingual student with a high frequency of comprehension strategies (i.e., third quartile) and, (d) bilingual student with a low frequency of comprehension strategies (i.e., first quartile). The monolingual/high-comprehension strategies line is directly under the monolingual/low-comprehension strategies line in the figure.

Discussion

With the added emphasis on vocabulary and comprehension in the CCSS (NGACBP & CCSSO, 2010), it is important to understand relationships between teachers' instruction and students' vocabulary and comprehension to inform teacher preparation and curriculum and intervention development, especially for students who are at risk for experiencing difficulty in reading. In this study, we sought to add to the research base by exploring the relationships between instruction and literacy outcomes among monolingual and bilingual students using a fine-grain unit of analysis (teacher utterance) and robust indicators of students' vocabulary and comprehension.

The results indicated that instruction devoted to definitions, word relations, and morphosyntax had a positive relationship with change in vocabulary, whereas instruction that included application across contexts and a focus on literal comprehension had a negative relationship with change in vocabulary. The findings also showed that attention to inferential comprehension was related to positive change in comprehension and that comprehension strategies instruction was related to positive change in comprehension for bilinguals but not monolinguals. The results showed no association between instruction that targeted context clues, text elements, or decoding/fluency and vocabulary or comprehension. These findings are subsequently discussed in depth.

Findings Related to Vocabulary Instruction

Because observational research has consistently shown the prevalence of definitional instruction in upper elementary school classrooms, we predicted the extensive use of this type of instruction in our study (e.g., Blachowicz, 1987; Scott et al., 2003; Watts, 1995). That we found a positive relationship between definitions and student vocabulary growth was expected. Explicit definitions are a tried-and-true method of vocabulary instruction, and most interventions (with monolingual and bilingual populations) that target vocabulary and have shown positive effects include attention to explicit definitions (e.g., Apthorp et al., 2012; Baumann, Ware, & Edwards, 2007; Beck, Perfetti, & McKeown, 1982; Carlo et al., 2004; Dalton, Proctor, Uccelli, Mo, & Snow, 2011; McKeown, Beck, Omanson, & Perfetti, 1983; Nash & Snowling, 2006; NICHNational Institute of Child Health and Human Development, 2000). The present study provides further evidence of the value of providing definitions as part of vocabulary instruction.

As in the observational research by Scott et al. (2003), we saw relatively frequent attention to word relations, which was positively associated with change in latent vocabulary across the academic year. This positive influence of attention to word relations aligns with recent correlational and intervention research that suggests that understanding relations among words may provide students with leverage as they encounter new words in their environment (Ouellette, 2006; Tannenbaum, Torgesen, & Wagner, 2006). Practically speaking, the present study provides evidence for attention to word relations in classroom instruction for monolinguals and bilinguals alike.

Similar to previous studies (Scott et al., 2003; Watts, 1995), we found limited instructional attention devoted to morphology and syntax. Yet, despite limited evidence of their instructional use, attention to morphology and syntax was related to change in vocabulary. This finding also is aligned with recent intervention and correlational studies on the importance of these skills for monolinguals and bilinguals (e.g., Baumann, Edwards, Bolan, Olejnik, & Kame'enui, 2003; Kieffer & Lesaux, 2012; Kuo & Anderson, 2006; Nagy, Berninger, & Abbott, 2006). These findings suggest that if students are taught to break down words into meaningful word parts and to

analyze how words are used in various syntactic contexts, they may be able to learn new words, improve language proficiency, and perhaps ultimately, use those skills to comprehend text.

Just as other studies have found that teachers regularly attend to application of words across contexts (Scott et al., 2003; Watts, 1995), we found that teachers implemented contextual instruction often. Given that many effective interventions include application of words in various contexts (e.g., Apthorp et al., 2012; Carlo et al., 2004; Taboada & Rutherford, 2011), it was surprising that there was a negative relationship in our study between application across contexts and latent vocabulary change for monolingual and bilingual students. This result may be related to the nature of the application that we observed. Often, teachers intended to support students' understanding of words by providing examples of how words could be used in various contexts and how words could be applied to students' own experiences.

We believe, however, that sometimes this application was not helpful and led students off track from learning the target word. For example, when teaching the word *delivery*, a teacher defined the word and then demonstrated how to apply it by talking about how she ordered pizza the night before and asking students what kind of food they like to order. The discussion veered into the realm of preferred foods rather than to the process of having those foods brought to one's home (i.e., delivered). Students were not encouraged to use the specific word *delivery* in their responses, and the teacher did not bring the discussion back to the word *delivery* and its meaning at the end of the conversation. Although the present study did not investigate the quality of instruction, future research should look more closely at what constitutes a supportive versus a distracting application so teacher preparation and professional development can guide teachers in understanding when and how to focus on application of new words.

Despite the fact that recent intervention research suggests that there is a positive effect of attending to word-learning strategies on students' vocabulary (Baumann, 2005; Nash & Snowling, 2006; Taboada & Rutherford, 2011), we found little attention in our study to the use of context clues to figure out unknown words. The null effect of context clues instruction on vocabulary may be attributable to the fact that instruction on context clues was not often observed. When it was seen, teachers tended to point out that students should use context clues to figure out a new word, without explaining, modeling, and/or guiding how to employ the strategy. It may be that more intensive and structured instruction on context clues, as seen in the work of Baumann et al. (2003), is needed for positive effects of attention to context clues to materialize. Future intervention research should focus on the conditions under which context clues instruction is supportive of students' vocabulary and comprehension development.

In evaluating the relationship between comprehension-related instruction and vocabulary, we found that attention to literal comprehension showed a negative relationship to vocabulary, which suggests that greater use of this type of instruction by teachers is related to decreased change in students' vocabulary. There is little research on the effects of comprehension instruction on vocabulary, although comprehension and vocabulary are considered reciprocal processes (Baumann, 2005). Future research should continue to explore the effect of comprehension instruction on vocabulary and vice versa. That this relationship is negative, and the correlations between the literal comprehension instruction variable and the vocabulary instructional variables are relatively weak, may suggest that if teachers spend too much time on literal comprehension, they may not provide enough support for deeper word learning. More research is needed to understand this relationship.

Findings Related to Comprehension Instruction

For decades, observational and intervention research has demonstrated the benefits of explicit comprehension instruction (e.g., Duke & Pearson, 2002; Palincsar & Brown, 1984; Pearson & Dole, 1987). The present study adds to this research base by its investigation of the relationship between specific types of instruction implemented in natural classroom settings and monolingual and bilingual students' comprehension. When investigating the relationship between instructional variables and change in latent comprehension, we found a positive relationship between attention to inferential comprehension and student change in comprehension. Thinking about what is being implied in text is an important skill for being able to navigate text independently, and one for which teachers should provide guidance.

The positive relationship between instruction focused on inferential comprehension and comprehension outcomes is aligned with research that shows the benefits of interventions, such as Questioning the Author (Beck et al., 1996) and Concept-Oriented Reading Instruction (Guthrie & Cox, 1998), that focus on supporting students in inferencing by guiding them to think deeply about texts and discuss texts with their peers. Because the CCSS (NGACBP & CCSSO, 2010) highlight the need for students to interpret text and find evidence for their interpretations, attention to inferential comprehension will become more critical than ever in the classroom. Importantly, the present study shows that instructional attention to inferential

comprehension in the context of regular school settings is related to student outcomes for monolingual and bilingual students alike.

In addition to the effect of attention to inferential comprehension, we found an interaction between language status (monolingual vs. bilingual) and comprehension strategies instruction such that this type of instruction was related to positive gains in comprehension for bilinguals but not monolinguals. Previous intervention research has shown positive effects of comprehension strategy instruction for both monolinguals and bilinguals (e.g., Guthrie & Cox, 1998; Taboada & Rutherford, 2011); however, there is scant research that compares how comprehension strategy instruction is related to comprehension for monolingual and bilingual students in regular classroom settings. The differential relationship between instruction related to comprehension strategies and change in comprehension found for bilingual learners shows that these learners, in particular, may benefit from explicit attention to strategies that are designed to scaffold text comprehension.

Note that in the present study, we coded teachers' attention to comprehension strategies in general, regardless of whether teachers were implementing explicit comprehension strategy instruction as discussed by Duke and Pearson (2002). However, the present study suggests that even mentioning comprehension strategies during instruction could be important for bilingual students who are developing comprehension skills.

There was no relationship between any of the other comprehension instruction variables and change in comprehension over the course of the academic year. This may be related to how often or simply how the instruction was implemented. For example, attention to literal comprehension mainly followed an Initiate-Respond-Evaluate model (Fisher & Frey, 2007), in which teachers asked questions, students provided brief responses, and teachers countered with quick evaluations before asking more questions. Unlike the "who, what, when, and where" approach (McMaster et al., 2012) of guiding students to find answers to literal questions in texts or the question–answer relationships approach (Raphael & Au, 2005) of supporting students in analyzing question types and finding answers to questions (e.g., in the book, in their head), teachers did not spend time on teaching the process of answering literal questions or remembering explicit details in text.

Although literal comprehension instruction of the type seen in the present study could, theoretically, help with comprehension of the text at hand, the likelihood that this type of instruction would lead to general improvement in answering literal questions across texts is slim. Indeed, one implication of this finding is that perhaps time spent on literal comprehension instruction could be better spent on other, more generative types of instruction, such as instruction on inferential comprehension skills, which was shown to be associated with positive comprehension gains. More research is needed to determine the optimal amount of instruction that should be spent on literal versus inferential comprehension in schools. In addition, examination of the quality of comprehension instruction may provide more information about why such instruction was not related to student comprehension growth.

Another type of instruction that is widely held as supportive of student comprehension is attention to text elements. Intervention research has shown that calling attention to characters and settings in narrative texts and headings and boldface words in informational texts helps students comprehend (e.g., Williams et al., 2005, 2007; Williams, Stafford, Lauer, Hall, & Pollini, 2009). Consistent with previous research (Ness, 2011), teachers in this study focused on identifying the text elements with relative frequency. Teachers did not, however, pay a great deal of attention to explaining how to use text elements to support comprehension.

For example, one activity that we observed was cutting out labels for different informational text features and pasting them to a photocopy of an informational text. Although students may have been able to name the text features when they were done, they may not have known how to use them to support comprehension while reading authentic text. As with other instructional variables in this study, for instruction to show positive effects, it may be that as in the interventions conducted by Williams and colleagues (Williams et al., 2005, 2007; Williams, Stafford, Lauer, Hall, & Pollini, 2009), instruction needs to be more structured, with explicit guidance for learning how and why recognizing text elements can be useful for comprehension.

Finally, although decoding and fluency are essential to comprehension, and interventions that focus on these skills have shown positive effects on comprehension (e.g., Gunn, Biglan, Smolkowski, & Ary, 2000; Kim, Capotosto, Hartry, & Fitzgerald, 2011; O'Connor, White, & Swanson, 2007; Vadasy & Sanders, 2008), there was relatively little instruction on decoding and fluency observed in this study, and instruction in attending to these skills did not show positive relations with comprehension. The limited attention to fluency and decoding instruction is consistent with the recent observational work by Ness (2011) and may be driven by the fact that our observations were conducted in upper elementary–grade classrooms, where decoding and fluency instruction typically begin to taper in their predictive strength relative to comprehension (Gough & Tunmer, 1986; Hoover & Gough, 1990) and where instruction tends to shift in focus to reading for learning (Chall, 1996).

The finding that decoding and fluency instruction had no effect on vocabulary or comprehension for either monolingual or bilingual students may be due to the fact that teachers' attention to the construct in this study was not optimal. Often, teachers prompted students to "sound it out" or "read it with fluency," but they did not give explicit and systematic instruction on how to do this.

Alternatively, it may be that instruction focused on decoding and fluency skills has an impact on decoding and fluency, which, in turn, has an effect on comprehension, but this indirect effect was not explored in this study. Of note, teachers were not observed providing differentiated instruction for students who may have needed extra support with decoding or fluency. Although teachers may believe that attention to decoding and fluency is unnecessary in upper elementary grades, research on late-emerging reading difficulties suggests that many students, both monolingual and bilingual, still struggle with these skills, even in upper elementary grades, especially with multisyllabic words (Kieffer & Vukovic, 2013; Torgesen et al., 2001). Thus, attention to these skills may be needed, especially for the most struggling students. Decoding and fluency instruction could be connected with vocabulary instruction focused on morphology and syntax, but no such connection was seen in this study.

Given the role of vocabulary in comprehension (Baumann, 2005), it is surprising that vocabulary instruction did not relate to comprehension in this study. It could be that the role of vocabulary instruction on comprehension is indirect, an effect that we did not test in this study. It is also possible that the vocabulary instruction that we observed here was not of high enough quality to have an impact on comprehension, although vocabulary intervention research with bilingual and monolingual populations has resulted in similar findings. Proctor et al. (2011), in an online vocabulary and comprehension intervention, found significant effects on both researcher-developed and standardized measures of vocabulary; however, their findings did not hold for standardized comprehension measures. More research is needed on the relationship between the quality and quantity of vocabulary instruction practices and comprehension, under experimental and nonexperimental conditions, in elementary schools.

Findings on Differences for Monolinguals and Bilinguals

On both latent variables, monolinguals performed significantly higher than did bilinguals in the fall. Specifically, monolinguals scored 0.41 of a standard deviation higher on vocabulary and 0.28 of a standard deviation higher on comprehension than did bilinguals. Overall, rates of fall–spring change did not differ for monolinguals and bilinguals, although, as noted, growth for monolinguals and bilinguals differed on comprehension, depending on the amount of one instructional variable, attention to comprehension strategies. The finding that the bilingual students in the sample showed lower levels of initial vocabulary and comprehension than did their monolingual counterparts is a calcified replication that is common to most studies that undertake comparative analysis using standardized measures (e.g., August & Shanahan, 2006).

Further, the finding that, for the most part, instruction is not differentially related to vocabulary and comprehension for monolingual and bilingual students has been well documented in the research base (e.g., Shanahan & Beck, 2006). This study's finding, however, that comprehension strategies instruction may be differentially associated with comprehension for bilingual and monolingual students makes an important contribution to the research base and warrants further investigation.

In general, we witnessed very little instructional differentiation for bilingual students in the classrooms in this study. For example, there was minimal attention to cognates, translation, and nonverbal aids (e.g., gestures, pictures, videos), which have been used in effective intervention studies with bilingual populations (e.g., Carlo et al., 2004; Dalton et al., 2011). Thus, more research is needed on how to support teachers in differentiating instruction and on the effects of such differentiated instruction in regular classroom settings for monolinguals and bilinguals alike.

Limitations and Future Directions

There are several limitations inherent in the present study. First, because the study was correlational in design, we cannot draw causal inferences from the findings. For example, just as students' growth in vocabulary and comprehension could have been a result of teachers' instruction, the instruction could have been responsive to students' growth in vocabulary and comprehension (e.g., Foorman & Schatschneider, 2003). Additional research is needed to clarify the directionality of relationships discovered in this study.

Second, our decision to code teacher utterances could conceal a number of important aspects of teacher instruction. For example, we did not code at the level of the turn or series of turns in teacher–student dialogue, which could reveal important information about how the context and intent of the instruction as well as the depth of instruction relate to student outcomes. Additionally, in choosing to explore teacher utterances, we did not indicate whether the utterance was an explanation, prompt, or feedback. These different types of

teacher utterances could make a marked difference in the effectiveness of instruction (e.g., Hattie & Timperley, 2007; Shute, 2008). For example, telling students about a strategy, asking them to implement a strategy, and giving them feedback on how they implemented a strategy could lead to different results. Finally, we did not code student language or document the nature of teacher–student interaction, both of which could be integral to the extent to which teacher instruction is internalized by students.

Third, in developing a parsimonious coding scheme, we did not include *all* important facets of instruction. For example, research shows that effective vocabulary instruction includes multiple exposures, multiple strategies, and active processing of word meanings (e.g., Hairrell et al., 2011). Yet, we did not capture these aspects of vocabulary instruction in our approach to coding. Additionally, in including individual types of instruction (e.g., monitoring, visualizing, summarizing) in a broader category of instruction (e.g., comprehension strategies), we may have masked important relationships between discrete types of instruction and students' vocabulary and comprehension. Further, because we investigated specific types of instruction as separate, we did not consider how the different types of instruction that we observed overlap and interplay.

Fourth, our resources limited the breadth and depth of observation that we could conduct across classrooms. For example, we did not have the resources to observe individual students, which is ideal for capturing instruction received by specific students with particular characteristics (e.g., Greenwood et al., 2003). This limitation is mitigated, however, by the fact that we observed little differentiation in the target classrooms. Future research should investigate how differentiated instruction in natural upper elementary school classroom settings affects individual students. Additionally, we did not have the resources to observe more than two or three times across the school year. Although we were aligned with previous research that uses only three observations to obtain a snapshot of instruction, we recognize that we were not able to capture the full picture of teacher instruction over the course of the academic year.

Finally, although we focused mainly on the quantity of types of instruction in this study, we appreciate the need for in-depth analyses to investigate the quality of instruction and how it relates to student outcomes for monolinguals and bilinguals alike. Using quality rating scales and qualitative data analysis methods, such as comparative case studies, may shed light not only on which types of instruction are most effective for student vocabulary and comprehension growth but also under what conditions those types of instruction are optimally supportive (e.g., Michener, Sengupta-Irving, Proctor, & Silverman, in press). We believe that correlational studies across a relatively larger number of classrooms, such as the present study, serve to indicate relationships that could be explored in rich qualitative analyses in a comparatively smaller number of classrooms, and as such, we maintain that investigating questions of instruction using different analytic lenses could provide a more robust foundation of research with which to inform teacher preparation and curriculum and intervention development.

Conclusion

To support monolingual and bilingual students in meeting the CCSS, which focus heavily on vocabulary and comprehension, educators need to understand the relationship that instruction, as it is currently implemented, has with students' vocabulary and comprehension. Although more research is needed to provide a comprehensive picture of the current educational landscape and the effects of instruction as it is implemented in everyday classroom settings, the present study adds to the research base by its investigation of the relationships between instructional variables and vocabulary and comprehension outcomes for both monolingual and bilingual students in the current upper elementary school context.

NOTES

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A090152 to the University of Maryland. 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 wish to acknowledge Jennifer Gray, Sarah Zelinke, Catherine Michener, Christine Leider, and Minori Nagahara for their contributions in coding and analysis for this project.

REFERENCES

Al Otaiba, S., Connor, C., Lane, H., Kosanovich, M.L., Schatschneider, C., Dyrlund, A.K., … &Wright, T.L. (2008). *Reading First* kindergarten classroom instruction and students' growth in phonological awareness and letter naming–decoding fluency. *Journal of School Psychology*, *46*(3), 281–314. doi:10.1016/j.jsp.2007.06.002

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

Applebee, A.N., Langer, J.A., Nystrand, M., & Gamoran, A. (2003). Discussion-based approaches to developing understanding: Classroom instruction and student performance in middle and high school English. *American Educational Research Journal*, *40*(3), 685–730. doi:10.3102/00028312040003685

Apthorp, H., Randel, B., Cherasaro, T., Clark, T., McKeown, M., & Beck, I. (2012). Effects of a supplemental vocabulary program on word knowledge and passage comprehension. *Journal of Research on Educational Effectiveness*, *5*(2), 160–188. doi:10.1080/19345747.2012.660240

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

Aukrust, V.G. (2007). Young children acquiring second language vocabulary in preschool group-time: Does amount, diversity, and discourse complexity of teacher talk matter? *Journal of Research in Childhood Education*, *22*(1), 17–37. doi:10.1080/02568540709594610

Baumann, J.F. (2005). Vocabulary–comprehension relationships. In B. Maloch, J.V. Hoffman, D.L. Schallert, C.M. Fairbanks, & J. Worthy (Eds.), *54th yearbook of the National Reading Conference* (pp. 117–131). Oak Creek, WI: National Reading Conference.

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

Baumann, J.F., Ware, D., & Edwards, E.C. (2007). "Bumping into spicy, tasty words that catch your tongue": A formative experiment on vocabulary instruction. *The Reading Teacher*, *61*(2), 108–122. doi:10.1598/RT.61.2.1

Beck, I.L., McKeown, M.G., Sandora, C., Kucan, L., & Worthy, J. (1996). Questioning the author: A yearlong classroom implementation to engage students with text. *The Elementary School Journal*, *96*(4), 385–414. doi:10.1086/461835

Beck, I.L., Perfetti, C.A., & McKeown, M.G. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology*, *74*(4), 506–521. doi:10.1037/0022-0663.74.4.506

Bellack, C., Kleibard, H., Hyman, R., & Smith, F. (1966). *The language of the classroom*. New York, NY: Teachers College Press.

Blachowicz, C.L.Z. (1987). Vocabulary instruction: What goes on in the classroom? *The Reading Teacher*, *41*(2), 132–137.

Bowers, E.P., & Vasilyeva, M. (2011). The relation between teacher input and lexical growth of preschoolers. *Applied Psycholinguistics*, *32*(1), 221–241. doi:10.1017/S0142716410000354

Boyd, M.P., & Rubin, D. (2002). Elaborated student talk in an elementary ESoL classroom. *Research in the Teaching of English*, *36*(4), 495–530.

Brophy, J.E. (1973). Stability of teacher effectiveness. *American Educational Research Journal*, *10*(3), 245–252. doi:10.3102/00028312010003245

Brophy, J.E., & Good, T.L. (1970). Teachers' communication of differential expectations for children's classroom performance: Some behavioral data. *Journal of Educational Psychology*, *61*(5), 365–374. doi:10.1037/h0029908

Calderón, M., Slavin, R., & Sanchez, M. (2011). Effective instruction for English learners. *The Future of Children*, *21*(1), 103–127. doi: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. doi:10.1017/S0142716400007839

Carlisle, J., Kelcey, B., Berebitsky, D., & Phelps, G. (2011). Embracing the complexity of instruction: A study of the effects of teachers' instruction on students' reading comprehension. *Scientific Studies of Reading*, *15*(5), 409–439. doi:10.1080/10888438.2010.497521

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. doi:10.1598/RRQ.39.2.3

Catts, H.W., Compton, D., Tomblin, J.B., & Bridges, M.S. (2012). Prevalence and nature of late-emerging reading disabilities. *Journal of Educational Psychology*, *104*(1), 166–181. doi:10.1037/a0025323

Cazden, C.B. (1998). *Classroom discourse: The language of teaching and learning*. Portsmouth, NH: Heinemann.

Chall, J.S. (1996). *Stages of reading development* (2nd ed.). New York, NY: Harcourt Brace.

Chall, J.S., & Jacobs, V.A. (2003). Poor children's fourth-grade slump. *American Educator*, *27*(1), 14–15.

Chaudron, C. (1988). *Second language classrooms: Research on teaching and learning*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9781139524469

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

Connor, C.M., Morrison, F.J., Fishman, B., Giuliani, S., Luck, M., Underwood, P.S., & Schatschneider, C. (2011). Testing the impact of child characteristics × instruction interactions on third graders' reading comprehension by differentiating literacy instruction. *Reading Research Quarterly*, *46*(3), 189–221. doi:10.1598/RRQ.46.3.1

Connor, C.M., Morrison, F.J., & Katch, E.L. (2004). Beyond the reading wars: The effect of classroom instruction by child interactions on early reading. *Scientific Studies of Reading*, *8*(4), 305–336. doi:10.1207/s1532799xssr0804_1

Connor, C.M., Morrison, F.J., & Petrella, J. (2004). Effective reading comprehension instruction: Examining child × instruction interactions. *Journal of Educational Psychology*, *96*(4), 682–698. doi:10.1037/0022-0663.96.4.682

Connor, C.M., Morrison, F.J., & Slominski, L. (2006). Preschool instruction and children's literacy skill growth. *Journal of Educational Psychology*, *98*(4), 665–689. doi:10.1037/0022-0663.98.4.665

Crookes, G. (1990). The utterance, and other basic units for second language discourse analysis. *Applied Linguistics*, *11*(2), 183–199. doi:10.1093/applin/11.2.183

Cutting, L.E., & Scarborough, H.S. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading*, *10*(3), 277–299. doi:10.1207/s1532799xssr1003_5

Dalton, B., Proctor, C.P., Uccelli, P., Mo, E., & Snow, C.E. (2011). Designing for diversity: The role of reading strategies and interactive vocabulary in a digital reading environment for fifth-grade monolingual English and bilingual students. *Journal of Literacy Research*, *43*(1), 68–100. doi:10.1177/1086296X10397872

Dickinson, D.K., & Porche, M.V. (2011). Relation between language experiences in preschool classrooms and children's kindergarten and fourth-grade language and reading abilities. *Child Development*, *82*(3), 870–886. doi:10.1111/j.1467-8624.2011.01576.x

Duffy, G.G., Roehler, L.R., & Rackliffe, G. (1986). How teachers' instructional talk influences students' understanding of lesson content. *The Elementary School Journal*, *87*(1), 3–16. doi:10.1086/461476

Duke, N.K., & Pearson, P.D. (2002). Effective practices for developing reading comprehension. In A.E. Farstrup, & S.J. Samuels (Eds.), *What research has to say about reading instruction* (3rd ed. pp. 205–242). Newark, DE: International Reading Association.

Durkin, D. (1978). What classroom observations reveal about reading comprehension instruction. *Reading Research Quarterly*, *14*(4), 481–533. doi:10.1598/RRQ.14.4.2

Ferrer, E., Balluerka, N., & Widaman, K.F. (2008). Factorial invariance and the specification of second-order latent growth models. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *4*(1), 22–36. doi:10.1027/1614-2241.4.1.22

Fisher, D., & Frey, N. (2007). *Checking for understanding: Formative assessment techniques for your classroom*. Alexandria, VA: Association for Supervision and Curriculum Development.

Flanders, N. (1970). *Analyzing teacher behavior*. New York, NY: Addison-Wesley.

Fletcher, T.V., Bos, C.S., & Johnson, L.M. (1999). Accommodating English language learners with language and learning disabilities

in bilingual education classrooms. *Learning Disabilities Research & Practice*, *14*(2), 80–91. doi:10.1207/sldrp1402_2

Foorman, B.R., & Schatschneider, C. (2003). Measurement of teaching practices during reading/language arts instruction and its relationship to student achievement. In W.R. Vaughn, & K.L. Briggs (Eds.), *Reading in the classroom: Systems for the observation of teaching and learning* (pp. 1–30). Baltimore, MD: Brookes.

Gersten, R., Baker, S.K., Shanahan, T., Linan-Thompson, S., Collins, P., & Scarcella, R. (2007). *Effective literacy and English language instruction for English learners in the elementary grades: A practice guide* (NCEE 2007–4011). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Goldenberg, C., & Patthey-Chavez, G. (1995). Discourse processes in instruction conversations: Interactions between teacher and transition readers. *Discourse Processes*, *19*(1), 57–73. doi:10.1080/01638539109544905

Goodwin, A.P., Huggins, A.C., Carlo, M., Malabonga, V., Kenyon, D., Louguit, M., & August, D. (2012). Development and validation of Extract the Base: An English derivational morphology test for third through fifth grade monolingual students and Spanish-speaking English language learners. *Language Testing*, *29*(2), 261–285. doi:10.1177/0265532211419827

Gough, P.B., & Tunmer, W.E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, *7*(1), 6–10. doi:10.1177/074193258600700104

Greenwood, C.R., Abbott, M., & Tapia, Y. (2003). Ecological strategies: Observing, measuring, and analyzing behavior and reading interventions. In S.R. Vaughn, & K.L. Briggs (Eds.), *Reading in the classroom: Systems for the observation of teaching and learning* (pp. 53–82). Baltimore, MD: Brookes.

Greenwood, C.R., Arreaga-Mayer, C., & Carta, J. (1994). Identification and translation of effective teacher-developed instructional procedures for general practice. *Remedial and SpecialEducation*,*15*(3),140–151.doi:10.1177/074193259401500302

Greenwood, C.R., Carta, J.J., Arreaga-Mayer, C., & Rager, A. (1991). The behavior analyst consulting model: Identifying and validating naturally effective instructional procedures. *Journal of Behavioral Education*, *1*(2), 165–191. doi:10.1007/BF00957003

Grosjean, J. (2010). *Bilingual: Life and reality*. Cambridge, MA: Harvard University Press.

Gunn, B., Biglan, A., Smolkowski, K., & Ary, D. (2000). The efficacy of supplemental instruction in decoding skills for Hispanic and non-Hispanic students in early elementary school. *The Journal of SpecialEducation*,*34*(2),90–103.doi:10.1177/002246690003400204

Guthrie, J.T., & Cox, K. (1998). Portrait of an engaging classroom: Principles of Concept-Oriented Reading Instruction for diverse students. In K. Harris (Ed.), *Teaching every child every day: Learning in diverse schools and classrooms* (pp. 70–130). Cambridge, MA: Brookline.

Hairrell, A., Rupley, W., & Simmons, D. (2011). The state of vocabulary research. *Literacy Research and Instruction*, *50*(4), 253–271. doi:10.1080/19388071.2010.514036

Hancock, G.R., Kuo, W., & Lawrence, F.R. (2001). An illustration of second-order latent growth models. *Structural Equation Modeling*, *8*(3), 470–489. doi:10.1207/S15328007SEM0803_7

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of EducationalResearch*,*77*(1),81–112.doi:10.3102/003465430298487

Hoffman, J., Maloch, B., & Sailors, B. (2011). Researching the teaching of reading through direct observation: Tools, methodologies, and guidelines for the future. In M.L. Kamil, P.D. Pearson, E.B. Moje, & P.P. Afflerbach (Eds.), *Handbook of reading research* (Vol. *4*, pp. 3–33). New York, NY: Routledge.

Hoover, W.A., & Gough, P.B. (1990). The simple view of reading. *Reading and Writing*, *2*(2), 127–160. doi:10.1007/BF00401799

Hu, L., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55. doi:10.1080/10705519909540118

Kamil, M.L., Borman, G.D., Dole, J., Kral, C.C., Salinger, T., & Torgesen, J. (2008). *Improving adolescent literacy: Effective classroom and intervention practices* (NCEE 2008–4027). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Keenan, J.M., Betjemann, R.S., & Olson, R.K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading*, *12*(3), 281–300. doi:10.1080/10888430802132279

Kelcey, B., & Carlisle, J.F. (2013). Learning about teachers' literacy instruction from classroom observations. *Reading Research Quarterly*, *48*(3), 301–317. doi:10.1002/rrq.51

Kieffer, M.J., & Lesaux, N.K. (2012). Knowledge of words, knowledge about words: Dimensions of vocabulary in first and second language learners in sixth grade. *Reading and Writing*, *25*(2), 347–373. doi:10.1007/s11145-010-9272-9

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.

Kim, J.S., Capotosto, L., Hartry, A., & Fitzgerald, R. (2011). Can a mixed-method literacy intervention improve the reading achievement of low-performing elementary school students in an after-school program? Results from a randomized controlled trial of READ 180 Enterprise. *Educational Evaluation and Policy Analysis*, *33*(2), 183–201. doi:10.3102/0162373711399148

Krippendorf, K.H. (2012). *Content analysis: An introduction to its methodology*. Thousand Oaks, CA: Sage.

Kuo, L.J., & Anderson, R.C. (2006). Morphological awareness and learning to read: A cross-language perspective. *Educational Psychologist*, *41*(3), 161–180. doi:10.1207/s15326985ep4103_3

Leider, C.M., Proctor, C.P., Silverman, R.D., & Harring, J.R. (2013). Examining the role of vocabulary depth, cross-linguistic transfer, and types of reading measures on the reading comprehension of Latino bilinguals in elementary school. *Reading and Writing*, *26*(9), 1459–1485.

Lesaux, N.K. (2006). Building consensus: Future directions for research on English language learners at risk for learning difficulties. *Teachers College Record*, *108*(11), 2406–2438.

Lesaux, N.K., & Geva, E. (2006). Synthesis: Development of literacy in language-minority students. In D. August, & T. Shanahan (Eds.), *Developing literacy in second-language learners* (pp. 53–74). Mahwah, NJ: Erlbaum.

Lindsay, J.S. (1990). Classroom discourse analysis: A review of the literature with implications for educational evaluation. *Journal of Research and Development in Education*, *23*(2), 107–116.

Long, M.H. (1985). Input and second language acquisition theory. In S.M. Gass, & C.G. Madden (Eds.), *Input in second language acquisition* (pp. 377–393). Rowley, MA: Newbury House.

Maas, C.J.M., & Hox, J.J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *1*(3), 86–92. doi:10.1027/1614-2241.1.3.86

MacGinitie, W.H., MacGinitie, R.K., Maria, K., & Dreyer, L.G. (2000). *Gates–MacGinitie reading tests: Forms S and T* (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. doi:0.1037/a0019135

McArdle, J.J. (2001). A latent difference score approach to longitudinal dynamic structural analysis. In R. Cudeck, S.H.C. du Toit, & K.G. Jreskog (Eds.), *Structural equation modeling: Present and future* (pp. 341–380). Lincolnwood, IL: Scientific Software.

McKeown, M.G., Beck, I.L., Omanson, R.C., & Perfetti, C.A. (1983). The effects of long-term vocabulary instruction on reading comprehension: A replication. *Journal of Reading Behavior*, *15*(1), 3–18. doi:10.1080/10862968309547474

McMaster, K.L., van den Broek, P., Espin, C.A., White, M.J., Rapp, D.N., Kendeou, P., … & Carlson, S. (2012). Making the right connections: Differential effects of reading intervention for subgroups of comprehenders. *Learning and Individual Differences*, *22*(1), 100–111. doi:10.1016/j.lindif.2011.11.017

Michaels, S., O'Connor, C., & Resnick, L.B. (2008). Deliberative discourse idealized and realized: AccounTable talk in the classroom and in civic life. *Studies in Philosophy and Education*, *27*(4), 283–297. doi:10.1007/s11217-007-9071-1

Michener, C.J., Sengupta-Irving, T., Proctor, C.P., & Silverman, R.D. (in press). Culturally sustaining pedagogy within monolingual language policy: Variability in instruction. *Language Policy*.

Murphy, P.K., Wilkinson, I.A.G., Soter, A.O., Hennessey, M.N., & Alexander, J.F. (2009). Examining the effects of classroom discussion on students' comprehension of text: A meta-analysis. *Journal of Educational Psychology*, *101*(3), 740–764. doi:10.1037/a0015576

Muthén, L.K., & Muthén, B. (1998–2010). *M*plus *user's guide: Statistical analysis with latent variables*. Los Angeles, CA: Muthén & Muthén.

Nagy, W., Berninger, V.W., & Abbott, R.D. (2006). Contributions of morphology beyond phonology to literacy outcomes of upper elementary and middle school students. *Journal of Educational Psychology*, *98*(1), 134–147. doi:10.1037/0022-0663.98.1.134

Nagy, W.E., & Scott, J.A. (2000). Vocabulary processes. In M.L. Kamil, P.B. Mosenthal, P.D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. *3*, pp. 269–284). Mahwah, NJ: Erlbaum.

Nash, H., & Snowling, M. (2006). Teaching new words to children with poor existing vocabulary knowledge: A controlled evaluation of the definition and context methods. *International Journal of Language & Communication Disorders*, *41*(3), 335–354. doi:10.1080/13682820600602295

National Center for Education Statistics (2011). *The Nation's Report Card: Reading 2011* (NCES 2012–457). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

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

National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups* (NIH Publication No. 00-4754). Washington, DC: U.S. Government Printing Office.

Ness, M. (2011). Explicit reading comprehension instruction in elementary classrooms: Teachers' use of reading comprehension strategies. *Journal of Research in Childhood Education*, *25*(1), 98–117. doi:10.1080/02568543.2010.531076

Neuendorf, K.A. (2002). *The content analysis guidebook*. Thousand Oaks, CA: Sage.

Nystrand, M. (2006). The social and historical context for writing research. In C. MacArthur (Ed.), *Handbook of writing research* (pp. 11–27). New York, NY: Guilford.

Nystrand, M., Wu, L., Gamoran, A., Zeiser, S., & Long, D. (2003). Questions in time: Investigating the structure and dynamics of unfolding classroom discourse. *Discourse Processes*, *35*(2), 135–198. doi:10.1207/S15326950DP3502_3

O'Connor, R.E., White, A., & Swanson, H.L. (2007). Repeated reading versus continuous reading: Influences on reading fluency and comprehension. *Exceptional Children*, *74*(1), 31–46.

Ouellette, G. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology*, *98*(3), 554–566. doi:10.1037/0022-0663.98.3.554

Palincsar, A.S., & Brown, A.L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, *1*(2), 117–175. doi:10.1207/s1532690xci0102_1

Pearson, P.D., & Dole, J.A. (1987). Explicit comprehension instruction: A review of research and a new conceptualization of instruction. *The Elementary School Journal*, *88*(2), 151–165. doi:10.1086/461530

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

Pressley, M., Wharton-McDonald, R., Mistretta-Hampston, J., & Echevarria, M. (1998). Literacy instruction in 10 fourth-grade classrooms in upstate New York. *Scientific Studies of Reading*, *2*(2), 159–194. doi:10.1207/s1532799xssr0202_4

Proctor, C.P., Dalton, D., Uccelli, P., Biancarosa, G., Mo, E., Snow, C.E., & Neugebauer, S. (2011). Improving comprehension online: Effects of deep vocabulary instruction with bilingual and monolingual fifth graders. *Reading and Writing*, *24*(5), 517–544. doi:10.1007/s11145-009-9218-2

Proctor, C.P., Silverman, R.D., Harring, J., & 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. doi:10.1007/s11145-011-9336-5

Raphael, T.E., & Au, K.H. (2005). QAR: Enhancing comprehension and test taking across grades and content areas. *The Reading Teacher*, *59*(3), 206–221. doi:10.1598/RT.59.3.1

Satorra, A. (2000). Scaled and adjusted restricted tests in multi-sample analysis of moment structures. In R.D.H. Heijmans, D.S.G. Pollock & A. Satorra (Eds.), *Innovations in multivariate statistical analysis: A Festschrift for Heinz Neudecker* (pp. 233–247). London, UK: Kluwer Academic. doi:10.1007/978-1-4615-4603-0_17

Schreier, M. (2012). *Qualitative content analysis in practice*. Thousand Oaks, CA: Sage.

Schumm, J.S., Moody, S.W., & Vaughn, S. (2000). Grouping for reading instruction: Does one size fit all? *Journal of Learning Disabilities*, *33*(5), 477–488. doi:10.1177/002221940003300508

Scott, J.A., Jamieson-Noel, D., & Asselin, M. (2003). Vocabulary instruction throughout the day in 23 Canadian upper-elementary classrooms. *The Elementary School Journal*, *103*(3), 269–286. doi:10.1086/499726

Semel, E., Wiig, E., & Secord, W. (2003). *Clinical Evaluation of Language Fundamentals* (4th ed.). San Antonio, TX: PsychCorp.

Shanahan, T., & Beck, I. (2006). Effective literacy teaching for English-language learners. In D.L. August, & T. Shanahan (Eds.), *Developing literacy in a second language: Report of the National Literacy Panel* (pp. 415–488). Mahwah, NJ: Erlbaum.

Shanahan, T., Callison, K., Carriere, C., Duke, N.K., Pearson, P.D., Schatschneider, C., & Torgesen, J. (2010). *Improving reading comprehension in kindergarten through 3rd grade: A practice guide* (NCEE 2010–4038). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Shute, V.J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153–189. doi:10.3102/0034654307313795

Sinclair, J.M., & Coulthard, R.M. (1975). *Towards an analysis of discourse: The English used by teachers and pupils*. Oxford, UK: Oxford University Press.

Smolkowski, K., & Gunn, B. (2012). Reliability and validity of the Classroom Observations of Student–Teacher Interactions (COSTI) for kindergarten reading instruction. *Early Childhood Research Quarterly*, *27*(2), 316–328. doi:10.1016/j.ecresq.2011.09.004

Soter, A.O., Wilkinson, I.A., Murphy, P.K., Rudge, L., Reninger, K., & Edwards, M. (2008). What the discourse tells us: Talk and indicators of high-level comprehension. *International Journal of Educational Research*, *47*(6), 372–391. doi:10.1016/j.ijer.2009.01.001

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. doi:10.1080/02702711003604468

Tannenbaum, K.R., Torgesen, J.K., & Wagner, R.K. (2006). Relationships between word knowledge and reading comprehension in third-grade children. *Scientific Studies of Reading*, *10*(4), 381–398. doi:10.1207/s1532799xssr1004_3

Taylor, B.M., Pearson, P.D., Clark, K., & Walpole, S. (2000). Effective schools and accomplished teachers: Lessons about primary-grade reading instruction in low-income schools. *The Elementary School Journal*, *101*(2), 121–165. doi:10.1086/499662

Torgesen, J.K., Alexander, A.W., Wagner, R.K., Rashotte, C.A., Voeller, K.K.S., & Conway, T. (2001). Intensive remedial instruction for children with severe reading disabilities: Immediate and long-term outcomes from two instructional approaches. *Journal of LearningDisabilities*,*34*(1),33–58.doi:10.1177/002221940103400104

Vadasy, P.F., & Sanders, E.A. (2008). Benefits of repeated reading intervention for low-achieving fourth- and fifth-grade students. *Remedial and Special Education*, *29*(4), 235–249. doi:10.1177/0741932507312013

Veldman, D.J., & Brophy, J.E. (1974). Measuring teacher effects on pupil achievement. *Journal of Educational Psychology*, *66*(3), 319–324. doi:10.1037/h0036500

Wagner, R.C., Torgesen, J.K., Rashotte, C.A., & Pearson, N.A. (2010). *Test of Silent Reading Efficiency and Comprehension*. Austin, TX: ProEd.

Wasik, B.A., & Bond, M.A. (2001). Beyond the pages of a book: Interactive book reading and language development in preschool classrooms. *Journal of Educational Psychology*, *93*(2), 243–250. doi:10.1037/0022-0663.93.2.243

Watts, S.M. (1995). Vocabulary instruction during reading lessons in six classrooms. *Journal of Reading Behavior*, *27*(3), 399–424. doi:10.1080/10862969509547889

Williams, J.P., Hall, K.M., Lauer, K.D., Stafford, K.B., DeSisto, L.A., & deCani, J.S. (2005). Expository text comprehension in the primary grade classroom. *Journal of Educational Psychology*, *97*(4), 538–550. doi:10.1037/0022-0663.97.4.538

Williams, J.P., Nubla-Kung, A.M., Pollini, S., Stafford, K.B., Garcia, A., & Snyder, A.E. (2007). Teaching cause–effect text structure through social studies content to at-risk second graders. *Journal of Learning Disabilities*, *40*(2), 111–120. doi:10.1177/00222194070400020201

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

Woodcock, R.W., & Dahl, M.N. (1971). *A common scale for the measurement of person ability and test item difficulty* (AGS Paper No. 10). Circle Pines, MN: American Guidance Service.

Woodcock, R.W., Muñoz-Sandoval, A.F., Ruef, M.L., & Alvarado, C.G. (2005). *Woodcock–Muñoz Language Survey–revised*. Rolling Meadows, IL: Riverside.

REBECCA D. SILVERMAN (corresponding author) is an associate professor in the Department of Counseling, Higher Education, and Special Education at the University of Maryland, College Park, USA; e-mail rdsilver@umd.edu.

C. PATRICK PROCTOR is an associate professor in the Department of Teacher Education/Special Education, Curriculum and Instruction at Boston College, Chestnut Hill, MA, USA: e-mail proctoch@bc.edu.

JEFFREY R. HARRING is an associate professor in the Department of Human Development and Quantitative Methods at the University of Maryland, College Park, USA; e-mail harring@umd.edu.

BRIE DOYLE and **MARISA A. MITCHELL** are doctoral students and **ANNA G. MEYER** a research associate in the Department of Counseling, Higher Education, and Special Education at the University of Maryland, College Park, USA; e-mail cdoyle12@umd.edu, mamitch@umd.edu, and agmeyer@umd.edu.

APPENDIX A

Demographics on the Districts and Schools in the Study

Demographic category		District 1			District 2		
		School A	School B	School C	School D	School E	School F
Total enrollment		472	384	321	644	432	507
Race/ethnicity	Asian/Pacific Islander	0.10	0.07	0.02	0.11	0.03	0.00
	Black, non-Hispanic	0.09	0.09	0.06	0.45	0.33	0.48
	Hispanic	0.19	0.34	0.58	0.16	0.59	0.41
	Multiracial	0.01	0.03	0.03	0.05	0.02	0.07
	White, non-Hispanic	0.62	0.46	0.32	0.23	0.03	0.03
Additional services	Limited English proficient	0.19	0.13	0.25	0.16	0.43	0.26
	Federal school lunch program	0.26	0.39	0.56	0.35	0.89	0.83

Total enrollment is the number of students enrolled in the school. All other numbers are percentages of the total enrollment number.

APPENDIX B

Classroom Context

Teachers at both sites reported using basal reading programs and supplemental curricula for instruction. The programs included narrative and expository text selections with boldfaced and defined vocabulary and comprehension questions throughout. Teachers primarily used their basal reading programs for text selections rather than lesson planning. The supplemental curricula included explicit instruction on what words mean (although not strategies for figuring out unknown words) and additional practice with comprehension strategies in workbooks. Teachers also reported that they used guided reading (i.e., differentiated small-group instruction) and writers' workshop (i.e., teacher-supported independent writing using the process approach), although these instructional models were not often seen in observations.

As has been found in previous research (e.g., Fletcher, Bos, & Johnson, 1999; Schumm, Moody, & Vaughn, 2000), most (60%) of the reading/language arts instruction observed in the present study across sites, schools, and classrooms was delivered by the instructor to the whole class. Roughly 20% of instruction included small-group work, in which students collaborated on activities introduced during whole-class instruction. Nearly 15% of instruction followed the guided reading format, although teachers typically used the same texts and taught the same content across guided reading groups. Very little (5%) differentiated instruction was observed within the classrooms. Less than 1% of instruction included partner activities. Three classrooms in the Northeastern site were considered Sheltered English Immersion classrooms, which served primarily recently arrived immigrant students. Instruction in these classrooms looked remarkably similar to instruction in the other classrooms despite a programmatic designation that might suggest otherwise.

APPENDIX C

Codebook for Vocabulary Instruction Variables

Variable	Description	Example(s)
Definition	• The teacher provides a brief definition or asks students what a word means. • The teacher asks students to use the dictionary, glossary, or in-text definition to look up a word. • The teacher labels a picture, an object, or a concept.	• "A union is a group that is formed to protect workers in certain fields."
Application across contexts	• The teacher guides or shows students how to apply a word's meaning by providing an example of the word, using the word in a sentence, illustrating the word (drawing or acting it out), or making a connection to the word by using other text, personal experiences, or previous lessons.	• In discussing the word *exhausted*, "OK, so maybe like after you've just worked out really hard, and then you're just like, oh, so tired, and you're at the gym, and you spread out."
Word relations	• The teacher calls attention to the relations among words by discussing synonyms, antonyms, related words, multiple meanings of a word, native-language translations, or cognates of a word.	• "*Confused* is a synonym for *bewildered*." • "The word *explore* has something to do with *adventure*." • "Is a bolt of fabric like a bolt of lightning?"
Morphosyntax	• The teacher calls attention to morphological or syntactical facets of a word. • The teacher may include attending to inflectional endings of words; breaking apart or putting together compound words; attending to prefixes, suffixes, and word derivations; discussing how a word fits syntactically in a sentence; or explaining how to using a word in a grammatically correct way.	• "What does that prefix *sub-* mean?" • "What's the subject of the sentence?"
Context clues	• The teacher teaches or asks students to use an explicit strategy to determine word meaning from clues in the text.	• "There's a context clue right in the text that tells us what a trial is."

APPENDIX D

Codebook for Comprehension Instruction Variables

Variable	Description	Example(s)
Literal comprehension	• The teacher guides or asks students to ask or answer questions about literal details in the text.	• "Then, what was the last thing that she did after she removed the rocks out from in front of the canoe?"
Inferential comprehension	• The teacher guides or asks students to use context clues to figure out the meaning of a sentence or event in the text. • The teacher provides an inference or has students provide an inference about the text.	• "We don't really know why she is. We can just make an inference why maybe based off of your experiences."
Comprehension strategies	• The teacher models or has students use one of the following comprehension strategies: previewing, activating background knowledge/ making connections, monitoring, visualizing, or summarizing.	• "I'm starting to wonder why the author named the title of this chapter 'The Day Things Went Wrong.'" • "Remember, we said yesterday, visualizing is like a movie picture in our minds when we're reading." • "So, you need to stop and tell me what you've read up until that point."

continued

Variable	Description	Example(s)
Text elements	• The teacher guides students to discuss features of text, including story elements (setting, mood, conflict, etc.), genre, organization of text, and text structures (boldfaced font, captions, titles, headings, etc.).	• "The main event is what I really want you to be writing about, some things that happened, like the climax that involves all the characters, involves a problem, a solution, right?"
Decoding/fluency	• The teacher calls students' attention to letter–sound correspondence/phonics skills to read a word, directs students' attention to reading with fluency, or asks students to read with fluency.	• "The /k/ sound just like *career*, *occurred*, and then you have the *U* and the double *R*." • "Remember, stop at a period, you take a short breath, and you keep reading so I know each sentence is a complete thought, all right?"

APPENDIX E

Three-Minute Excerpt of a Coded Vocabulary-Focused Transcript

Speaker	Utterance	Code
Teacher	"The next word is *curiosity*."	Noninstruction
Teacher	"What's the word that this is based on?"	Morphosyntax
Teacher	"*Curiosity* is not the root word."	Morphosyntax
Teacher	"It's not the word that—it's not the word that you build all these words off of."	Morphosyntax
Teacher	"What is the root word for this particular word, *curiosity*?"	Morphosyntax
Teacher	"What's the root word?"	Morphosyntax
Teacher	"What word do you know about that you kind of hear inside *curiosity*, [Student]?"	Morphosyntax
Student	"*Curious*."	
Teacher	"So, what does that mean when you're curious?"	Definition
Teacher	"What do you think, [Student]?"	Noninstruction
Student	"Happy or crazy?"	
Teacher	"No, not happy or crazy so much."	Definition
Teacher	"You know there's an expression that cats are very curious."	Application
Teacher	"You ever heard that expression, cats are very curious?"	Application
Student	"Yeah."	
Teacher	"What does that mean, [Student]?"	Application
Student	"*Curious* means like when you see something and you're curious because you want to know what it is but you don't know."	
Teacher	"For instance, some people could be really curious about how chairs go together with desks."	Application
Student	"Yeah."	
Teacher	"And that might stop them."	Application
Teacher	"People are curious."	Application

continued

Speaker	Utterance	Code
Teacher	"Now, if you want to know about something—for instance, if it is your birthday, and your mom wrote you a note and put it in your lunch on Thursday: 'I have a great present for you when you get home.'"	Application
Teacher	"Do you think you would be curious about that present?"	Application
Student	"A lot."	
Teacher	"Thumb in the sky if you think you would be curious about that present."	Application
Teacher	"Like if before you come to school—before you come to school, your dad says, 'I'm going to get a new car,' would you think about that car all day?"	Application
Teacher	"Would you be curious about what kind of car your father was getting?"	Application
Teacher	"Thumb in the sky if you would be curious about what kind of car your father was getting."	Application
Student	"My brother got a new car like two days ago."	

APPENDIX F

Three-Minute Excerpt of Coded Comprehension-Focused Transcript

Speaker	Utterance	Code
Teacher	"Yesterday, who did we read about?"	Literal comprehension
Teacher	"Raise your hand. Let's see if anybody remembers."	Noninstruction
Student	"Gilbert."	
Teacher	"Chuck Yeager."	Literal comprehension
Teacher	"What kind of attributes do we give to Chuck Yeager?"	Text elements
Teacher	"Raise your hand."	Noninstruction
Teacher	"You said one thing was that he was brave, and why did we say that he was brave?"	Inferential comprehension
Student	"He risked his life so he could blow up and he would go boom."	
Teacher	"OK, what other attribute or characteristic did we give him?"	Text elements
Student	"Intelligent."	
Teacher	"Why did you say that, that he was intelligent?"	Inferential comprehension
Student	"Because he knew that he was hurt, so he just took like a broomstick so he could flip the switch because he knew that his ribs were hurting, and he just lean forward."	
Teacher	"Good."	Noninstruction
Teacher	"He broke his ribs."	Literal comprehension
Teacher	"Did he show people, oh, I have a broomstick? No, what did he do?"	Literal comprehension
Teacher	"Where did he hide that broomstick?"	Literal comprehension
Student	"In the jet."	
Teacher	"OK, what other attribute?"	Text elements

continued

Speaker	Utterance	Code
Student	"Determined."	
Teacher	"Why did you say that?"	Inferential comprehension
Student	"Because he wanted to break the record of sound."	
Teacher	"Good."	Noninstruction
Teacher	"He wanted to break the sound barrier record."	Literal comprehension
Teacher	"How fast was that?"	Literal comprehension
Student	"That might be 61 miles."	
Teacher	"Very good, better than 61 miles, that's how fast sound travels."	Literal comprehension
Teacher	"So, he's gonna go full speed ahead, and he's gonna actually break that sound barrier."	Literal comprehension