REMIE
Multidisciplinary Journal of Educational Research

Hipatia Press
www.hipatiapress.com

Instructions for authors, subscriptions and further details:

http://remie.hipatiapress.com

The Effects of a Web-Based Vocabulary Development Tool on Student Reading Comprehension of Science Text

Karen Thompson[1], Diego X. Román, Greg Wientjes & Kenji Hakuta[2]

1) College of Education, Oregon State University, United States of America.

2) Stanford University School of Education, United States of America.

Date of publication: October 15th, 2012

To cite this article: Thompson, K., Román, D.X., Wientjes, G. & Hakuta, K. (2012). The Effects of a Web-Based Vocabulary Development Tool on Student Reading Comprehension of Science Text. *Multidisciplinary Journal of Educational Research*, *3*(2), 272-300. doi: 10.4471/remie.2012.15

To link this article: http://dx.doi.org/10.4471/remie.2012.15

REMIE - Multidisciplinary Journal of Educational Research Vol. 2 No. 3 October 2012 pp. 272-300

The Effects of a Web-Based Vocabulary Development Tool on Student Reading Comprehension of Science Texts

Karen Thompson
Oregon State University

Diego X. Román
Stanford University

Greg Wientjes
Stanford University

Kenji Hakuta
Stanford University

Abstract

The complexities of reading comprehension have received increasing recognition in recent years. In this realm, the power of vocabulary in predicting cognitive challenges in phonological, orthographic, and semantic processes is well documented. In this study, we present a web-based vocabulary development tool that has a series of interactive displays, including a list of the 50 most frequent words in a particular text, Google image and video results for any combination of those words, definitions, and synonyms for particular words from the text, and a list of sentences from the text in which particular words appear. Additionally, we report the results of an experiment that was performed working collaboratively with middle school science teachers from a large urban district in the United States. While this experiment did not show a significant positive effect of this tool on reading comprehension in science, we did find that girls seem to score worse on a reading comprehension assessment after using our web-based tool. This result could reflect prior research that suggests that some girls tend to have a negative attitude towards technology due to gender stereotypes that give girls the impression that they are not as good as boys in working with computers.

Keywords: Vocabulary development, reading comprehension, technology, science texts

2012 Hipatia Press
ISSN 2014-2862
DOI: 10.4471/remie.2012.15

REMIE - Multidisciplinary Journal of Educational Research Vol. 2 No. 3 October 2012 pp. 272-300

Efectos de una herramienta-web de desarrollo del vocabulario para la mejora de la comprensión lectora de los textos de ciencias

Karen Thompson
Oregon State University

Diego X. Román
Stanford University

Greg Wientjes
Stanford University

Kenji Hakuta
Stanford University

Resumen

Las complejidades de la comprehensión lectora está recibiendo un mayor reconocimiento en los últimos años. En este aspecto, el poder del vocabulario en predecir retos en procesos cognitivos fonológicos, ortográficos y semánticos está bien documentado. En este estudio, presentamos una herramienta-web de desarrollo del vocabulario que tiene una serie de pantallas interactivas, incluyendo una lista de las cincuenta palabras más frecuentes en un texto específico, imágenes de Google y resultados de video por cada combinación de estas palabras, definiciones y sinónimos, así como frases que aparezcan en el texto. Adicionalmente, destacamos los resultados de un experimento que se llevó a cabo trabajando colaborativamente con profesores de ciencias de escuelas de secundaria de grandes distritos urbanos en los Estados Unidos. Mientras que el experimento no identificó efectos positivos de forma significativa debido al uso de esta herramienta, hallamos que las chicas parecen puntuar peor en las pruebas de comprehension lectora después de haber usado nuestra herramienta-web. Este resultado complementa lo que algunas investigaciones anteriores han destacado acerca de algunas chicas que suelen tener una actitud negativa hacia la tecnología debido a los estereotipos de género que existen y que señalan que ellas no son tan buenas como los chicos en las tareas que se lleven a cabo a través de computadoras.

Palabras claves: Desarrollo del vocabulario, comprensión lectora, tecnología, textos de ciencias

2012 Hipatia Press
ISSN 2014-2862
DOI: 10.4471/remie.2012.15

Researchers have long explored ways to support students' reading comprehension by fostering vocabulary development. We have developed a web-based vocabulary development tool that has a series of interactive displays, including a list of the 50 most frequent words in a particular text, Google image results for any combination of those words, definitions, and synonyms for particular words from the text, and a list of sentences from the text in which particular words appear. Working collaboratively with middle school science teachers from a large urban district, we designed a series of randomized experiments to test the effects of this interactive web-based vocabulary development tool on students' reading comprehension of content-area texts. Here, we report the results of the first in this series of experiments.

Theoretical background

The challenges and complexities of reading comprehension have received increasing recognition in recent years. The most recent data from the National Assessment of Educational Progress indicates that only one-third of 4th and 8th graders in the U.S. qualify as proficient readers. Furthermore, there are sizeable achievement gaps in reading ability by race and socioeconomic status. For example, approximately 15% of black and Latino 4th and 8th graders read at the proficient level compared to 40% of white students; an almost identical gap exists when comparing the reading scores of economically disadvantaged students with the scores of their peers (National Center for Education Statistics, 2010). Despite repeated calls to focus energy on improving the reading comprehension abilities of the nation's students, reading scores have changed relatively little in the past 20 years.

The RAND Reading Study Group (2002) provides a useful definition of reading comprehension as "the process of simultaneously extracting and constructing meaning through interaction and involvement with written language" (p. 11), in which three key components interface: the reader, the text, and the activity. To each reading task, readers bring their prior conceptual knowledge, their vocabulary knowledge, their prior experiences as readers, and a variety of cognitive abilities including decoding skills, memory, and attention. Texts, meanwhile, have their own set of characteristics. As the RAND Reading Study Group writes, "In addition to content, the vocabulary load of the text and its linguistic structure, discourse style, and genre also interact with the reader's knowledge. When too many of these factors are not matched to a reader's knowledge and experience, the text may be too difficult for optimal comprehension to occur" (2002, p. 14). Finally, features of the specific reading task - the activity – impact comprehension, as well, including the purposes the reader has set for herself. These three components – the reader, the text, and the activity – interface in a broader sociocultural context that also impacts comprehension. Economically disadvantaged students and students of color may be less likely to see their experiences reflected in the content of reading materials, for example, and literacy practices in these communities may be less valued in school settings.

Vocabulary is a central arena in which the discrete skills of reading (decoding, sight-word recognition, reading fluency and accuracy) come together with the top-down cognitive processes involved in comprehension. As researchers have noted, a deficit in any of these areas may prevent readers from comprehending grade-level text, but deficits in vocabulary knowledge and the semantic knowledge that it represents may be the most widely shared problem among struggling adolescent readers (Kamil, 2003; Snow, Porche, Tabors, & Harris, 2007).

The power of vocabulary in predicting cognitive processing in phonological, orthographic, and semantic processing as well as reading rates and other tasks are well-documented (Nation & Snowling, 2004; Yang & Perfetti, 2006). Catherine Snow and her colleagues have also demonstrated increasing correlations between vocabulary scores and reading comprehension scores as student move from primary to secondary grades (Snow et al., 2007).

In its comprehensive review of the reading research base, the National Reading Panel (National Institute of Child Health and Human Development, 2000) also emphasized that vocabulary plays a central role in reading comprehension, positing vocabulary instruction as one of the two key components of comprehension instruction, with the other being comprehension strategy instruction. From its review of the literature, the Panel drew a variety of conclusions about the features of effective vocabulary instruction. Among these conclusions were: (1) repetition and rich support are necessary for vocabulary learning; (2) effective vocabulary instruction requires active engagement on the part of students; (3) vocabulary should be taught both directly and indirectly; and (4) technology can support vocabulary learning. However, the Panel noted that much more research was needed to understand how vocabulary growth could best be fostered in instructional contexts.

Along with this general focus on the need for improved vocabulary instruction, researchers have paid increasing attention to the unique needs of adolescent readers. A recent report about adolescent literacy pointed to numerous features that make the texts readers encounter in the middle and high school grades much more challenging than those encountered by younger students. Texts for adolescents have greater word complexity, sentence complexity, and structural complexity, they present more conceptually challenging information, and they are of greater length (Carnegie Corporation of New York, 2010). In addition, unlike elementary teachers who must have training in reading instruction, many middle and high school content-area teachers have had no such training. Thus, although the content-area teachers from whom adolescents receive much of their instruction often rely heavily on texts to present essential information, many content-area teachers presuppose adequate literacy skills among their students and … are typically not well prepared to teach students with below-average literacy skills" (RAND Reading Study Group, 2002, pp. 5-6).

Learning academic words (Coxhead, 2000) that cut across subject matter (e.g., affirm, interpret, deny, evidence, conclusion, theory, factor, process) presents particular challenges for adolescent students in content-area classrooms. Within science classrooms such as those where our collaborating teachers work, Snow (2008) has found that non-specialized academic words are crucial to understanding written and oral science texts. In discussing the importance of teaching academic words, Snow (2008) advocates for explicit vocabulary instruction in science classrooms that not only focuses on subject-specific words, but also addresses the cross-discipline academic vocabulary. As these words do not commonly appear in conversations, it is very likely that non-mainstream readers lack sufficient exposure to the more complex meanings of these words. Snow asserts that many content-area teachers assume that students, even English learners, already know general academic vocabulary and only focus on teaching subject-specific vocabulary.

Reading in Science

The new Common Core State Standards For English Language Arts in the United States indicates that one of its main goals is for all students to comprehend independently and proficiently the complex texts found in college and career settings (CCSSI, 2010). The view presented in this document indicates that although the reading demands in college and careers have consistently increased in sophistication, "little attention has been paid to students' ability to read texts independently" (CCSSI, 2010, p. 2). This situation has left an enormous gap between the reading that occurs in colleges and professions compared to the reading that takes place in K-12 schools. This gap is even more noticeable when looking at the reading proficiency scores of English learners and underrepresented racial groups (Schleppegrell, 2004; Fang, 2005).

Among the literacy skills students require to access science concepts, Yore and Shymansky (1991) state that the ability to "read about science is a critical skill to have in order to develop scientific literacy" (p. 29). Likewise, other literacy experts posit that the ability to comprehend expository texts is essential for students to succeed in science classes and science-related careers (Moore, Readence, & Rickleman, 1983; Yore & Shymansky, 1991; Halliday & Martin, 1993; CCSSI, 2010). Furthermore, according to Yore and Shimansky, science teachers could significantly improve their effectiveness with "contemporary knowledge about the reading process and its application in classrooms" (p. 29).

By including literacy standards to accompany science content standards for each K-12 grade-level, the Common Core State Standards have finally acknowledged what research has unequivocally shown: language should play a more prominent role in science instruction (Lemke, 1990; Halliday & Martin, 1993; Wellington & Osborne, 2001; Norris & Phillips, 2003; Pearson, Moje, & Greenleaf, 2010). Infusing literacy in the teaching of science, however, can be "viewed as a radical proposal" (Alberts, 2010, p. 405) since literacy instruction has traditionally been confined to language arts classes and focused on fictional rather than expository texts.

Pearson et al. (2010) echo Alberts' concern, pointing out that science teachers perceive implementing literacy instruction as contradictory to hands-on inquiry strategies, conceptualizing the latter as more science-specific. Furthermore, Pearson et al. posit that this discrepancy in literacy and science instruction perspectives has generated two distinct interpretations of scientific literacy in the education community. The first definition of scientific literacy –perhaps the most accepted one by science teachers– focuses mostly on the knowledge of the science concepts, skills, and processes people should possess to be considered scientifically literate. The second definition of scientific literacy, although also addressing scientific knowledge and skills, also considers how the development of basic literacy skills affects the growth of scientific literacy in individuals.

Norris and Phillips (2003) call the essential notion of scientific literacy (i.e., related to the ability to read and write science texts) the fundamental sense of literacy and the knowledge-specific domain the derived sense of literacy. In their analysis, these authors indicate that the fundamental sense of literacy has been neglected in favor of the derived sense of literacy (i.e., scientific literacy). Similarly, Wellington and Osborne (2001) state that the focus of secondary science education has mostly been on teaching science as an empirical subject, even though for many students "the greatest obstacle in learning science –and also the most important achievement – is to learn its language" (p. 3).

The implications for not addressing language issues in science education are daunting. As different authors have mentioned, almost all of what we call "knowledge" is based on language (Wellington & Osborne, 2001; Hines, Wible, & McCartney, 2010) and, in the end, doing science depends on being able to talk science to ourselves and to others (Lemke, 1990). Science teachers, however, lack the knowledge about "the vital role [that] literacy plays in enhancing rather than replacing science learning" (Pearson et al., 2010, p. 462) and, thus, are not able to mentor students in these practices.

The Common Core State Standards For English Language Arts and Literacy in History/Social Studies, Science and Technical Subjects (CCSI, 2010) has renewed the interest in the United States years in the field of content-area reading instruction. This is not a new area of research, as it has been present in the American system for over a hundred years (Hall, 2005). For instance, in a paper that presents a comprehensive review of this topic, Moore et al. (1983) mention that recommendations for content-area reading started to proliferate in the United States around the first half of the twentieth century. These authors indicate that this field was born out of the recognition that readers require specific strategies when working with subject-specific materials. Thus, the main objective of content-area reading is "to develop students' reading to learn strategies" (Moore, et al., 1983, p. 420).

According to Chall (1983), the difference in reading demands between young children and adolescents is similar to the distinction between learning to read (from pre-K through third grade) and using reading to learn (from fourth grade until the end of secondary schooling). This author indicates that children are only able to use reading to learn after mastering six distinct stages: (0) getting ready to learn (becoming acquainted with letters, words, and how books are used), (1) learning beginning reading skills, (2) practicing beginning reading skills, (3) reading to learning the new, (4) reading multiple points of view, and (5) constructing and reconstructing meaning. Chall points out that the first three stages occur between pre-Kindergarten through third grade, whereas the rest start emerging in grade fourth and continue throughout schooling, and even during college and beyond (stage 5, especially).

Although being able to read was initially considered an accumulation of specific skills that allowed for the decoding of words (Moore, et al., 1983; Jacobs, 2008), the definition of reading gradually evolved over time to include the understanding of the semantic, pragmatic, and sociocultural domains in which these words occur. According to Moore et al. (1983), prior to the 1900s, U.S. reading instruction consisted mainly of elocution and memorization.

Jacobs (2002) states that the gradual shift from reading as an accumulation of specific skills to reading as a meaning-based process came together with "a shift in responsibility for secondary reading instruction from the reading specialist to the content-area teacher" (p. 19). Jacobs points out that although there was not much debate about the benefits of teaching reading at different grade levels, content area teachers were uncomfortable seeing reading as part of their instructional responsibilities. Moreover, even the content area teachers that felt sympathetic to teaching reading in their classrooms complained about the lack of training they had in reading instruction and the time away from teaching their subject these practices entailed (Jacobs, 1999).

Although researchers have agreed that students can profit from having reading instruction incorporated into their content-area classes (Anders & Guizzetti, 1996), researchers and teacher educators have struggled to help content area teachers see the benefits of implementing reading practices in their practices (O'Brien, Stewart, & Moje, 1995). Studies that analyze the implementation of literacy practices in content-area classrooms encountered teachers resisting the use of these strategies (Hall, 2005). Moreover, these papers report that content-area teachers had a number of misconceptions about the usefulness of incorporating literacy in the subject-areas (Yore, 1991).

Hall (2005) points out that the lack of implementation of reading strategies in content-area classes might be the product of deficient explanations by researchers and teacher educators about the role that content-area teachers play as teachers of reading. Moreover, Wineburg (2001) has argued that teacher preparation programs often portray reading as a general skill without providing specific ways in which reading strategies could be incorporated into the content areas. Furthermore, this author mentions that this generic description of reading results in teachers not understanding the nuances of reading in their subjects, as well as the different demands that texts they use place upon their students.

The Web-Based Vocabulary Tool

Given the need to foster vocabulary development, particularly among adolescents, and the need to provide content-area teachers with strategies to foster literacy development, we developed a web-based vocabulary tool called WordSift, which aligns with the features of effective vocabulary instruction identified by the National Reading Panel (2000). WordSift is a free tool, accessible to anyone at http://www.wordsift.com. Upon arriving at the site, users see a blank box into which they can paste any text they choose from a website, pdf, or other electronic document. After clicking a button that says "Sift," users are taken to a new page that presents a variety of interactive displays, including: (1) a list of the 50 most frequent words in the text sifted by the user; (2) Google image results for any combination of those words; (3) definitions and synonyms for particular words from the text; and (4) a list of sentences from the text in which particular words appear. Figure 1 shows a screenshot of WordSift's displays, and we will describe each in greater detail.

Figure 1. A screenshot of WordSift's initial display for a text about solar energy

The first thing users see after sifting a text is a "tag cloud" of the list of the 50 most common words in the text they sifted (http://en.wikipedia.org/wiki/Tag_cloud). These words are sized according to their frequency in the text, with the more frequent words appearing the largest. For example, in the text sifted to create Figure 1, the most frequent word is "solar." The words in the tag cloud are initially arranged in alphabetical order and appear in green. However, users have several options to alter the appearance of the tag cloud. For example, users may sort the words from common to rare. If this option is selected, the words no longer appear in alphabetical order but instead appear in the order of their frequency in the English language. (Although the order in which the words are displayed can be altered, the words retain their sizing to indicate their relative frequency in the text sifted by the user.) In addition, users have the option to mark various words in the cloud that appear on specific word lists. For example, users may mark words in the cloud that appear on the Academic Word List (Coxhead, 2000), which consists of words that surface frequently in academic texts across a variety of disciplines. In addition, users may mark subject-specific words, highlighting words that appear frequently in just science or social studies, for example. When users mark a specific set of words, those words turn from green to orange in the tag cloud.

Below the tag cloud on the left, users see a display of images. Initially, these images show results from a Google image search for the two most frequent words in the text sifted by the user. In Figure 1, for example, the two most frequent words in the text sifted are solar and energy, so the images that initially appear are Google image results for the combination of these two words. Users can click on any of these images to see them at full size.

Below the tag cloud on the right is a display from the Visual Thesaurus®. The most frequent word in the text the user sifted is entered into the Visual Thesaurus® and the result is displayed as a word web. The Visual Thesaurus® is a product based on WordNet, a digital dictionary and thesaurus created by Princeton psychologist George Miller. The Visual Thesaurus® display is interactive: the definition of each word on the display pops up when the cursor scrolls over it, and a click on any word on the web re-configures the display to bring that word to the center.

Finally, below the Visual Thesaurus® is a list of sentences from the text. Initially, WordSift displays the sentences from the text that contain the most frequent word. Thus, in Figure 1, WordSift displays sentences from the text that contain the word solar. Clicking on any of the sentences will take the user to a display of the entire text, with that particular sentence highlighted so that the user can see the sentence in the context of the rest of the text.

Importantly, WordSift's four displays are not static. By interacting with each of them, the user can explore words and concepts of particular interest. For example, clicking on a new word in the tag cloud will alter the images, Visual Thesaurus® display, and the sentences so that the displays focus on the new word selected by the user.

WordSift is intended for use by teachers, students, and the general public for a variety of uses. The website guide encourages using it as a "visual playground" to explore text. Teachers might use WordSift on their own to preview a text they are going to read with students and identify vocabulary words for instruction. In addition, teachers might use WordSift in whole class settings, projecting WordSift's display for the class and exploring words and their meanings together. Students themselves might use WordSift in a computer lab or at home to preview or review a text they are reading and explore word meanings. Through its multi-faceted displays, WordSift is designed to provide students with repetition and rich support for word learning, to promote active engagement among students, and to facilitate both direct and indirect vocabulary development – all features of effective vocabulary instruction identified by the National Reading Panel (2000).

As Figure 2 illustrates, WordSift is designed to affect reading comprehension by influencing the three key components of the reading task: the reader, the text, and the activity, all within the broader sociocultural context. Previewing a text with WordSift influences the reader by presenting her with an opportunity to learn words that are unfamiliar to her. Furthermore, WordSift can influence the reader by activating prior knowledge and providing a schema for the text she is going to encounter via the tag cloud. Exposure to WordSift can also reduce the complexity of the text for the reader by allowing the reader to interact with the text in a variety of ways prior to a full reading. Finally, WordSift can influence the activity of reading itself by shaping students' purposes for reading. By testing WordSift's effect, we seek to add to the research knowledge base about strategies for fostering vocabulary development and thus improving reading comprehension, particularly for adolescents.

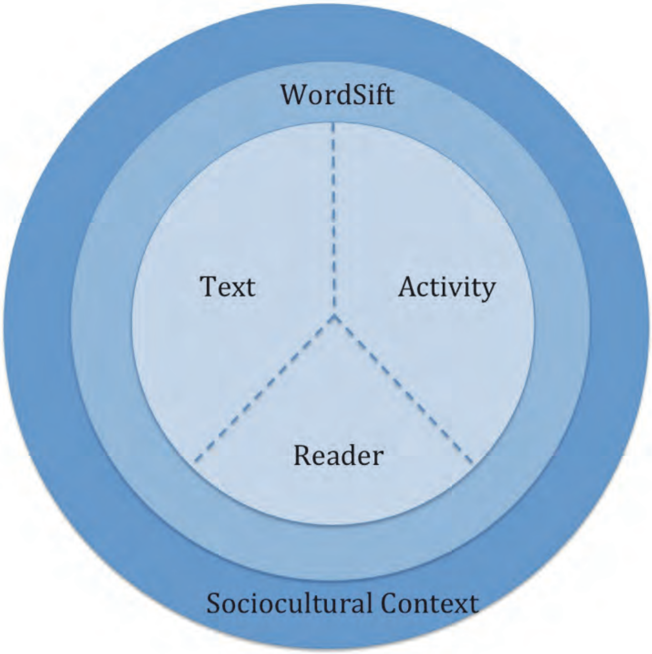


Figure 2. Schematic illustration of WordSift's relationship to reading comprehension
Note: Adapted from RAND Reading Study Group, 2000 (Figure S.1, p. xiv)

Methods

To test the effect of this web-based vocabulary development tool, in partnership with nine middle school science teachers with whom we collaborated for two years, we designed a within-subjects experiment, spread out over two class sessions of approximately 50 minutes. Each student experienced both the treatment and control conditions. In the treatment condition, teachers led students in interacting with WordSift.

Teachers pasted a passage from a grade-level science textbook into WordSift, and with the teachers' guidance, students completed a preview worksheet based on the web-based vocabulary development tool's displays. Then students read the passage individually and answered a series of comprehension questions about what they read. In the control condition, teachers also led students in completing a preview worksheet for a science textbook passage, this time without the benefit of WordSift's displays. Then, as in the treatment condition, students read the passage and answered comprehension questions about what they read. The texts and conditions were counterbalanced to eliminate text and order effects. Table 1 illustrates the four conditions to which the classrooms were assigned (because we focus here on 6th grade assessment results, as explained below). Assignment to conditions occurred at the class level. All aspects of the experimental design process, from the selection of the textbook excerpts to the development of the comprehension assessments and scoring rubrics, were conducted collaboratively with the nine middle school teachers.

Table 1
The four experimental conditions to which 6th grade classrooms were randomly assigned

Condition 1 Day 1: Solar text with WordSift Day 2: Fossil fuels text without WordSift	*Condition 2* Day 1: Fossil fuels text with WordSift Day 2: Solar text without WordSift
Condition 3 Day 1: Solar text without WordSift Day 2: Fossil fuels text with WordSift	*Condition 4* Day 1: Fossil fuels text without WordSift Day 2: Solar text with WordSift

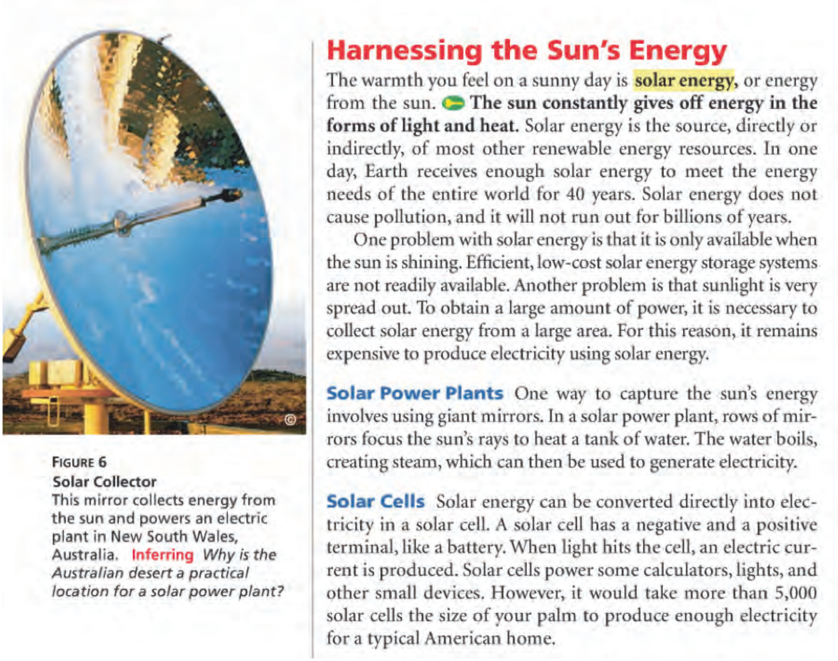
Data Sources

We collected assessment data from 479 students, for a total of 958 observations, nested within 19 classrooms of eight different teachers (see Table 2).

Table 2
Number of students participating in the experiment, by grade level, teacher, period, and condition

Grade	Teacher	Period	Condition	N
6	Teacher 1	2	2	17
6	Teacher 1	7	1	22
6	Teacher 2	1	4	28
6	Teacher 2	2	2	30
6	Teacher 3	2	4	31
6	Teacher 3	3	1	30
6	Teacher 4	5	3	12
6	Teacher 4	2	2	27
7	Teacher 5	2	2	28
7	Teacher 5	3	1	30
8	Teacher 5	5	4	30
8	Teacher 5	6	3	22
8	Teacher 5	7	3	19
8	Teacher 6	3	4	14
8	Teacher 6	7	2	33
8	Teacher 7	6	3	28
8	Teacher 8	1	1	22
8	Teacher 8	4	4	33
8	Teacher 8	6	3	23

Here we report results for the sixth grade assessment data, which consists of 394 observations nested within 197 students who are enrolled in eight different classes taught by four teachers. Two assessments were designed, one for each of the two reading passages, one on fossil fuels and one on solar energy. Figure 3 presents an excerpt from the passage on solar energy. Each assessment consisted of two open-response questions and two multiple-choice questions. Questions for the assessments were drawn from the science textbook itself, as well as the suggestions of teachers and district content specialists.



Harnessing the Sun's Energy

The warmth you feel on a sunny day is **solar energy**, or energy from the sun. **The sun constantly gives off energy in the forms of light and heat.** Solar energy is the source, directly or indirectly, of most other renewable energy resources. In one day, Earth receives enough solar energy to meet the energy needs of the entire world for 40 years. Solar energy does not cause pollution, and it will not run out for billions of years.

One problem with solar energy is that it is only available when the sun is shining. Efficient, low-cost solar energy storage systems are not readily available. Another problem is that sunlight is very spread out. To obtain a large amount of power, it is necessary to collect solar energy from a large area. For this reason, it remains expensive to produce electricity using solar energy.

Solar Power Plants One way to capture the sun's energy involves using giant mirrors. In a solar power plant, rows of mirrors focus the sun's rays to heat a tank of water. The water boils, creating steam, which can then be used to generate electricity.

Solar Cells Solar energy can be converted directly into electricity in a solar cell. A solar cell has a negative and a positive terminal, like a battery. When light hits the cell, an electric current is produced. Solar cells power some calculators, lights, and other small devices. However, it would take more than 5,000 solar cells the size of your palm to produce enough electricity for a typical American home.

FIGURE 6
Solar Collector
This mirror collects energy from the sun and powers an electric plant in New South Wales, Australia. **Inferring** *Why is the Australian desert a practical location for a solar power plant?*

Figure 3. Excerpt from 6th grade reading passage on solar energy
Note: From Earth Science, Prentice Hall, © 2007, p. 486

Figure 4 presents the four questions used in the solar energy assessment. Each open-response question was scored on a four-point rubric, which was developed and refined over the course of repeated meetings with teachers. Multiple-choice questions were scored as correct or incorrect. These assessments were scored by teachers and researchers, after a scoring calibration process, including an inter-rater reliability check. Scores for assessment items one and two (the open-ended response items) were each converted to separate z-scores. Scores for assessment items three and four (the multiple-choice items) were summed, and this sum was also converted to as z-score. We then checked the reliability of the three scales by computing Cronbach's alpha separately for each of the two assessments. For the fossil fuel assessment, Cronbach's alpha was 0.567, and for the solar energy assessment Cronbach's alpha was is 0.672. Finally, the z-scores were summed to create our final outcome measure, sumz, which represents an overall score for the comprehension assessment. This outcome measure ranged from -6.423 to 3.996 with a mean of -.002 and a standard deviation of 2.265.

Part 1: Short Answer Questions

1. Define solar cells and give an example of how they are used.
2. Describe three features of a solar home.

Part 2: Multiple Choice

3. Which of the following is a problem with solar energy?
    A. There is not enough solar energy for the world's energy needs.
    B. Solar energy causes pollution.
    C. It is expensive to produce electricity using solar energy.
    D. The supply of solar energy will eventually run out.
4. The interior of your car heats up on a sunny day because of:
    A. Passive solar heating.
    B. Solar cells.
    C. Active solar heating.
    D. Indirect solar heating.

Figure 4. Questions from solar energy assessment.

Results

We used hierarchical linear modeling (HLM) to analyze the assessment results. HLM allows us to account for the nested structure of our data because it makes no assumption that observations are independent. Since our data consists of observations (i.e. reading comprehension assessment scores) nested within students nested within class periods, we constructed a three-level HLM model.

First, to determine the portion of variability associated with each level – observations, students, and class periods – we constructed an unconditional HLM model:

$$Level-1: Y_{ijk} = \pi_{0jk} + e_{ijk}$$
$$Level-2: \pi_{0jk} = \beta_{00k} + r_{0jk}$$
$$Level-3: \beta_{00k} = \gamma_{000} + u_{00k}$$

Our estimated fixed effect for this model was -.337, which represents the estimated grand mean for students' reading comprehension assessment scores. (See Table 3 for all model results.) Our fixed effect is not significant ($p<.05$), but this simply suggests that the estimated grand mean is not significantly different from 0, which we might expect since the assessment score is comprised of z-scores with a mean of 0. Of greater interest are the variance components, which are significant ($p<.001$) at both Level 2 (students within classrooms) and Level 3 (classrooms). Furthermore, the inter-class correlation coefficients indicate that 17% of the variance in assessment scores is between teachers and 38% of the variance in scores is between classrooms. This provides further evidence that a multilevel model is suitable for these data.

Next, we constructed an ANCOVA model by inserting three dummy variables at the observation level related to the experimental design. The first of these variables indicates whether a particular observation occurred with WordSift (if so, WordSift=1) and is the key variable for estimating the treatment effect. The other variables at the observation-level seek to control for features of our experimental design. One indicates whether the observation occurred on day 1 of the experiment

Table 3
A series of hierarchical linear models predicting the reading comprehension assessment scores of sixth graders with observation-, student-, and school-level variables

		Model 1: Unconditional		Model 2: ANCOVA (Baseline)		Model 3: Random		Model 4: Level 2 variables	
Fixed effects		Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Intercept		-.337	.521	-.337	.521	-.336	.520	-.8386	.525
	Female							-.6145**	.197
	GATE							1.1505***	.249
Wordshift				-.429*	.168	-.396	.223	.065	.375
	Female							-.762*	.335
	GATE							-.169	.371
Day 1				.150	.177	.112	.231	-.646	.400
	Female							.292	.352
	GATE							.838	.397
Solar				.117	.171	.081	.224	-.275	.385
	Female							.003	.339
	GATE							.476	.386
Random effects: Variance Components									
Level-1		2.46010		2.38112		2.29629		2.27475	
Level-2	Intercept	.92131***		0.96081***		1.00329***		.71927***	

		Model 1: Unconditional	Model 2: ANCOVA (Baseline)	Model 3: Random	Model 4: Level 2 variables
Level-3	Intercept	2.07133***	2.07133***	2.07133***	2.07084***
	WordSift			.16027*	
Proportion of L-2 variance		.169	.179	.181	.151
Proportion of L-3 variance		.380	.382	.403	.372
Proportion of L-2 & L-3 variance		.549	.559	.584	.523
Proportion explained variance at Level-1			.032	.067	
Proportion explained variance at Level-2 (compared to baseline)					.143
Deviance		1608.271029	1601.842946	1599.590994	1563.762988
Parameters		4	7	9	15
Chi-square for change in deviance			6.42809+ (compared to unconditional)	2.25195 (compared to Model 2: ANCOVA -Baseline-)	38.07995*** (compared to Model 2: ANCOVA -Baseline-)
df			3	2	8

+p<.1 *p<.05. **p<.01. ***p<.001.
For fixed effects, italics indicates that the variable was centered around its group mean.

(if so, day1=1), and the other indicates whether the observation occurred after reading the text about solar energy (if so, solar=1). Model 2 with observation-level fixed effects is:

$$\begin{aligned}
&Level-1: Y_{ijk} = \pi_{0jk} + \pi_{1jk} wordsift_{ijk} + \pi_{2jk} day1_{ijk} + \pi_{3jk} solar_{ijk} + e_{ijk} \\
&Level-2: \pi_{0jk} = \beta_{00k} + r_{0jk} \\
&\quad \pi_{1jk} = \beta_{10k} \\
&Level-3: \beta_{00k} = \gamma_{000} + u_{00k} \\
&\quad \beta_{10k} = \gamma_{100}
\end{aligned}$$

The coefficient of interest is the coefficient on WordSift, which is -.429 and significant ($p<.01$), suggesting that, in this model, before accounting for student characteristics, students did somewhat worse on the comprehension assessment they took after using WordSift than on the comprehension assessment they took after previewing the text in a more conventional way. The other fixed effects accounting for the day of the assessment (day1) and the text used (solar) are not significant ($p>.05$), suggesting that these features of the experimental design did not significantly impact students' assessment results. In analyzing the variance components for this model, we see that the variance components at Level 2 (students within classrooms) and Level 3 (classrooms) are still significant ($p<.001$). However, when we compare the deviance statistics of our ANCOVA model and our unconditional model, we see that the reduction in the deviance in our unconditional model is only marginally significant ($p>.1$), suggesting that this model only marginally more explanatory power than our unconditional model. Nonetheless, since the fixed effect on WordSift is significant and is what we are interested in estimating, since both variance components at Level 2 and Level 3 are significant, and since we want to control for any variation related to our experimental design, we retain these observation-level variables in future models.

In our third model, we let the effect of WordSift vary randomly at Level 3 across classrooms:

$$
\begin{aligned}
&Level-1: Y_{ijk} = \pi_{0jk} + \pi_{1jk} wordsift_{ijk} + \pi_{2jk} day1_{ijk} + \pi_{3jk} solar_{ijk} + e_{ijk} \\
&Level-2: \pi_{0jk} = \beta_{00k} + r_{0jk} \\
&\quad \pi_{1jk} = \beta_{10k} \\
&Level-3: \beta_{00k} = \gamma_{000} + u_{00k} \\
&\quad \beta_{10k} = \gamma_{100} + u_{00k}
\end{aligned}
$$

(Since we have only two observations for each student, we cannot let the effect of WordSift vary randomly at Level 2 across students). In this model, the coefficient on WordSift is no longer significant ($p>.05$). The coefficients on the other variables related to experimental design remain non-significant, as in the previous model. Although the variance components at both Level 2 and Level 3 are significant ($p>.05$), because the reduction in the deviance statistic compared to the ANCOVA model is no greater than would be expected by chance given the number of parameters we are estimating, we do not retain the random effect for WordSift in future models and use Model 2 as our baseline model.

In our final model, we entered student-level variables, controlling for gender and participation in gifted programs[1]. Additionally, we checked whether these student-level variables mediated the treatment effect:

$$
\begin{aligned}
&Level-1: Y_{ijk} = \pi_{0jk} + \pi_{1jk} wordsift_{ijk} + \pi_{2jk} day1_{ijk} + \pi_{3jk} solar_{ijk} + e_{ijk} \\
&Level-2: \pi_{0jk} = \beta_{01k} + \beta_{01k} female_{0jk} + \beta_{02k} gate_{0jk} + r_{0jk} \\
&\quad \pi_{1jk} = \beta_{10k} + \beta_{11k} female_{1jk} + \beta_{12k} gate_{1jk} \\
&\quad \pi_{2jk} = \beta_{20k} + \beta_{21k} female_{2jk} + \beta_{22k} gate_{2jk} \\
&\quad \pi_{3jk} = \beta_{30k} + \beta_{31k} female_{3jk} + \beta_{12k} gate_{3jk} \\
&Level-3: \beta_{00k} = \gamma_{000} + u_{00k} \\
&\quad \beta_{10k} = \gamma_{100}
\end{aligned}
$$

In this model, the coefficients on both student-level variables were significant on the intercept. Controlling for other factors, girls scored .615 points lower, on average, than boys on the comprehension assessments ($p<.01$), while students in the gifted program scored 1.151 points higher, on average, than other students ($p<.001$). Interestingly, it appears that gender significantly mediated the treatment effect ($p<.05$), resulting in a significant negative treatment effect for girls and slightly positive (though not significant) effect for boys ($p>.05$) (See Figure 5).

After controlling for the other observation- and student-level factors in our model, boys scored an average of .065 points higher on reading comprehension assessments after using WordSift. Girls, however, scored an average of .698 points lower after using the web-based tool, approximately one-third of a standard deviation lower.

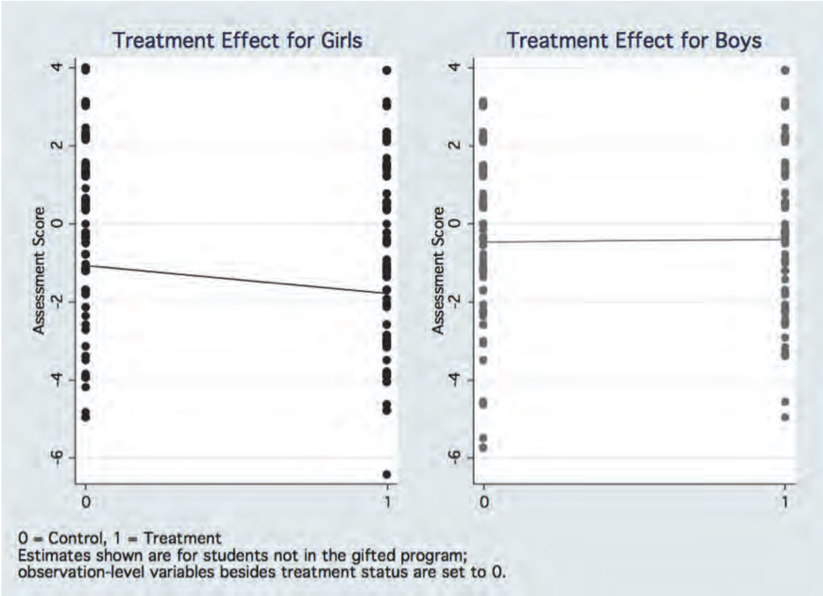


Figure 5. Effect of web-based vocabulary tool on reading comprehension assessment scores by gender.

The other interactions of student characteristics and observation-level variables were not significant ($p>.05$).[2] The variance components at both Level 2 and Level 3 were still significant ($p>.001$) and the reduction in the deviance statistic compared to our baseline model was greater than we would predict by chance based on the number of parameters we were estimating. Introducing student-level characteristics explained an additional 14% of the variance at Level 2 compared to our baseline model. Yet substantial variation at all levels remained unexplained.

Discussion

Although results from this first set of experiments did not show a significant positive effect for the web-based vocabulary tool on students' reading comprehension, the experiment described here represented only a first attempt to investigate WordSift's impact on learning. Later experiments will investigate whether WordSift shows an effect on students' reading comprehension when students use the tool themselves in a computer lab rather than in a whole-class setting. Furthermore, the limited number of English learners in the sample analyzed here prevented us from fully exploring how the web-based vocabulary tool affects the reading comprehension of this population. Later experiments will focus on classrooms with more English learners.

Our finding that girls seem to score worse on the reading comprehension assessment after using WordSift merits further discussion. Prior research suggests that girls tend to experience and interact with educational technology differently than boys and typically have less experience with and less positive attitudes towards computers (cf. Sanders, 2005). The negative attitudes that some girls have towards technology might in part be the result of gender stereotypes that give girls the impression that they are not as good as boys in working with computers (Adya & Kaiser, 2005). Parents and teachers can generate and perpetuate these stereotypes by giving boys priority on computer usage and, thus, reinforcing perceptions that boys are inherently better at working with technology than girls (Sanders & Stone, 1986; Volman & van Eck, 2001).

Additionally, there are studies that have found that females' confidence level in their computer ability is significantly lower than that of males, even when females were more successful than the males in the group studied (Gurer & Camp, 1998). The lack of confidence in their computer skills can result in girls deciding at young ages that they do not want to pursue careers in technology-related fields (Adya & Kaiser, 2005). The American Association of University Women (2000), for instance, reports that girls as young as 11 years old might have already decided that they do not like math, science, and technology. While WordSift itself cannot eliminate disparities in girls and boys confidence with and interest in technology, teachers can be mindful of these issues when using WordSift in their classrooms, ensuring that girls have plenty of time to explore the site and become comfortable using its multiple features.

While this experiment did not show a significant positive effect of WordSift on reading comprehension in science, students in both the treatment and control conditions received content-area reading instruction, having the opportunity to learn vocabulary and grapple with science texts. Research reports that students who receive content-area reading instruction are more likely to improve their comprehension of content-area texts and develop a wider range of reading skills that allow them to infer the meaning of content-area texts (Hall, 2005). In other words, some content-area literacy instruction is better than none. If WordSift provides teachers with a entry point for implementing literacy instruction in science classrooms, this may eventually have positive benefits for students' reading comprehension skills.

Finally, from a methodological perspective, the research reported here demonstrates how classroom teachers can be full collaborators in an iterative experimental design process, planning and carrying out scientifically rigorous experiments within their own classrooms to answer questions of real instructional interest.

Notes

[1] Our sample consists of 103 girls (52%) and 94 boys (48%). One hundred forty-eight students in our sample (75%) qualify for the district's gifted program, an unusually high proportion. Although we are particularly interested in the effect of the web-based vocabulary tool for English learners, this sample of sixth graders contained only seven ELs (3.6% of the sample), limiting our ability to detect treatment effects specific to this population. In future experiments, we will include classrooms with more ELs.

[2] The coefficient for the intercept and for the main effects of the observation-level variables appear to have shifted in this model. That is because unlike the observation-level variables, the student-level variables have not been group mean centered. Thus, the estimate for the intercept now reflects the estimated mean assessment score for a boy not in the gifted program.

References

Adya, M., & Kaiser, K. (2005). Early determinants of women in the IT workforce: A model of girls' career choices. *Information Technology & People, 18*(3), 230-259. doi: 10.1108/09593840510615860

Alberts, B. (2010). Prioritizing science education. *Science, 23*(328), 405. doi:10.1126/science.1190788

American Association of University Women. (2000). *Tech-Savvy: Educating girls in the new computer age*. Washington DC: American Association of University Women.

Anders, P., & Guizzetti, B. (1996). *Literacy instruction in the content areas*. NY: Harcourt.

Carnegie Council on Advancing Adolescent Literacy. (2010). *Time to act: An agenda for advancing adolescent literacy for college and career success*. New York: Carnegie Corporation of New York.

Chall, J. (1983). *Stages of reading development*. NY: McGraw-Hill.

Common Core Standards Initiative (2010). *Common Core State Standards for English Language Arts & Literacy in History/Social Studies, Science, and Technical Subjects*. Retrieved from http://www.corestandards.org/the-standards

Coxhead, A. (2000). A New Academic Word List. *TESOL Quarterly, 34* (2), 213-238. doi:10.2307/3587951

Fang, Z. (2005). Scientific literacy: a systemic functional linguistics perspective. *Science Education, 89*, 335-347. doi: 10.1002/sce.20050

Gurer, D., & Camp, T. (2002). An ACM-W literature review on women in computing. *Women and Computing, 34*(2), 121-127. doi: 10.1145/543812.543844

Hall, L. (2005). Teachers and content area reading: Attitudes, beliefs, and change. *Teaching and Teacher Education, 21*, 403-414. doi: 10.1016/j.tate.2005.01.009

Halliday, M., & Martin, J. (1993). *Writing science: literacy and discursive power.* PA: University of Pittsburgh Press.

Hines, P., Wible, B., & McCartney, M. (2010). Learning to read, reading to learn. *Science, 23*(328), 447. doi:10.1126/science.328.5977.447

Jacobs, V. (1999). What secondary teachers can do to teach reading: a three-step strategy for helping students delve deeper into texts. *Harvard Education Letter*, July/August, 4-5.

Jacobs, V. (2002). Reading, writing, and understanding. *Educational Leadership 60*(3), 58-61.

Kamil, M. L. (2003). *Adolescents and literacy: Reading for the 21st century*. Washington, DC: Alliance for Excellent Education.

Lemke, J. (1990). *Talking science: language learning, and values*. CT: Ablex.

Moore, D., Readence, J., Rickelman, R. (1983). A historical exploration of content area instruction. *Reading Research Quarterly, 18*(4), 419-438.

Nation, K., & Snowling, M. (2004). Beyond phonological skills: Broader language skills contribute to the development of reading. *Journal of Research in Reading, 27*(4), 342-356. doi: 10.1111/j.1467-9817.2004.00238.x

National Center for Education Statistics. (2010). *Projections of Education Statistics to 2010*. Retrieved from: http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2000071

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*. Washington DC: Author.

Norris, S. & Phillips, L. (2003). How literacy in its fundamental sense is central to scientific literacy. *Science Education, 87*(2), 224-240. doi:10.1002/sce.10066

O'Brien, D., Stewart, R., & Moje, E. (1995). Why content literacy is difficult to infuse into secondary curriculum: Strategies, goals, and classroom realities. *Reading Research Quarterly 30*(3), 442-463.

Pearson, D., Moje, E., & Greenleaf, C. (2010). Literacy and science: Each in the service of the other. *Science, 23*(328), 459-463. doi:10.1126/science.1182595

RAND Reading Study Group. (2002). *Reading for understanding: Toward a research and development program in reading comprehension*. Santa Monica, CA: Author.

Sanders, J. (2005). *Gender and technology: A research review.* Retrieved July 21, 2009 from http://www.josanders.com/pdf/gendertech0705.pdf

Sanders, J. & Stone, A. (1986). *The neuter computer: Computers for girls and boys*. New York: Neal-Schuman Publishers.

Schleppegrell, M. (2004). *The Language of schooling: a Functional Linguistics perspective*. NJ: Erlbaum.

Snow, C. (2008). What is the vocabulary of science? In A. Rosebery and B. Warren (Eds.), *Teaching science to English language learners* (pp. 71-83). Arlington, VA: NSTA Press.

Snow, C. E., Porche, M. V., Tabors, P., & Harris, S. (2007). *Is literacy enough?: Pathways to academic success for adolescents.* Baltimore, MD: Paul H. Brookes Publishing Co.

Volman, M., & van Eck, E. (2001). Gender equity and information technology in education: The second decade. *Review of Educational Research, 17*(4), 613-634. doi: 10.3102/00346543071004613

Wellington, J., & Osborne, J. (2001). *Language and literacy in science education*. Glasgow: Open University Press.

Wineburg, S. (2001). *Historical thinking and other unnatural acts*. PA: Temple University.

Yang, C. L., & Perfetti, C. (2006). *Reading skill and the acquisition of high quality representations of new words*. Paper presented at the Society of the Scientific Study of Reading, Vancouver, Canada.

Yore, L. (1991). Secondary science teachers' attitudes toward and beliefs about science reading and science textbooks. *Journal of Research in Science Teaching, 28*(1), 55-72. doi: 10.1002/tea.3660280106

Yore, L., & Shymansky, J. (1991). Reading in science: developing an operational conception to guide instruction. *Journal of Science Teacher Education, 2*(2), 29-36. doi:10.1007/BF02962849

[a]**Karen Thompson** is an Assistant Profesor at the College of Education, Oregon State University, United States of America.

Diego X. Román is a Ph.D. Candidate in the Educational Linguistics Program at the Stanford University School of Education, United States of America.

Greg Wientjes graduated with a Ph.D. from the Stanford Center for Innovations in Learning, Stanford University School of Education, United States of America.

Kenji Hakuta is the Lee L. Jacks Professor of Education, Stanford University School of Education, United States of America.

Contact Address: Oregon State University, 301B Furman Hall, Corvallis, Oregon, 97331-3502, United States of America. Email: Karen.Thompson@oregonstate.edu