International Electronic Journal of Elementary Education, 2011, 4(1), 107-125.

The effects of syntactic and lexical complexity on the comprehension of elementary science texts

Diana J. ARYA*
University of Oslo, Norway

Elfrieda H. HIEBERT
TextProject and University of California, Santa Cruz, United States

P. David PEARSON
University of California, Berkeley, United States

Abstract

In this study we examined the effects of syntactic and lexical complexity on third-grade students' comprehension of science texts. A total of 16 expository texts were designed to represent systematic differences in levels of syntactic and lexical complexity across four science-related topics (*Tree Frogs, Soil, Jelly Beans and Toothpast*e). A Latin-square design was used to counterbalance the order of administration of these 16 texts. After reading each text, students responded to a post-test comprehension measure (without access to the text). External measures of reading achievement and prior vocabulary knowledge were also gathered to serve as control variables. Findings show that lexical complexity had a significant impact on students' comprehension on two of the four topics. Comprehension performance was not influenced by the syntactic complexity of texts, regardless of topic. Further, no additional effects were found for English language learners. Potentially moderating and confounding issues, such as the inference demand of syntactically simple texts and the role of topic familiarity, are discussed in order to explain the inconsistency of the findings across topics.

Keywords: Text complexity, reading comprehension, science literacy

Introduction

Recently, scholars have highlighted the need for increased attention to informational texts in elementary schools, especially primary-level classrooms (Donovan & Smolkin, 2001; Duke, 2000). The argument for this shift in textual diet is that increased attention to informational texts will improve many of the things that matter in students' later development: world

* Diana J. Arya, Norwegian Center for Science Education at the University of Oslo, Norway.
E-mail: d.j.arya@naturfagsenteret.no

ISSN:1307-9298

www.iejee.com

knowledge, monitoring and problem-solving strategies, and dispositions toward academic reading. While all disciplines have benefited from this shift in emphasis, science has received the most attention. For example, in 2010, an entire special issue of the Journal, *Science*, was devoted to the literacy-science interface—a remarkable departure for a public access journal that normally focuses on research and policy for the hard sciences. Science requires numerous firsthand experiences; however, appropriate texts can have a critical role in science learning (Cervetti & Barber, 2009; Guthrie, McRae & Klauda, 2007). Science texts provide readers with a purpose for reading and additional exposure to key science concepts that lead to deeper conceptual understanding (Guthrie, Anderson, Alao, & Rinehart, 1999; Palincsar & Magnusson, 2001; Romance & Vitale, 1992; 2006).

Although the academic benefits of science texts are evident, they pose challenges to teaching and learning. In particular, the vocabulary of science texts can be dense and complex (Armstrong & Collier, 1990; Schleppegrell, 2004; Snow, 2010). Elementary science texts have been criticized for being inaccessible because they introduce the reader to many unfamiliar words yet fail to explain them in ways that connect with students' experiences (Armbruster, 1993; Armstrong & Collier, 1990; Norris & Phillips, 2003; Rutherford, 1991). One of the benefits of having a science text is to help clarify and extend scientific concepts that students encounter during firsthand investigations (Duke & Bennett-Armistead, 2003; Donovan & Smolkin, 2001). However, for young students who are still developing literacy skills, as well as academic vocabulary, science texts containing unfamiliar terms can be very difficult to comprehend.

In the development of student science texts, there is a tension between conceptual explicitness (which often requires more complex syntactical realizations and rare, concept-oriented vocabulary words) and linguistic simplicity (which generally requires less complex syntactic realizations and simpler vocabulary). Matters of syntactic complexity were salient in the text comprehension research of the 1970s and even into the early 1980s (Pearson & Camparell, 1981), but text structure yielded to other emphases, most notably comprehension strategy work, in the 1990s and early 2000s (see Pearson, 2009). The need to re-examine these factors is greater than ever in light of two recent developments. First the dramatic increase in the numbers of students with diverse linguistic and cultural backgrounds (US Census, 2000) and the challenges many linguistically diverse students experience on tasks such as the NAEP Science assessment (Gutierrez & Rogoff, 2003; Lee, & Luykx, 2005; Shaw, 1997) requires us to take a closer look at text features that may prove especially challenging or supportive for English language learners. A major challenge is identifying text features that make information more accessible for ELLs. Second, the advent of the new Common Core State Standards in English language arts (CCSS, 2010) has upped the ante on standards of text complexity; educators are being challenged by these standards to increase the complexity of texts students read at every grade level by at least a half grade in measured readability. This means that all students are going to be asked to read texts with more complex syntax and more difficult vocabulary. In this study, we investigate the extent to which lexically and syntactically complex realizations of content hinder or help comprehension—and whether these two factors interact to provide either unique scaffolds or barriers to acquiring important science concepts.

Gauging Text Complexity

Syntactic Complexity. The "simple" view of syntactic complexity evident in readability formulas such as the *Flesch Ease of Reading* formula (Flesch, 1948) holds that the fewer words in a sentence, the less difficult it is for readers to comprehend (Klare, 1984). This perspective, however, may be misleading; more words may simply be an alias for more ideas or, even more likely, more complex ideas. In other words, conceptual complexity could be driving difficulty, and the number of words in a sentence simply indexes (rather than causes) that complexity. In psychological terms, the longer the sentence, the greater the likelihood that multiple discrete ideas, called propositions, are embedded in it (Kintsch, 1998). Examples 1-3 illustrate this point.

1. Tree frogs have red eyes.
2. Tree frogs have red eyes that help them see and find food.
3. Tree frogs have red eyes that help them see and find food at night.

Example 1 conveys two complete ideas or propositions: (a) tree frogs have eyes, and (b) these eyes are red. Example 2 has two additional propositions: these red eyes help the frog to (a) *see* food and (b) to *find* food. Example 3 actually adds three more propositions, one explicit and two implicit. The explicit proposition is that the frogs do the seeing and finding at night. That proposition invokes two more entailments: that (a) that frogs are *awake* at night and (b) that their eyes help them to see *in the dark*. The amount and explicitness of the information provided in each sentence increases as the number of embedded structures (e.g., adjectives, relative clauses, and prepositional phrases) increases. Readers must be able to unpack the propositions within complex sentences and establish their logical relations to one another to understand all of the information presented.

Any account of text difficulty that uses sentence length to establish the readability of texts assumes, at least implicitly, that unpacking the propositions within a complex sentence is more difficult than making connections across related propositions stated in simple sentences. A short sentence in itself may be easier to comprehend than a complex one. However, the challenge may come when the reader needs to construct a cohesion model of meaning from a series of short sentences. To illustrate this distinction, a complex sentence such as the third example above can be broken up into five simple sentences, as in Example 4:

4. Tree frogs have eyes. These eyes are red. These eyes help them see. They help them find food. The tree frogs are awake at night.

Just as the complex sentence required readers to unpack propositions within the sentence, having to connect ideas across discrete, simple sentences may place other task demands on readers. Connective cues (e.g., conjunctives, conjunctive adverbs, and relative clauses) and other embedded structures serve as markers to guide readers to a full understanding of the ideas presented. Eliminating these connective cues may increase the inference burden on readers (Bowey, 1986; Pearson & Camperell, 1981); relationships, such as cause-effect or problem-solution, or sequence, that were explicitly cued in the more complex versions have to be inferred in the less complex versions. Ozuru, Dempsy, Sayroo and McNamara (2005) found that adding cohesive devices such as connectives that made relationships between sentences more explicit were beneficial for students reading science texts about unfamiliar topics. Students were able to correctly answer more questions when texts had syntactic structures that made meaning more explicit than when texts were less cohesive. Similarly, Rawson (2004) found that texts that presented more ambiguous syntactic structures with unmarked, reduced relative clauses (*the girls told about the movie were excited*) were more difficult for college students (all of whom had high reading abilities) than texts with more explicit structures, containing marked clauses (*the girls who were told about the movie were excited).*

While some complexity in sentences can support readers in comprehending text, presumably there is a point where sentences can become too complex for novice or inexperienced readers. Several factors likely influence where this tipping point occurs. Developmental level and reader proficiency appear to be two such factors in that older readers and more proficient readers demonstrate greater comprehension of grammatically complex structures than younger and less proficient readers (Nation & Snowling, 2000; Willows & Ryan, 1986).

Background knowledge or conceptual familiarity of the topic may also influence readers' abilities to comprehend the embedded structures of complex sentences. Goldman and Bisanz (2002) reported that novice and less proficient readers who did not have background knowledge of a topic were less able than more knowledgeable or more proficient readers to avail themselves of embedded structural cues. However, McNamara, Kintsch, Songer, and Kintsch (1996) found that science texts containing greater number of embedded structures that clarify or highlight information (e.g. use of connectives or embedded explanations) benefit readers with less knowledge of the concepts whereas texts with "cohesive gaps" (i.e., fewer connectives and embedded clauses) that require students to make inferences about relationships and concepts benefit readers with a strong level of prior knowledge (McNamara, 2001; McNamara et al., 1996); in short, less knowledgeable readers are aided by the scaffolding of explicit cues but more knowledgeable readers are aided by the challenge of a text that needs "fixing". The current study attempts to address this conflict. Thus when it comes to the issue of syntactic complexity, a trade-off may well exist: What is made more easily accessible by complexity (seeing the relations among propositions) is made more esoteric by simplicity. What is made easier to comprehend by simplicity (getting unitary ideas through the veil of working memory) is rendered complex by the addition of embedded structures.

Lexical Complexity. Sentence length, typically an alias for syntactic complexity, is often, indeed almost universally, coupled with vocabulary difficulty in readability formulas in order to determine overall accessibility of a given text (Flesch, 1948, 1979; Lennon & Burdick, 2004). Vocabulary difficulty is generally indexed by how frequently a particular word generally appears in texts (Zeno, Ivens, Millard, & Duvvuri, 1995). The assumption is that the more exposures a reader has to a particular word, the more a reader learns about it and, in turn, the more accessible that word (and the message in which it is embedded) becomes. Indicators of familiarity have long been used to estimate the readability of text (e.g., Cunningham & Stanovich, 1998; Snow & Sweet, 2003; Stahl, 1999).

Word frequency is strongly correlated with word knowledge, which is a crucial aspect of reading comprehension (NICHD, 2000; RAND Reading Study Group, 2002); simply put, the more frequently a word occurs in a language the greater the likelihood that students will know its meaning. Research on vocabulary suggests that texts containing few unknown words provide readers with an appropriate source from which to develop fluency and word knowledge (Beck & McKeown, 1991; Qian, 2002; Vellutino, 2003). Thus, the more students read texts with very few rare words, the greater their chances in developing a solid understanding of the unfamiliar concepts that are present, allowing them to comprehend texts with additional lexical complexity in the future (Nagy & Scott, 2000; Stanovich, 2000). However, it may be argued that the more frequent a word, the greater the likelihood that, while students will know its meaning, its meaning may be less precise (Carey, 1985; Gopnik, 1996). This may especially be so with words in science where a less frequent word such as *astronaut* conveys a level of precision that a generic word like *man* does not. Conversely, too many unfamiliar or complex vocabulary words within science texts may inhibit readers' ability to learn concepts through reading (Shymansky, Yore, & Good, 1991; Stahl, 1999). We expect students to infer word meanings from context; it is a required part of skilled, strategic reading. However, if there are too many unknown words in the surrounding context, there may be no meaning base from which a student could infer the meaning of a particular word. Contrast the challenge of inferring the meaning of *habitat* in Examples X and Y:

X. The soil in the alluvial plane, rich in nutrients and decomposers, provided an optimal habitat for our earthworms.

Y. The soil along the river provided a good habitat for our earthworms.

Science texts are purported to have more than twice the number of rare words as texts from any other discipline, thus creating a vexing challenge for developers of science literacy curricula: How can they create considerate and accessible texts for young readers that also do justice to the concepts students are supposed to acquire (Hayes & Ahrens, 1988)? Just as with syntactic complexity, there is a potential trade off in lexical complexity. Rare words have a level of precision that high frequency words do not. However, the presence of too many rare words may make a text inaccessible to readers.

Vocabulary familiarity (complexity) has a direct relationship to readers' knowledge about the topic, which has a great impact on comprehension (Kintsch, 1998; RAND Reading Study Group, 2002; Smagorinsky, 2001; Snow & Sweet, 2003; Stahl, 1999). As one becomes more familiar and experienced with a topic, knowledge of contextualized meanings of words develops as well (Anderson & Freebody, 1981; Kintsch, 1998). In experiments that use association and priming tasks, skilled readers have been found to approach a text with an organized network of knowledge called schemata. These allow readers to integrate new information with prior knowledge (Kintsch, 1998; RAND Reading Study Group, 2002; Smagorinsky, 2001; Snow & Sweet, 2003) and, in the process, enhance their schemata even more. The stronger one's prior knowledge about a particular subject, the greater one's ability to read and comprehend texts quickly and efficiently (Kintsch, 1998). The connections that readers make with text are dependent on their knowledge base and ability to retrieve the most relevant meaning from alternatives in their mental lexicons (Kintsch, 1998; Smagorinsky, 2001; Wilson & Sperber, 1987).

Just as students' prior knowledge about particular concepts facilitates comprehension, a lack of knowledge about concepts within a text can have a detrimental impact on understanding. Bailey (2007) conducted a language analysis of American standardized achievement tests and found that academic language (i.e., words often used in tests such as *examine* or *cause*) confounds the ability of English Language Learners (ELLs) to demonstrate their understanding of the construct that is being assessed in English. Similarly, Droop and Verhoeven (1998) found in their study of third grade students learning Dutch as a first or second language that lexical complexity (defined in terms of word frequency) as well as cultural relevance impacts text comprehension. However neither of these studies examined the impact of syntactic complexity or its interaction with lexical complexity in academic language.

The Current Study

The aim of the present investigation was to compare the effects of syntactic and lexical complexity on students' understanding of science content. Students' comprehension of texts was examined as a function of two dimensions of syntactic complexity (simple, complex) and two dimensions of lexical complexity (simple, complex); additionally, the main and interaction effects of syntactic and lexical complexity were examined through the lenses of reading ability and prior knowledge.

Language status was also considered as a potential confounding factor on the comprehension of these texts. Text accessibility is an important issue for ELLs because they must have the opportunity to read extensively in texts at their level of reading ability in order to improve comprehension and fluency (Cunningham & Stanovich, 1998; Elley, 1996; Grabe, 1991; Snowling & Nation, 1997). However, few studies of readability have investigated the effects of text difficulty on the comprehension of ELLs. Further, no such study has focused on both lexical and syntactic complexity while holding issues of cultural relevance constant.

Specifically, the following questions are addressed in this investigation:

1. Do syntactic and lexical complexity affect comprehension of science texts for third graders?
2. How do these two forms of complexity interact to produce unique combination effects on comprehension?
3. Are there any additional effects of syntactic and lexical complexity for ELLs?

We anticipated that, the greater the complexity of a given science text (as measured by embedded clauses and difficult vocabulary), the more skilled a reader must be to successfully understand the text. Thus, scores on general reading assessments such as informal reading inventories and state tests should predict scores on an assessment of comprehension of science content. Based on a long history of readability research, we also hypothesized that lexical complexity might have a greater impact on performance than syntactic complexity, but that the interaction of syntactic and lexical complexity would have the most debilitating effect on comprehension. In short, only the very best readers as defined by reading test scores would be able to handle the difficulty imposed by texts that are complex on both syntactic and lexical criteria.

Questions about the manner in which prior knowledge of words and lexical complexity influence students' comprehension and how these constructs contrast with syntactic complexity merit particular attention with science texts. It is possible that, when complex ideas are communicated with accessible (i.e., high frequency) vocabulary, syntactic complexity does not matter as much as it does when technical (low frequency) vocabulary is used. Further, a reader's experience with certain subject matter may determine the degree to which lexical complexity, syntactic complexity, or both affect understanding.

Method

Participants

This study included all 142 third-graders who had returned parental consent in 10 classrooms in four non-charter public schools. According to California state regulations operating at the time of the data collection, no K-3 classroom could enroll more than 20 students. An average of 14.3 students per classroom (minimum = 12, maximum = 19), which was approximately 75% of total third grade enrollment across these schools, participated in the study. All four schools were located in northern California and varied according to urbanicity, ethnicity, and percentage of ELLs, defined by language spoken at home. The reason for this ELL distinction is that three of the participating schools had no English language program, thus having no school language designation for students who are learning English as a second language. This information was reported by the participants and confirmed by the parental consent letters. Students who spoke only English at home were not considered to be ELL. All other students (49 students, 34% of total participants) were considered to be ELL. Of the 49 ELL students, 23 (16%) spoke only Spanish at home while 11 (8%) spoke a mix of English and Spanish. The remaining 15 students (11%) spoke one of various Asian or European languages at home.

Two of the four participating schools (3 classrooms total) were urban, while one was located within a suburban area (one classroom) and one is rural (6 classrooms total). The four schools ranged in percentage of ethnic minority (i.e., other than White, 44%--73%) as well as language minority (other than English, 12%--59%) students. Half of the total population of participants were Caucasian (71 students) while the other half represented Hispanic (44 students, 30% of total participants), African American (15 students, 10% of total participants), and Asian (eight students, 6% of total participants) ethnicities with a remaining few (five students, 4% of total participants) representing other or mixed ethnicities. Note that SES was not reported for this study: as of 2005, it is illegal to access this particular demographic information according to California state regulations, even for classroom teachers, regardless of consent or approval of the university's internal review board.

In the initial phase of the study, each participant read a narrative passage orally from the *Qualitative Reading Inventory (QRI)* (Leslie & Caldwell, 2000) and answered questions about it. Students who were not able to read at least 75% of the narrative text (five participants total) were excused from continuing on with the assessment procedure in order to prevent potential frustrations. Thus, a total of 142 participants continued with the study. Participants demonstrated a broad range of abilities in fluency (words read within a minute, WPM) and comprehension (number of correct responses to questions about the passage reading) on the *QRI*. The mean WPM performance was 97 with a standard deviation of 36. The mean comprehension score was 5.8 (based on a total of 8) with a standard deviation of 1.7.

Assessments

Assessments were administered across three sessions that spanned a three-week period. In the first session, the QRI and a measure of students' knowledge on the specific topics of the experimental texts were administered. In the second and third sessions occurring three weeks later, students were given the experimental texts. Two additional measures of student reading achievement were obtained: (a) teachers' ranking of student reading proficiency and (b) student scores on the Standardized Testing and Reporting Program (STAR) (California Department of Education, 2007) from the previous spring. The mean score of the STAR was 351 with a standard deviation of 62. All assessments and scores described here were used to establish a baseline of reading abilities for all participants.

Qualitative Reading Inventory (QRI). As already described, students individually read a third-grade, narrative passage of the *QRI* and answered explicit and inferential questions about it (Leslie & Caldwell, 2000). A student's oral reading of the text was timed and miscues recorded. The oral readings and responses to comprehension questions were tape-recorded to establish fidelity of different investigators' on-the-spot recording of miscues. The authors of the QRI report very high alternate form reliability ($r = .9$) as well as high correlation with an unidentified standardized reading test ($r = .7$). This form of assessment was used not only for its reliability, but also for the fact that participating teachers and students are familiar with this more qualitative format of assessment. Thus, the teachers could also use student performance on the QRI formatively for general educational purposes.

Prior vocabulary knowledge. A prior knowledge measure was developed by identifying six words for each of the four topics that were the focus of the experimental portion of the study: tree frogs, toothpaste, jelly beans, and soil. Sixteen of the words in the 24-item measure represented highlighted science concepts in the lexically academic forms of the experimental texts (four items per topic). All words were within the same general range of frequency, from 46 to 53 on the SFI index (Zeno et al., 1995). The remaining eight items consisted of either words representing science concepts that were not part of the experimental texts (e.g., *terrarium)* and or cross-disciplinary words (e.g., *determine*) that were not in the experimental texts but were within the same range of frequency. This second group of words was included in order to obtain a measure of general lexicon as well as of the specific topics in the study.

The 24 words were then randomly organized into six groups. A student's task was to match a word with its definition for each group of words. Definitions were short, everyday descriptions of the words, such as *to dig* for *burrow*. This task was not timed and was completed in small, investigator-supervised groups.

Teacher rankings of students. Teachers were asked to rank students, beginning with 1 (the strongest reader in the classroom). Teachers completed these rankings without receiving feedback on students' performances on the *QRI* or the prior knowledge measure.

State assessment. Students' performances on the state's Standardized Testing and Reporting Program (STAR) (California Department of Education, 2007) from the end of the prior academic year were obtained as an external measure. This measure was used as a covariate with the *QRI* to establish external validity of the vocabulary pre-assessment and comprehension measure of the experimental texts. Forty-two scores were missing due to record unavailability (15 total) and missing teacher files (two of the six classroom teachers within the rural school were unable to locate scores).

Experimental texts. Sixteen texts of approximately 200 words in length were written, with four versions of each of four different science topics. Topics were identified from the national science education standards (NRC, 2001) to represent the three strands of life, earth, and physical science.

The creation of the experimental texts began with a single text for each topic. This initial text had three sections: (a) an introductory section of 50 words that was common across all conditions, (b) a manipulated section of approximately 100 words (within a 9-word range) that differed according to condition, and (c) a concluding section of 50 words that was common to all conditions. The introductory and concluding sections of the text used "simple" syntactic forms and "everyday" lexical content.

Table 1. *Example of Manipulated Version of Topical Texts: Jelly Beans*

	Everyday Vocabulary	*Academic Vocabulary*
Syntactically Simple	A scientist wanted to make a new flavor. He wanted to make grass flavor. Grass is not safe food. He could not use real grass. He used other things. These things are safe. His new jelly bean smells like grass. It tastes like grass.	One scientist wanted to **invent**[1] a flavor. This was grass flavor. Grass is not **edible.** He could not **manufacture** the flavor. He used different **ingredients.** This jelly bean had the **odor** of grass. It had the taste of grass.
Syntactically Simple	A scientist wanted to make a new flavor. He wanted to make grass flavor. Grass is not safe food. He could not use real grass. He used other things. These things are safe. His new jelly bean smells like grass. It tastes like grass.	One scientist wanted to **invent**[1] a flavor. This was grass flavor. Grass is not **edible.** He could not **manufacture** the flavor. He used different **ingredients.** This jelly bean had the **odor** of grass. It had the taste of grass.
Syntactically Embedded	One scientist wanted to make a new flavor, grass flavor, *by*[2] using other things *because* grass is not safe to eat. He could not use real grass to make the flavor, *but* it smelled *and* tasted like grass.	One scientist wanted to invent a flavor, grass flavor. by using different ingredients because grass is not edible. He could not use grass to manufacture the flavor, but this jelly bean had the odor and taste of grass.

[1]Academic vocabulary
[2]Word indicating an embedded structure

Table 2. *Indexed Features of Syntactic and Lexical Complexity*

	Syntactic Complexity (Average Number of Propositions within Version)		*Lexical Complexity (Average Standard Frequency Index (SFI) within Version)*	
	Simple	Embedded	Everyday	Academic
Tree Frogs	2.9	7.3	65.9	52.1
Soil	2.9	7.3	63	48.3
Jelly Beans	2.6	7	67	47.5
Toothpaste	2.7	6.8	63.4	47

The middle section of the text was rewritten so that there were four texts for each topic: (a) syntactically simple with everyday vocabulary (simple/everyday), (b) syntactically complex with everyday vocabulary (embedded/everyday), (c) syntactically simple with academic vocabulary (simple/academic), and (d) syntactically complex with academic vocabulary (embedded/academic).

For this study, a high level of syntactic complexity was defined as the presence of two or more embedded structures within a sentence; sentences with one or no embedded structures were deemed as low in syntactic complexity. Embedded structures included relative clauses, nominalizations, appositives and multiple modifiers. An illustration of the "treated" portion of a text and the nature of embedded structures appears in Table 1.

A propositional analysis (Kintsch, 1998) was used to determine the difference between syntactically simple and complex texts. The average number of propositions per sentence for the simple and embedded versions of the texts is summarized in Table 2. Across the four topics, the difference between the simple and complex version is consistently about 4 propositions per topic.

Lexical complexity was indexed by the presence or absence of academic (cross-disciplinary or scientific) words that directly relate to science concepts or processes and are beyond the 1000 most frequent words according to Zeno et al. (1995). High-frequency words (words within 1000 most frequent) are referred to as "everyday words." To verify the differences across these passages, the standard frequency index (SFI) of the words in each passage were computed. The higher the SFI, the more frequently the word is used in texts (e.g., *the* = 88.3; *sanitize* = 25.6). The average SFIs for academic and everyday versions across the four topics are reported above in Table 2. Averaged across the four topics, the difference between the mean SFI values of the academic and everyday versions was 16 (the difference between the mean SFI value of each of the individual topics were within three points of this). Since the remaining portions of the texts (i.e., the first and last 25% of each text) are equivalent, and since the function words for all manipulated versions are high in frequency, the focus of the analyses was on the everyday and complex version of academic words.

For each topic, 10 questions were constructed to measure students' comprehension. Half of the questions were multiple-choice and half required short-answer responses. An example of a multiple-choice question is the following: *What makes plants grow? a. rocks; b. vitamins; c. bugs; d. wind.* The short-answer responses were constructed to elicit a specific response, such as the following: *Write two ways that animals help plants*. The questions for a given topic were the same, regardless of the manipulated condition that students received. Four of the 10 questions targeted the content of the manipulated portion of the text; the remaining six questions referred to the first and last 25% portion of the text (three questions for each portion). Two of the four questions for the treated portion were explicit recall of information from the text and two required the student to make inferences based on what they read from this portion. The remaining six questions also consisted of both direct recall and inferential questions.

The short-answer questions (e.g., *How do frogs get away from their enemies?*) were scored on a scale of 0-1-2. A rubric was constructed to assign no, partial or full credit. No credit was given to responses that were irrelevant (e.g., *they like to swim*). Partial credit was given to responses that included part of the intended answer (e.g., *they hop around*). Full credit was given to complete and accurate answers (e.g., *they hop around really fast*). A sample of 20% of the responses was double-scored; the inter-rater agreement was 95%.

Reliability and validity of measure

All experimenter-designed assessments were piloted to determine validity and reliability. After revision, the prior knowledge assessment had a Cronbach's alpha coefficient of .85 and correlated strongly with the QRI timed miscue measure (.65, $p < .01$), teacher ranking of reading ability (.57, $p < .01$) and performance on the STAR (.67, $p < .01$).

The comprehension assessments for the experimental texts on the four topics, *Tree Frogs, Soil, Jelly Beans* and *Toothpaste,* had a Cronbach's alpha coefficient of .86. These comprehension assessments strongly correlated with the state reading assessment (.56, .67, .74, .63; $p < .01$) and the QRI timed miscue measure (.51, .50, .51, .51; $p < .01$).

Procedures

Three experienced researchers collected all of the data for the present study. To reduce the possibility of priming the participants on key vocabulary, the prior knowledge measure was administered individually in one session, along with the QRI task, at least three weeks prior to the experimental reading task. The passage reading/comprehension tasks took place in two sessions as whole-class events on two separate days; each of these sessions lasted approximately 50 minutes.

All participants read four passages with the constraint that each student received each topic and each version once and only once. There were 4 topics and 4 versions per topic, yielding 16 unique reading tasks (a passage followed by the comprehension items connected with that particular topic). These reading tasks were assigned to participants using a Latin-square design, which resulted in complete counterbalancing for the order in which both topic and version were presented. In other words, each of the 16 reading tasks was completed equally often in the first through fourth testing positions across students. To avoid fatigue, participants completed two reading tasks on the first day and two on the second day of testing. As an example, one student might have read *Tree Frogs* in the syntactically simple/everyday vocabulary version and *Soil* in the syntactically simple/academy vocabulary version on day one, followed by *Toothpaste* in the syntactically complex/everyday vocabulary version and *Jelly Beans* in the syntactically complex/academically vocabulary version on day two. It required a total of 64 participants to complete one complete replicate of the 4 topics X 4 versions X four serial testing positions design.

Participants were given as much time as needed to read the text and then answer the questions, but each text was collected directly before distributing questions. They were required to answer each set of questions based on memory of what had been read, without the opportunity to look back at the text. Tables 3a and 3b show the total performance on each text by version and topic as well as specific performance on only the treated portions.

Results

A series of 2-step (students were level 1 and classrooms, level 2), hierarchical linear models were fit to the data to examine the relationship between treatment (syntactic and/or lexical complexity) and performance on the treated sections of the text, while simultaneously accounting for variance due to the clustering of students within classrooms. A random intercept was included in the model; it permitted different mean performance levels across classrooms. No random slopes were included in this model due to the small number of classrooms ($N = 10$) as well as the implausibility and irrelevance of classroom-specific effects of treatment on performance. No additional classroom variables were considered in the present study. Such analyses, which would have allowed for more level-2 covariates, would have required a much larger sample of classrooms than was available.

Error-variance histograms revealed that the error variance from each of the regression models fit was normally distributed. Also, predicted-versus-observed scatterplots of the outcome variables revealed that the error variance was constant across the range of data. Thus, the assumptions of regression modeling were met for the data used in this study.

This study uses a modest form of HLM, with a random intercept only and no level-2 covariates. In Raudenbush and Bryk's (2002) notation, our full model (which corresponds to Model 3 described below) is described by this formula:

Table 3a. *Means and SDs for Total performance on Designed Texts*

Topic→	*Tree Frogs*	*Soil*	*Jelly Beans*	*Toothpaste*
Version 1 (simple/everyday)	10.6 (3.3)	9.5 (3.3)	6.2 (3.6)	9.7 (3.2)
Version 2 (complex/everyday)	10.7 (3.4)	8.7 (3.1)	5.1 (3.6)	9.7 (3.2)
Version 3 (simple/academic)	10 (3.2)	7.5 (3)	6.3 (2.8)	10.2 (3.7)
Version 4 (complex/academic)	9.6 (3.1)	7.3 (2.6)	6.1 (2.9)	9.6 (3.1)

Table 3b. *Means and SDs for Treated Portion of Designed Texts*

Topic→	*Tree Frogs*	*Soil*	*Jelly Beans*	*Toothpaste*
Version 1 (simple/everyday)	4.4 (1.7)	3.4 (1.6)	2.2 (1.8)	3.4 (1.4)
Version 2 (complex/everyday)	4.5 (1.6)	3.2 (1.5)	1.8 (1.6)	3.5 (1.2)
Version 3 (simple/academic)	4.0 (1.7)	2.6 (1.7)	2.1 (1.5)	3.9 (1.6)
Version 4 (complex/academic)	3.5 (1.7)	2.6 (1.6)	2.2 (1.5)	3.5 (1.3)

In the present study, the model form described above was fit four times, once for each of the four topics. Although the multiple models were fit using the same participants, a Bonferonni-like correction was not applied in this situation given that the same question was asked four times, once for each topic. Naturally, we hoped that results from the four model sets would converge.

The first model fit (Model 1) is a variance-components model with no covariates and is presented to illustrate the amount of total variance in performance that can be attributed to classroom-level effects. Model 2 adds the control variables, and Model 3 adds the independent variables. Since the interaction between syntactic and lexical complexity was not significant, it was dropped for the final model (Model 4). This variance components model indicates that a significant amount (6.4%, $p < .05$) of variation in performance is between-classrooms. Since the various text conditions were assigned randomly to students within classrooms, it was important to control for classroom-level effects in order to accurately assess treatment differences within all ten classrooms included in the analysis.

Model 2 adds in the covariates, which are home language (i.e., ELL status) and four pretest scores (STAR from grade 2, prior vocabulary knowledge, and the fluency and comprehension scores for the 3rd grade QRI passage). Pretest scores were a highly significant predictor of performance; ELL status was not, after controlling for pretest scores. Thus, ELL status did not explain any additional variance in performance on the designed texts. The random intercept variance remained significant, but its share of the variance was reduced greatly in comparison to Model 1, indicating that much of the variance between classrooms is attributable to student background characteristics and prior achievement.

Model 3 adds in the independent variables: presence of syntactic complexity, presence of lexical complexity, and an interaction term between the two. All three of these variables were non-significant. We then dropped the interaction term from the model, leaving model 4, in which lexical complexity affected performance but syntactic complexity did not.

Model 4 explains a significant amount of variance for only two of the four topics, *Tree Frogs* and *Soil*. Similar results were not obtained for *Jelly Beans* and *Toothpaste*; for the latter two topics, neither lexical nor syntactic complexity affected performance.

This final model suggests that high lexical complexity (i.e., more low frequency words) in the text is associated with lower performance on the test ($p < .05$). As would be predicted by the design of the passages, the impact of lexical complexity was limited to items in the middle 50% of the passage (the manipulated portions); lexical complexity did not explain any significant portion of variance in responses for comprehension items relating to the first and final sections of the texts. A model with only syntactic complexity as a predictor variable was also fit to the data, but was not significant at the 0.05 level. Model 4, with lexical complexity predicting comprehension differences across forms, is presented in Table 4 for all four topics.

These inconsistent results prompted a series of post-hoc investigations into the particular conditions under which lexical complexity of a text may affect comprehension of that text. The most obvious candidate to explain the inconsistent patterns is background knowledge of particular concepts across the four topics. The knowledge of concepts explanation was explored in two ways. The first was an examination of the SFI indices of frequency from the Zeno et al.'s (1995) corpus; these data appear in Table 2. Differences between the SFIs for the academic and everyday versions of the texts for the four topics were calculated. The observed average SFI differences between levels of lexical complexity, which were (in order of magnitude), *Jelly Beans*: 19.5; *Toothpaste*: 16.8; *Soil*: 14.7; and *Tree Frogs*: 13.8, would have predicted the greatest between-version differences in comprehension on the *Jelly beans* and *Toothpaste* passages. Ironically, just the opposite pattern was evident in the data, with the greatest differences between academic and everyday versions on *Soil* and *Tree Frogs*, the two topics with the smallest differences between the everyday and academic versions. Thus, SFI index does not provide a suitable explanation for the apparent interaction between topic and lexical complexity.

The second way in which background knowledge was considered was to examine the relationship of the prior knowledge vocabulary measure to comprehension of the topics. Recall that the prior knowledge vocabulary measure correlated strongly with students' comprehension of the manipulated portions of the texts: *Tree Frogs*: .52; *Soil*: .59; *Jelly Beans*: .65; *toothpaste*: .67 ($p < .01$). The mean scores (out of a maximum of 4) and standard deviations of the prior vocabulary assessment items for the four topics are as follows: *Tree Frogs:* 2.3 (*sd*, 1.3); *Soil*: 1.7 (*sd*, 1.3); *Jelly Beans*: 2.9 (*sd*, 1.1); *Toothpaste*: 2.8 (*sd*, 1.0). When the simple effects were calculated across these four means, the analysis showed that "academic vocabulary" used to create the complex versions of the passages yielded significantly different pre-test vocabulary results across the four topics. The pre-test academic vocabulary performances for *Toothpaste* and *Jelly Beans*, which did not differ from one another, were significantly easier than either *Soil* or *Tree Frogs*; additionally, *Tree Frogs* was easier than *Soil* ($p < .01$, in all cases); in sum: (*Jelly Beans= Toothpaste*) > (*Tree Frogs* > *Soil*). Thus, the empirical measure of students' prior knowledge of words was a more accurate predictor of lexical complexity than the SFI index. It is the only plausible explanation of the differential effect of lexical complexity across topics.

Discussion

The present study was designed to address the question of whether lexical or syntactic factors exert greater influence on the comprehension of elementary science texts. Based on previous research on text accessibility, it was expected that syntactic and lexical complexity would each affect students' performance on science texts, and that these two types of text complexity together would additionally impact student performance. In order to test this hypothesis, 16 texts that varied in syntactic and lexical complexity across four different topics were constructed. Students read texts that ranged in complexity, each from a different topic.

Contrary to our hypotheses, syntactic complexity did not explain variance in performance across any of the four topics. It is difficult to interpret our results on syntactic complexity. As established in the review of research on this topic, opinions are divided as to whether or not explicitness, as defined by embedded clauses and connective cues, hinders or aids comprehension. It is possible that different sorts of cognitive loads effectively canceled out differences between the syntactically simple and complex versions: our syntactically simple versions required students to engage in a great deal of inferencing to create the logical links between sentences (e.g., A caused B or A happened before B). By contrast, the syntactically complex versions required readers to hold many embedded constructions and cues in short term memory to unpack those logical links. However, since reading ability (as measured by the QRI and STAR test) and the prior knowledge assessment did not interact with syntactic complexity, it is difficult to sort out what was happening across levels of syntactic complexity. We certainly were not able to replicate the McNamara et al (1996) finding of an interaction between students' level of prior knowledge and the cohesion of the texts as indexed by strong use of cohesive ties between clauses and sentences. Future studies might include gradations of syntactic complexity in order to begin to unpack this mystery. The other possibility is that the methodology used for measuring comprehension obscured the real impact of syntax. It may be that syntax achieves its effect on comprehension in the "search" process readers engage in when they consult the text to find exact answers to explicit questions or clues to help them draw inferences. By taking away the texts during the comprehension assessment, we may have pre-empted the very mechanism (text search) through which syntactic explicitness achieves its effect.

Table 4. *Regression Results without Interaction Terms (Model 4)*

Predictor	Topic 1 (*Tree frogs*)	Topic 2 (*Soil*)	Topic 3 (*Jelly Beans*)	Topic 4 (*Toothpaste*)
Intercept	1.82** (.52)	1.23** (.34)	.80 (.53)	3.77** (.64)
Home language	-.47 (.34)	-.46 .(23)	-.53 (.35)	-.61 (.47)
QRI (pretest)	.46** (.07)	.27** (.05)	.24** (.08)	.54** (.10)
Syntactic complexity	.06 (.25)	-.22 (.17)	.06 (.26)	-.46 (.35)
Lexical complexity	-.55* (.24)	-.54** (.16)	.13 (.25)	.36 (.34)
Variance component of:				
Classroom mean, u_{ij}	.044*	.185*	.086*	.228*
Level-1 effect, r_{ij}	1.999	.854	2.103	3.599

Note: The results represent a set of non-nested multilevel models, fit to the same participants using different topics. Standard errors are given in parentheses.
* $p < .05$; ** $p < .01$.

Lexical complexity significantly influenced comprehension performance for texts on two of the four topics, *Tree Frogs* and *Soil*, but not for texts on *Jelly Beans* and *Toothpaste*. This finding was consistent across all participant groups, including ELLs. A possible explanation is that prior knowledge of vocabulary, rather than any established index of word frequency, determines how difficult a lexically complex text will be for a student. Although, for example, *bacteria* is considered a very low frequency word, 62% of the participants were able to correctly identify its meaning. Further, *essential*, a word with a comparable SFI value to *bacteria* (SFI=56) was a much less familiar word, at least for our sample of students, in that less than half (42%) of the students were able to correctly identify its meaning. Assuming that world and word knowledge is shaped by experience, it is plausible to assume that most eight year olds (the average age of our sample) have visited the dentist several times and have learned about dental hygiene, including words such as *bacteria*. The role of conceptual familiarity as a predictor of text comprehension has been commented upon in previous research (Cunningham & Stanovich, 1998; Kintsch, 1998; Smagorinsky, 2001; Snow & Sweet, 2003; Stahl, 1999), thus giving strength to this admittedly speculative explanation for the interaction between topic familiarity and lexical complexity. However, it is important to note that our explanation of the inconsistent lexical complexity effect are tentative at best and require further investigation. Future studies on the effects of lexical complexity should include measures of students' prior knowledge in order to assess conceptual familiarity adequately.

A specific interest in the present study was the effect of variations in text complexity on the comprehension of ELLs. Language status did not explain any additional variance in performance beyond the general findings in this study. Thus, lexical complexity was the only significant factor in comprehension performance for ELLs. This finding is consistent with research by Proctor, August, Carlo and Snow (2005) who reported that L2 vocabulary knowledge was a significant predictor of L2 text comprehension of ELLs. Our findings did not reveal any significant differences in comprehension performance between native English speakers and ELLs, thus suggesting a global model of comprehension, seemingly contrary to Proctor et al.'s (2005) conclusion that we need an L2-only model of comprehension. However, due to differences in specific information about L1 proficiency, comparisons between this study and the work by Proctor et al. are speculative at best.

While the results of this study are intriguing, it is important to note significant limitations. First, the manipulated portions of the experimental texts (approximately 100 words in each of the 200-word texts) may not have been long enough to allow for the detection the effects of syntactic and lexical complexity across all four topics. Additionally, the fact that ELL status was dichotomously classified (ELL or non-ELL) could limit our ability to explore the effect of first language (L1) expertise on performance. A multitude of studies highlight the significant effects of L1 proficiency on L2 acquisition and comprehension (Jimenez, Garcia, & Pearson, 1996; Proctor et al., 2005). The hypothesis that students' command over their first language influences their ability to comprehend both syntactically and lexically complex features of texts was not considered in the present study. Further research is needed to determine possible effects of varying gradations in L1 proficiency on L2 text difficulty.

The findings within the present study have left questions regarding text accessibility unanswered. Does syntactic complexity have absolutely no effect on comprehension, or is there some gradation of difference that was not captured within the design of our texts? Does prior knowledge, as defined by conceptual familiarity, trump lexical complexity, as indexed by frequency, in determining comprehension? If so, how much familiarity is necessary to overcome difficult vocabulary? Finally, do EL learners face the same difficulties as native English speakers in terms of text accessibility, even when considering the effect of gradations in L1 proficiency? We hope that future studies will shed further light on these important questions. At the same time, our failure to elicit a syntactic complexity affect might give us pause, when we design curriculum, of being too rigid about keeping sentence length to an absolute minimum. Further, the lexical complexity effect, which seemed to be most powerful in situations in which students could not rely on prior knowledge from everyday experiences, merits attention for all students who struggle with unfamiliar content when reading in disciplinary settings.

• • •

Diana J. Arya is currently a research fellow working with the Norwegian Center for Science Education at the University of Oslo, Norway. The central focus of her research is on the quality of school science texts and on methods for improving this quality. Arya is a recent graduate from Berkeley (PhD) with additional degrees from Illinois (MA) and Michigan (MA). Before joining the academic research community, Arya was a reading specialist and middle school science teacher in Seattle, WA.

Elfrieda H. Hiebert is President and CEO of TextProject, Inc., a not-for-profit aimed at increasing student-reading levels through appropriate texts as well as a research associate at the University of California, Santa Cruz. Her research focuses on text design and instruction around the core and extended vocabularies of written English. She began her educational career as a teacher in California and, after receiving the Ph.D. in Educational Psychology from the University of Wisconsin-Madison, taught at the Universities of Kentucky, Colorado-Boulder, Michigan, and California, Berkeley.

P. David Pearson is a Professor in the Language, Literacy and Culture program in the Graduate School of Education at the University of California, Berkeley, where he pursues research on assessment, instruction, and curriculum reform in literacy. With colleagues at Lawrence Hall of Science, he is building and validating an integrated science and literacy curriculum for grades K-8. With degrees from Berkeley (BA) and Minnesota (PhD) and professorial stints at Minnesota, Illinois, and Michigan State, he began his career as a 5th grade teacher in Porterville, CA.

References

Anderson, R.C., & Freebody, P. (1981).Vocabulary knowledge. In J.T. Guthrie (Ed.), *Comprehension and teaching: Research reviews* (pp.77–117). Newark, DE: International Reading Association.

Armstrong, J.E. & Collier, G.E. (1990). *Science in biology: An introduction*. Prospect Heights, IL: Waveland Press.

Armbruster, B. (1993). Science and reading. *The Reading Teacher*, 46(4), 346-347.

Bailey, A.L. (2007). Introduction: Teaching and assessing students learning English in school. In A. L. Bailey (Ed.). *Language Demands of Students Learning English in School: Putting academic language to the test.* New Haven CT: Yale University Press.

Beck, I.L., & McKeown, M. (1991). Conditions of vocabulary acquisition. In R. Barr, M.L. Kamil, P. Mosenthal, & P.D. Pearson (Eds.), *Handbook of reading research* (Vol., II, pp. 789-814). White Plains, NY: Longman.

Bowey, J.A. (1986). Syntactic awareness in relation to reading skill and ongoing reading comprehension monitoring. *Journal of Experimental Child Psychology,* 41, 282-299.

California Department of Education (2007). *The California Standardized Testing and Reporting (STAR) Program*. Retrieved January 30, 2007 from http://star.cde.ca.gov/star2006/

California State Board of Education (April 17, 2006). *Criteria for evaluating instructional materials (Reading/Language Arts).* Retrieved January 30, 2007 from http://www.cde.ca.gov/ci/rl/im/

Carey, S. (1985). Are children fundamentally different thinkers than adults? In S. Chipman, J. Segal & R. Glaser (Eds.), *Thinking and learning skills* (pp. 436-517). Hillsdale, NJ: Lawrence Erlbaum.

Cervetti, G. and Barber, J. (2009). Bringing back books: Using text to supplement hands-on investigations for scientific inquiry. *Science and Children*, *47*(3), 20-23.

CCSS (2010). Common Core State Standards. Retrieved September 14, 2011, from http://www.corestandards.org/.

Cunningham, A. & Stanovich, K.E. (1998). What reading does for the mind. *American Educator,* 22(1& 2), 8-15.

Donovan, C.A., & Smolkin, L.B (2001). Genre and other factors influencing teachers' book selections for science instruction. *Reading Research Quarterly*, 36 (4), 412-440.

Droop, M. & Verhoeven, L. (1998). Background knowledge, linguistic complexity and second language reading comprehension. *Journal of Literacy Research,* 30(2), 253-271.

Duke, N. (2000). 3.6 minutes per day: The scarcity of informational texts in the first grade. *Reading Research Quarterly,* 35*,* 202-224.

Duke, N. K., & Bennett-Armistead, V. S. (2003). *Reading and writing informational text in the primary grades: Research-based practices.* New York: Scholastic.

Elley, W. (1996). Using book floods to raise literacy levels in developing countries. In V. Greaney (Ed.), *Promoting reading in developing countries: Views on making reading materials accessible to increase literacy levels* (pp. 148-163). Newark, DE: IRA.

Flesch, R. (1979). *How to write plain English*. New York, NY: Harper and Row.

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology, 32*, 221–23.

Goldman, S.R., & Bisanz, G. L. (2002). Toward a functional analysis of scientific genres: Implications for understanding and learning processes. In J. Ortero, J.A. Leon, & A.C. Graesser (Eds.), *The psychology of science text comprehension* (pp. 19-50). New Jersey: LEA.

Gopnik, A. (1996). The scientist as child. *Philosophy of science*, 63(4), 485-514.

Grabe, W. (1991). Current developments in second language reading research. *TESOL Quarterly*, 25(3), 375-406.

Gutierrez, K. , & Rogoff, B. (2003). Cultural ways of learning: Individual traits or repertoires of practice. *Educational Researcher*, 32(5), 19 - 25.

Guthrie, J. T., Anderson, E., Alao, S., & Rinehart, J. (1999). Influences of Concept- Oriented Reading Instruction on strategy use and conceptual learning from text. *Elementary School Journal, 99(4*), 343-366.

Guthrie, J. T., McRae, A. C., & Klauda, S. L. (2007). Contributions of Concept-Oriented Reading Instruction to knowledge about interventions for motivations in reading. *Educational Psychologist, 42,* 237-250.

Hayes, D. P., & Ahrens, M. (1988). Vocabulary simplification for children: A special case of 'motherese.' *Journal of Child Language,* 15*,* 395-410.

Jimenez, R.T., Garcia, G.E., & Pearson, P.D. (1996). The reading strategies of bilingual; Latina/o students who are successful English readers: Opportunities and obstacles. *Reading Research Quarterly,* 31, 90-112.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. NY: Cambridge University Press.

Klare, G.R. (1984). Readability. In P. D. Pearson (Ed.), *Handbook of Reading Research* (Vol. 1, pp. 681-744). New York: Longman.

Lee, O., & Luykx, A. (2005). Dilemmas in scaling up innovations in science instruction with nonmainstream elementary students. *American Educational Research Journal,* 42(5), 411–438.

Lennon, C. & Burdick, H. (2004). The lexile framework as an approach for reading measurement and success. MetaMetrics.

Leslie, L., & Caldwell, J. (2000). *Qualitative Reading Inventory-III*. New York: Longman.

McNamara, D. S. (2001). Reading both high-coherence and low coherence texts: Effects of text sequence and prior knowledge. *Canadian Journal of Experimental Psychology,* 55, 51-62.

McNamara, D.S., Kintsch, E., Songer, N.B., & Kintsch, W. (1996). Are good texts always better? Text coherence, background knowledge, and levels of understanding in learning from text. *Cognition & Instruction,* 14*,* 1-43.

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. III, pp. 269–284). Mahwah, NJ: LEA.

Nation, K., & Snowling, M.J. (2000). Factors influencing syntactic awareness skills in normal readers and poor comprehenders. *Applied Psycholinguistics,* 21, 229–241

National Institute of Child Health and Human Development (NICHD) (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* (NIH Publication No. 00-4769). Washington, DC: U.S. Government Printing Office.

National Research Council. (2001). *Classroom assessment and national science education standards*. Washington D.C.: National Academy Press.

Norris, S. P., & Phillips, L. M. (2003). How literacy in its fundamental sense is central to scientific literacy. *Science Education, 87*(2), 224-240.

Ozuru, Y., Dempsey, K., Sayroo J., & McNamara, D.S. (2005). Effect of text cohesion on comprehension of biology texts. Psychology Department, University of Memphis.

Palincsar, A. S. & Magnusson, S. J. (2001). The interplay of first-hand and text-based investigations to model and support the development of scientific knowledge and reasoning. In S. Carver & D. Klahr (Eds.), *Cognition and instruction: 25 years of progress* (pp.151-194). Mahwah, NJ: Lawrence Erlbaum.

Pearson, P.D. (2009). The roots of reading comprehension. In S.E. Israel & G.G. Duffy (Eds.), Handbook of research on reading comprehension (pp. 3–31). New York: Routledge.

Pearson, P.D., & Camperell, K. (1981). Comprehension of text structures. In J.T. Guthrie (Ed.), *Comprehension and teaching: Research reviews* (pp. 448-468). Newark, DE: International Reading Association.

Proctor, C.P., August, D., Carlo, M., & Snow, C. (2005). Native Spanish-speaking children reading in English: Toward a model of comprehension. *Journal of Educational Psychology,* 97(2), 246-256.

Qian, D.D. (2002) investigating the relationship between vocabulary knowledge and academic reading performance: an assessment perspective. *Language Learning,* 52, 513-536.

RAND Reading Study Group. (2002). *Reading for understanding: Towards an R&D program in reading comprehension.*

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models*. Thousand Oaks, CA: Sage Publications.

Rawson, K.A. (2004). Exploring automaticity in text processing: syntactic ambiguity as a test case. *Cognitive Psychology*, 49(4), 333-69.

Romance, N.R., & Vitale, M.R. (1992). A curriculum strategy that expands time for in-depth elementary science instruction by using science-based reading strategies: Effects of a year-long study in grade four. *Journal of Research in Science Teaching*, 29, 545-554.

Romance, N.R., & Vitale, M.R. (2006). Science IDEAS: Making the case for integrating reading and writing in elementary science as a key element in school reform. In R. Douglas,M. P. Klentschy, K. Worth and W. Binder (Eds.), *Linking science and literacy in the K–8 classroom* (pp. 391-405). Arlington, VA: National Science Teachers Association (NSTA) Press.

Rutherford, J.F. (1991). Vital connections: Children, books, and science. In S.Jagusch & W. Saul (Ed.), *Vital connections* (pp. 21-30). Portsmouth, NH: Heinemann.

Schleppegrell, M.J. (2004). *The language of schooling: A functional linguistics perspective*. Mahwah, NJ: LEA.

Shaw, J. M. (1997). Threats to the validity of science performance assessments for English language learners. *Journal of Research in Science Teaching,* 34(7), 721-743.

Shymansky, J.A., Yore, L.D., & Good, R. (1991). Elementary school teacher's beliefs about and perceptions of elementary school science, science reading, science textbooks, and supportive instructional factors. *Journal of Research in Science Teaching,* 28(5), 437-454.

Smagorinsky, P. (2001). If meaning is constructed, what is it made from? Toward a cultural theory of reading. *Review of Educational Research*, 71(1), 133-169.

Snow, C.E. (2010). Academic language and the challenge of reading for learning about science. *Science*, 328, 450-452.

Snow, C. E., & Sweet, A.P. (2003) Reading for Comprehension. In A. P. Sweet & C. E. Snow (Eds.), *Rethinking reading comprehension* (1-11). New York: Guilford Press.

Snowling, M., & Nation, K. (1997). Language, phonology and learning to read. In C. Hulme & M. Snowling (Eds.). *Dyslexia: Biology, cognition and intervention*. San Diego, CA: Singular Publishing Group.

Stahl, S.A. (1999). *Vocabulary development*. Cambridge, MA: Brookline.

Stanovich, K. (2000). *Progress in understanding reading: Scientific foundations and new frontiers*. New York: Guilford.

U.S. Census Bureau, (2000). http://www.census.gov/main/www/citation.html.

Vellutino, F.R. (2003). Individual differences as sources of variability in reading comprehension in elementary school children. In A.P. Sweet & C.E. Snow(Eds.), *Rethinking reading comprehension* (pp. 51-81). NY: Guilford Press.

Willows, D. M., & Ryan E. B. (1986). The development of grammatical sensitivity and its relationship to early reading achievement. *Reading Research Quarterly,* 21, 253–266.

Wilson, D, & Sperber, D. (1987). An outline of relevance theory. *Notes on Linguistics,* 39, 5-24.

Zeno, S.M., Ivens, S.H., Millard, R.T., & Duvvuri, R. (1995). *The Educator's Word Frequency Guide*. Brewster: Touchstone Applied Science Associates, Inc.