**Comprehension in a Scenario-based Assessment:
Domain and Topic-Specific Background Knowledge**

Kathryn S. McCarthy[1], Tricia A. Guerrero[1], Kevin M. Kent[1], Laura K. Allen[1], Danielle S. McNamara[1], Szu-Fu Chao[2], Jonathan Steinberg[2], Tenaha O'Reilly[2], John Sabatini[2]

[1]Arizona State University
[2]Educational Testing Service

Corresponding Author:
Kathryn S. McCarthy
Institute for the Science of Teaching and Learning
Arizona State University
P.O. Box 872111
Tempe, Arizona 85287-2111
USA
ksmccar1@asu.edu

Abstract

Background knowledge is a strong predictor of reading comprehension; yet little is known about how different types of background knowledge affect comprehension. The study investigated the impacts of both domain and topic-specific background knowledge on students' ability to comprehend and learn from science texts. High school students ($n = 3650$) completed two background knowledge assessments, a pretest, comprehension tasks, and a posttest, in the context of the Global, Integrated, Scenario-based Assessment (GISA) on Ecosystems. Linear mixed effects models revealed positive effects of background knowledge on comprehension and learning as well as an interactive effect of domain and topic-specific knowledge, such that readers with high domain knowledge, but low topic-specific knowledge improved most from pretest to posttest. We discuss the potential implications of these findings for educational assessments and interventions.

Comprehension in a Scenario-based Assessment:
Domain and Topic-Specific Background Knowledge

Background knowledge is one of the strongest predictors of comprehension (Alexander 1992; Dochy, Segers, & Buell, 1999; Goldman & Rakestraw, 2000; Shapiro, 2004). Information or experiences related to a text provide a more organized structure into which new information from the text can be integrated (Kintsch, 1988, 1998; Mandler, 1984) and also provide the necessary resources for inferences to be generated (McNamara & Kintsch, 1996). Consequently, readers who have more background knowledge, or prior knowledge, about the topic of a text are able to process the information more quickly, remember more of the information, understand the information at a deeper level, and more effectively ignore irrelevant information (e.g., Alexander, Kulikowich, & Schulze, 1994, Bransford & Johnson, 1972; Haenggi & Perfetti, 1994; McNamara & Kintsch, 1996; McNamara, Kintsch, Songer, & Kintsch, 1996; McNamara & McDaniel, 2004; Spilich, Vesonder, Chiesi, & Voss, 1979).

Background Knowledge in Comprehension

According to multiple theories of text comprehension, background knowledge is critical to the construction of a mental representation of text's content (e.g., Gernsbacher, 1997; Graesser, Singer, & Trabasso, 1994; Kintsch, 1988, 1998; van den Broek, Young, Tzeng, & Linderholm, 1999). An assumption underlying these theories is that the mental representation, or mental model (Johnson-Laird, 1983), can be decomposed into multiple levels. According to the Construction-Integration model (Kintsch, 1988), these levels include the *surface code* reflecting the exact wording of the text, the *textbase*, or gist of the text, and the *situation model* that reflects the meaning of the text. As such, a reader's mental representation of a text includes the facts and details presented in the text (i.e., surface code and textbase levels of the representation), as well as inferences that are generated to connect information across the text and integrate information

with background knowledge (i.e., situation model level). Thus, a reader can formulate a more elaborated mental representation of the text if they have adequate background knowledge on the topic.

Readers who are provided with relevant information about the material demonstrate better memory for the text (Bransford & Johnson, 1972), suggesting that background knowledge enhances both surface code and textbase construction. Providing background knowledge also improves situation model comprehension as indicated by performance on inference-based comprehension assessments in a variety of domains including science (e.g., Alvermann, Smith, & Readence, 1985; McNamara & Kintsch, 1996), history (e.g., McKeown, Beck, Sinatra, & Loxterman, 1992) and literature (e.g., McCarthy & Goldman, in press).

In addition to manipulating the availability of this background knowledge, research has also evaluated students' existing knowledge of text content prior to reading. While some studies ask students to self-report their familiarity with a topic, having students complete open-ended, multiple-choice, or recognition assessments yields estimates with greater validity (Dochy et al., 1999; Shapiro, 2004). Many studies have reported evidence that background knowledge is moderately to strongly-related to comprehension test performance (Dochy et al., 1999; Shapiro, 2004). Nonetheless, background knowledge is not always beneficial to comprehension, particularly if it solely comprises information provided by the researcher (McNamara & Kintsch, 1996) or the students' knowledge includes misconceptions (Kendeou & van den Broek, 2007).

Does the type of background knowledge matter?

Although relations between background knowledge and text comprehension have been widely investigated in a variety of domains (e.g., Alexander, Sperl, Buehl, Fives, & Chiu, 2004; Bigot & Rouet, 2007; Cromley & Azevedo, 2007; Dochy, Segers, & Buehl, 1999; Earthman, 1992; Fincher-Kiefer, Post, Greene, & Voss, 1988; Graves & Frederiksen, 1991; McCarthy &

Goldman, in press; McNamara, 2001; McNamara, Kintsch, Songer, & Kintsch, 1996; Murphy & Alexander, 2002; O'Reilly & McNamara, 2007a, 2007b; Ozuru, Best, Bell, Witherspoon, & McNamara, 2007; Peskin, 1998; Rouet, Favart, Britt, & Perfetti, 1997; Shapiro, 2004; Thompson & Zamboanga, 2004; Voss & Silfies, 1996; Walker, 1987; Wiley, George, & Rayner, 2016; Wineburg, 1991) there remains considerable uncertainty as to how to characterize the relations between background knowledge and reading comprehension. Background knowledge can be decomposed into a variety of types (Alexander, Kulikowich, & Schulze, 1994; de Jong & Ferguson-Hessler, 1996; Wineburg, 1997). For instance, a person may be able to identify a wide variety of vegetables, but have little knowledge about which ones to put together for a recipe; similarly, a person may know a great deal about American History in general, but have limited knowledge about the Seneca Falls Convention. Thus, a limitation in the extant work is that background knowledge is often considered as a single dimension, or, at the very least, tends to be assessed as such.

Take for example the science knowledge test used in O'Reilly and McNamara (2007a). The 18-item test was designed to assess general science knowledge and consisted of questions on a variety of scientific topics, including scientific tools, forms of energy, space, scientific inquiry, earth science, and mathematics. Students' performance on this background knowledge test predicted their performance on a standardized science test and a comprehension test on the specific topics of air mass and weather fronts. This work demonstrates the importance of science background knowledge for the comprehension of specific topics within the science domain. However, the study could not specify whether those relations were being driven by domain (e.g., science) or topic-specific (e.g., weather) knowledge. It could be the case that performance on a

particular topic within the background knowledge test (i.e., topic-specific knowledge) was driving the correlations with the comprehension assessment.

There are only a few studies that have differentiated the effects of domain and topic-specific background knowledge on comprehension and learning. Some have employed an expert/expert paradigm to describe how domain experts with different topic or content-specific knowledge leverage what they know to understand text (Wineburg, 1998; Warren, 2011). In science comprehension, Alexander and colleagues (1994; see also Alexander, Schallert, & Hare, 1991) measured undergraduate and graduate students' domain and topic-specific knowledge. They then asked the students to read two physics texts on different topics (quarks, grand unification theory). Analyses indicated that both domain and topic-specific knowledge played a role in students' performance. However, domain knowledge was a stronger predictor of interest and recall.

Though these findings demonstrate that both domain and topic-specific knowledge relate to reading comprehension, an important consideration is that an individual student may possess varying amounts of both types of knowledge, which could affect reading comprehension. High domain knowledge in science likely reflects that the student knows about a variety of science topics (See O'Reilly & McNamara, 2007b). However, it may be the case that a high domain knowledge student has, by chance, never encountered a particular science topic before. Alternatively, there are certainly some individuals who have specific knowledge about a topic, without having strong domain knowledge. Presumably having these different amounts of knowledge might affect the reader's mental representation of new texts.

Undoubtedly, students with both low domain and low topic knowledge are at a disadvantage, but it is an empirical question as to which type of knowledge is more critical for

learning from text. It could be hypothesized that a student with high topic-specific knowledge would be best suited to learn from the text because of their familiarity with the specific terminology, which is a particularly salient challenge for scientific texts (McNamara, Graesser, & Louwerse, 2012). Topic knowledge may be necessary for the generation of appropriate inferences needed for a coherent mental representation. Alternatively, one might predict that a student with extensive domain knowledge would be able to overcome limited familiarity with specific terms related to the topic due to a rich, coherent knowledge base that can provide a structure for new information to be organized. Thus, one purpose of the current study is to evaluate not only the unique contributions of different types of background knowledge, but also to consider how varying degrees of each type of knowledge interactively contribute to readers' comprehension of text content.

Assessing Reading Comprehension

Further complicating our understanding of the relations between background knowledge and comprehension is that educators, policy makers and researchers have advocated for an updated construct of reading comprehension that better reflects the types of materials and processes that characterize modern literacy demands (Goldman et al., 2016; Leu et al., 2013; NGA & CCSSO, 2010). In modern reading contexts, individuals often have specific purposes for reading a collection of source materials (Britt, Rouet, & Durik, 2017; van den Broek, Lorch, Linderholm, & Gustafson, 2001). Such goals help readers to set standards of coherence (van den Broek, Young, Tzeng, & Linderholm, 1999), which in turn help learners identify what information is relevant for their specific reading goals (McCrudden, Magliano, & Schraw, 2010). As a part of this process, people often have to evaluate, integrate, and synthesize multiple sources that are increasingly digital in nature (Britt et al., 2017; Goldman et al., 2016; Leu et al.,

2013; Metzger, 2007). The integration process may also involve complex reasoning in order to understand multiple perspectives on events, issues, and causal mechanisms (LaRusso et al., 2016). Furthermore, with the adoption of the Common Core State Standards, researchers and educators have argued that reading comprehension should encompass content area texts and disciplinary literacy (Goldman et al. 2016; NGA & CCSSO, 2010). Given these notions of what it takes to comprehend a text, the roles of background knowledge are likely highly complex and increasingly important.

Unfortunately, most existing high-stakes assessments are not designed to evaluate reading comprehension as a complex integration of various sources of content, and are generally constructed, explicitly, to ignore the contributions of background knowledge. In such assessments, students read a wide range of passages about topics that are intended to be familiar to most students. The logic driving this approach is that students may know more about some passages than others, but across the test as a whole, the impact of background knowledge "should" be mitigated over the course of the assessment (Cromley & Azevedo, 2007; Snow, 2002). In other words, these assessments treat background knowledge as construct-irrelevant, which is inconsistent with more modern notions of reading comprehension skill (Shapiro, 2004).

In contrast to these traditional standardized tests, the *Global, Integrated Scenario-based Assessment*, or GISA, provides a theory-driven assessment designed to target real-world reading and learning activities (O'Reilly & Sabatini, 2013; Sabatini, O'Reilly, & Deane, 2013). The GISA is a purpose-driven assessment that measures students' ability to integrate, evaluate, and synthesize multiple sources in a digital environment. Students are given a purpose for reading a collection of texts on specific scientific or social/historic topics. Given the nature of the topics, students come to the task with varying levels of background knowledge. Critically, the GISA

does not attempt to eliminate these effects, but rather accounts for them by evaluating students'

knowledge as part of the assessment. As such, the environment provides a rich forum to assess

the relations between background knowledge and comprehension that can provide insight into

contemporary discourse comprehension.

GISAs have been produced and evaluated for elementary (Sabatini, Halderman, O'Reilly,

& Weeks, 2016), middle (Sabatini, O'Reilly, Halderman, & Bruce, 2014) and high school

students (O'Reilly, Weeks, Sabatini, Halderman, & Steinberg, 2014). Collectively, these studies

indicate the assessments are reliable and produce a range of scores with no apparent floor or

celling effects. GISAs correlate with the prior year's English Language Arts state test scores

ranging from .52 to .68 (O'Reilly et al., 2014) and correlate with measures of academic

vocabulary, complex reasoning, and perspective taking (LaRusso et al., 2016).

In the GISA, test takers "interact" with simulated peer students to model, support and

assess test-taker understanding as they engage in complex reasoning, perspective taking,

application, and sometimes disciplinary literacy tasks. These tasks are designed to measure basic

understanding, such as the ability to locate information and to draw inferences, as well as more

demanding tasks that require applying concepts and principles to new situations, solving

problems, or making decisions. The students work in a "group", in which avatars act as

pedagogical agents, representing classmates or subject-matter experts to present information and

questions with varying levels of support. Notably, the agents provide different amounts of

scaffolding on each text and item, but this scaffolding is the same for every student (i.e., the

system is not adaptive) and students proceed through the assessment in uniform order.

Though generally similar, GISA forms are not identical in terms of the number and types

of items. Instead, each was designed to specifically target relevant content and skills for each

topic. In the assessment form used in this study (Ecosystems), there are seven sections that include a vocabulary test designed to assess knowledge in the domain of ecosystems, a "native species vs. invasive species" identification task to assess topic-specific background knowledge, a three-item multiple-choice pretest, a series of activities involving reading texts and answering questions (comprehension tasks), and a learning check. As a comprehension assessment, the GISA is designed to yield a single composite score that evaluates a variety of text and content-based skills. However, for the purposes of this empirical investigation, we decomposed the background knowledge score resulting in two predictors (a domain knowledge test score and a topic-specific knowledge test score) and three comprehension outcomes: pretest (Section 1), comprehension tasks (Sections 2-6), and posttest (Section 7).

The Current Study

The current study investigates the role of both domain and topic-specific background knowledge on performance in a content-based reading comprehension assessment. We first investigated the degree to which these two types of knowledge could be independent of one another, demonstrating that students were not exclusively considered high or low background knowledge, but that students could possess a large amount of one type of knowledge while simultaneously having little of the other type. Consistent with the extant work, it was predicted that both domain (ecosystems) and topic-specific (invasive species) background knowledge would positively predict the comprehension outcomes in the GISA. Critically, a large sample size of US high school students afforded the ability to test the hypothesis that varying amounts of each type of knowledge interactively contributes to students' comprehension. That is, we investigated the degree to which students are able to compensate for low knowledge of ecology by having familiarity with certain invasive species and, vice versa, the possibility that students

only benefit from having knowledge about invasive species if they also have sufficient general ecology knowledge. We predicted an additive effect of the two types of knowledge, such that students with more of both types would perform better than those with less knowledge. We also predicted that as topic knowledge decreased, domain knowledge would be more important for comprehension as it provides a means of organizing and integrating new information.

Method

Participants

Participants in this study were 9th to 12th grade students who completed the Ecosystems comprehension form in the context of a larger project conducted from 2011-2014 in both California and Pennsylvania (Fancsali et al., 2015). The Ecosystems form was administered to 4483 students across 37 schools. This analysis includes only those who completed all seven sections of the assessment ($n = 3560$). Of these students, 33.2% were in 9th grade, 50.7% in 10th grade, 11.7% in 11th grade, and 4.4% in 12th grade. Further demographic data for this subset is not available, but in the larger sample 39.5% of students qualified for free or reduced lunch, 49.3% identified as nonwhite, and 10.4% identified as English Language Learners (Fancsali et al., 2015).

Materials

Ecosystems GISA

The entire study was conducted within the GISA interface. The scenario for this form of the assessment is that the student is a member of a study group that is preparing for an upcoming Ecology test. Students complete seven sections. Section 1 consists of the two background knowledge tests and the pretest. Sections 2 through 6 consist of different comprehension

activities. Finally, Section 7 includes the learning check, which we refer to as the posttest (A sample of the GISA interface including texts and items appears in Appendix A)

Background Knowledge Tests. Section 1 of the GISA includes two background knowledge tests. Participants were told that these questions would not count towards their final score and were always able to select "I don't know" as an option. The *domain knowledge* test was a 44-item vocabulary recognition test (for more information the development of this assessment, see Deane, 2012; McKeown, Deane, Scott, Krovetz, & Lawless, 2017; O'Reilly, Sabatini, & Wang, under revision). In this test, participants identify each word as "related", "not related", or "don't know" to the topic of ecosystems. Students receive 1 point for correctly identifying the word as related or unrelated and receive 0 points for an incorrect identification or selecting "don't know", resulting in a possible score from 0-44 ($\alpha = .92$).The *topic-specific knowledge* test was comprised of 8 items. Participants were presented with a species and were tasked with identifying the species as "invasive" or "native". Again, participants were able to select "don't know". Student received 1 point for a correct categorization, resulting in a score from 0-8 ($\alpha = .66$).

Pretest and Posttest. The pretest (Section 1) and posttest (Section 7) consisted of the same three multiple-choice items on the topic of invasive species. These questions were answered without the texts present. Each question was designed to tap different levels of comprehension. The first question asked students to identify the appropriate definition of an invasive species. This information could be found directly from one of the texts read as they moved through the assessment. The second question required the students to make connections across multiple sentences to arrive at the answer. The third question asked about information that was not directly available in the texts and required readers to generate an elaborative inference

connecting information from outside the text to what was mentioned explicitly. Given the brevity

of the pretest and posttest, the raw reliability values were adjusted using the Spearman-Brown

formula using the domain knowledge test as a reference point (pretest: $\alpha = .76$, posttest: $\alpha = .56$).

Texts and Comprehension Task. Sections 2 through 6 of the GISA were comprised of

comprehension tasks in which students read a text and answered a series of questions with the

text available to them. The GISA was composed of five primary passages, ranging in length from

84 to 589 words. The reading ease levels of the passages varied widely, with Flesch Reading

Ease scores from 20.5 to 56.9 and Flesch-Kincaid Grade Level scores from 9.0 to 16.7. Students

were asked to summarize important information (Section 2), consider evidence and relate it to

scientific policy (Section 3), understand and apply scientific terms (Section 4), paraphrase

(Section 5), and review scientific data (Section 6). These questions were a mix of multiple-

choice and open-ended items. Human raters scored open-ended items from 0 to 3 and multiple-

choice questions were automatically scored as 0 or 1. The total points possible in the

comprehension task was 40 ($\alpha = .80$).

<div align="center">

Results

</div>

Preliminary Analyses

Table 1 provides the average score on each section of the assessment. These data show

normal distributions with neither floor nor ceiling effects. Importantly, pretest scores are

relatively low, suggesting that students do not appear to have prior mastery of the topic.

Table 2 displays a distribution of students as a function of high and low background

knowledge. As shown in the table, most participants had both low domain and topic-specific

background knowledge, but there was sufficient variability across the sample to investigate both

main effects and interactions between the two types of knowledge. Importantly, these median

splits were calculated only to provide a simplified description of the sample and not to conduct inferential analyses. The remaining analyses consider both types of background knowledge as continuous variables.

Central to our predictions are the relations between the two types of background knowledge and the three comprehension sections of the GISA (pretest, comprehension tasks, posttest). The correlations are provided in Table 3. Consistent with previous research (Alexander et al., 1994), students' domain and topic-specific background knowledge were both positively correlated with the comprehension outcomes (pretest, comprehension tasks, posttest). Hence, this correlational analysis supports the existing research demonstrating that background knowledge supports comprehension and learning from complex science text (e.g., Alexander et al., 1994; O'Reilly & McNamara, 2007a). It also demonstrates that students possess varying amounts of different types of background knowledge.

Interestingly, more advanced students did not consistently perform better on the background knowledge tests or the comprehension outcomes (Table 3). There was a weak, but significant positive correlation between grade and topic-specific knowledge indicating that the older students had more knowledge of invasive species than their younger counterparts. However, there was a weak negative correlation between grade and domain knowledge, indicating that the older students had less knowledge about ecosystems than the younger students. There was no relation between grade and pretest score. Surprisingly, grade was negatively related to performance on both the comprehension tasks in Sections 2-6 and the posttest. As described in more detail in the discussion, these negative correlations between grade and knowledge and grade and comprehension score may be a result of a sampling bias regarding the older students.

Interactive Effects of Background Knowledge

Analyses were conducted to examine the unique contributions of both types of background knowledge as well as their potential interactive effects. Given that both domain and topic-specific background knowledge were related to pretest performance, pretest was included as a covariate.

Comprehension Tasks. In Sections 2-6 of the GISA, participants read texts and answered corresponding questions with support from their pedagogical agent classmates. Students had the text available to them as they answered these open-ended and multiple-choice items. As shown in Table 1, the average percent correct was 43%.

We conducted a linear mixed-effects model analysis with domain knowledge z-score, topic-specific knowledge z-score, and grade as fixed factors and random intercepts for participants nested within school. Pretest score was included as a covariate. As suggested by Baayen, Davidson, and Bates (2008), Table 4 provides the significance tests for the fixed effects rather than a detailed description of the model.

Both grade and domain knowledge significantly predicted comprehension task performance, whereas topic-specific background knowledge did not. Further, the domain by topic-specific background knowledge interaction term was not a significant predictor. These results suggest that domain knowledge plays a more critical role in comprehension than knowledge about the specific topic. There was, however, a small, but significant three-way interaction between domain knowledge, topic-specific background knowledge, and grade. Follow-up analyses indicated that the interaction between domain and topic-specific background knowledge was only significant for the 12th graders, $t = 2.34$, $p < .05$.

Posttest. As shown in Table 1, average score increased from pretest to posttest. A linear mixed-effects analysis assessed the effects of the two types of background knowledge on posttest score. Domain knowledge z-score, topic-specific knowledge z-score, and grade were entered as fixed factors and participants nested within school were entered as random intercepts. Pretest was entered as a covariate. Significance tests for the fixed effects appear in Table 5.

Interestingly, neither domain nor topic-specific knowledge independently predicted posttest performance. Grade was the only significant main effect. However, as indicated by the negative beta coefficient, participants in the lower grades outperformed those in the upper grades. There was a significant domain by topic-specific knowledge interaction. Analyses of simple effects indicated that for participants with high domain knowledge, there was a non-significant, but negative slope of topic-specific knowledge ($t = -1.68$, *ns*). In contrast, for participants with low domain knowledge, there was a non-significant, but positive slope ($t = 1.75$, *ns*). To more clearly represent the nature of this interaction, Figure 1 plots average pretest and posttest score as a function of the four median-split groups: 1) high domain, high topic-specific, 2) high domain, low topic-specific, 3) low domain, high topic-specific, and 4) low domain, low topic-specific. This representation indicates that the domain by topic-specific knowledge interaction is driven by gains for those participants who had high domain knowledge, but low topic-specific knowledge. There was also a significant domain by topic-specific by grade level three-way interaction. Follow up analyses indicate that there was a significant positive interaction between domain and topic-specific background knowledge for the 9th graders, $t = -1.25$, $p < .05$ and a significant inverse interaction for the 12th graders, $t = 2.00$, $p < .05$. This interaction was not significant for the 10th and 11th graders.

To summarize, grade appeared to have an inverse relation with performance on the comprehension tasks and posttest. Both domain and topic-specific background knowledge were correlated with performance on the GISA. For the comprehension tasks in the GISA, domain background knowledge was the sole predictor of performance. At posttest, there was a significant domain by topic-specific interaction such that, students with high domain background knowledge, but low topic-knowledge gained the most from the texts and tasks.

Discussion

This study investigated the relative contributions of domain and topic-specific background knowledge on science comprehension in the context of the Global, Integrated Scenario-based Assessment (GISA). The GISA seeks to reflect a more modern view of reading comprehension through real-world, content-based activities. As such, the assessment provided a unique opportunity to examine the interactive effects of these two types of background knowledge in an ecologically valid task. The large sample of students allowed us to detect small, but significant contributions of background knowledge to performance.

Consistent with the extant lab research (e.g., O'Reilly & McNamara, 2007a), correlational analyses revealed positive relations between students' background knowledge and performance on all three aspects of the assessment (pretest, comprehension tasks, and posttest). Domain knowledge was more strongly correlated with comprehension scores than was topic-specific. The topic-specific knowledge assessment was substantially shorter than the domain knowledge assessment, which potentially contributes to its somewhat lower reliability. This may also contribute to the lower correlation with comprehension scores. Nonetheless, the stronger relation of domain knowledge is consistent with the previous work conducted by Alexander and colleagues (1994).

The novel contribution of this study is the investigation of the potential interactive effects of domain and topic-specific knowledge on reading comprehension and learning. In the comprehension tasks, students had the texts available and were provided scaffolding from the pedagogical agents. In these tasks, only domain background knowledge was a significant predictor of performance. The posttest required students to answer questions without the texts and without scaffolding. Neither domain nor topic-specific knowledge independently predicted posttest performance. Critically, this was qualified by a significant interaction between domain and topic-specific knowledge. Students with high domain background knowledge, but low topic knowledge "caught up" to those who had both high domain and high topic-specific knowledge at the outset. In this study, students with high domain and high topic-specific knowledge were not at ceiling; yet, those with high domain and low topic-specific knowledge performed equally well on the posttest. This suggests that domain prior knowledge may benefit reading comprehension and learning processes more strongly than topic-specific knowledge.

From a theoretical perspective, topic-specific knowledge provides the raw materials from which inferences can be generated, whereas domain knowledge affords a more coherent network of connections to related knowledge and general knowledge (e.g., McNamara & Kinstch, 1996). Inferences based on topic-specific knowledge are crucial to comprehension; however, when text-to-knowledge inferences are not well connected to one another, it is theoretically more difficult to apply the information in more complex comprehension tasks. Domain knowledge affords the generation of inferences, but also provides a more coherent structure, facilitating the incorporation of incoming, new information (e.g., Kintsch, 1988, 1998; Mandler, 1984). Hence, domain knowledge may contribute more to deep comprehension of the material.

Further, the results demonstrate the limitations of assessing background knowledge as a single construct. The differential and interactive effects of domain and topic knowledge highlight the need to conceptualize background knowledge as multi-faceted and to carefully describe the nature of the knowledge being evaluated in these investigations.

Domain and topic-specific knowledge represent just a few of the types of background knowledge that could contribute to comprehension. For example, one could imagine that both are part of a larger hierarchy that includes broader knowledge (e.g., general science) and more specific knowledge (e.g., knowledge about a particular invasive species). At any of these various "grain sizes", students may have basic knowledge (e.g., identify or recall specific dates or definitions), but they may lack a deep understanding of the relations between these ideas. As a first step into this exploration, we have developed and refined basic and conceptual background knowledge items for general history and science as well as for specific topics within these domains. We are compiling large-scale data sets that can address how students' basic and conceptual knowledge relate to GISA performance. Establishing the unique contributions of these various forms of background knowledge will help to better model how students process text in content-based comprehension tasks. This information can further refine discourse theories in terms of how background knowledge affects the quality of the mental model. It can also permit the development of interventions that scaffold support based on individuals' strengths or weaknesses.

One counterintuitive result was that students in the lower grades outperformed those in the higher grades on both the comprehension tasks and the posttest. There were also complex three-way interactions between grade and the two types of background knowledge. This pattern of results is consistent across other GISA data sets and may be indicative of issues of motivation

across the grade levels (O'Reilly, Wang, Sabatini, Steinberg, & Weeks, in preparation). There may also be more practically oriented explanations for these results. The inclusion criteria for this project was that the schools were required to include teachers who taught "9th grade ELA, 10th grade biology, or 11th grade U.S. History" (Fancsali et al., 2015). This explains the higher proportion of 10th grade students relative to 11th and 12th graders in this biology-related test. This imbalance of grade distribution likely contributes to our results. Without more specific student or classroom-level data, we cannot directly speak to whether there are further sampling biases responsible for the differences across grades.

Of course, these results should be replicated across other scientific domains as well as other subjects (e.g., history, art) before drawing strong conclusions. Not all GISAs follow the same structure and so we cannot immediately replicate these findings with extant data sets. For example, the social science GISA on Immigration includes a domain background knowledge vocabulary test, but does not include a topic-specific test. Nonetheless, we intend to evaluate the role of different types of background knowledge in these forms as data become available and we encourage others to investigate how these types of background knowledge relate to performance using other comprehension measures.

Notably, traditional reading comprehension assessments and verbal ability measures (e.g., Gates-MacGinitie, Nelson-Denny) were not collected within our current study. However, the GISA is strongly correlated with performance on other standardized reading comprehension assessments and performance varies as a function of lower level reading processes such as decoding and word fluency (Sabatini, O'Reilly, Halderman, & Bruce, 2014). Nonetheless, it will be informative in future work to collect such measures of reading skill in conjunction with the

GISA to better disentangle the contributions of both general reading skill and domain-specific content learning to reading comprehension.

In addition to further investigations with the GISA, it would be of value to pursue more targeted investigations in using traditional lab-based comprehension tests. Researcher-designed measures can be constructed to evaluate broader content coverage and include more specific items to assess the relative quality of the different levels (e.g., surface, textbase, situation model) of the readers' mental representation. Such analyses would further elucidate *how* background knowledge supports comprehension and learning from text.

In conclusion, this study demonstrated that students vary not only in their amount of background knowledge, but also the type of background knowledge. By understanding how different dimensions and combinations of background knowledge impact comprehension, researchers and educators can develop individualized supports that optimize learning content from texts. As such, this study provides foundational research on which to build a better understanding of the complex effects of knowledge on comprehension.

References

Alexander, P. A. (1992). Domain knowledge: Evolving themes and emerging

concerns. *Educational Psychologist*, *27*(1), 33-51.

Alexander, P. A., Kulikowich, J. M., & Schulze, S. K. (1994). The influence of topic knowledge,

domain knowledge, and interest on the comprehension of scientific exposition. *Learning

and Individual Differences, 6*, 379-397.

Alexander, P. A., Schallert, D. L., & Hare, V. C. (1991). Coming to terms: How researchers in

learning and literacy talk about knowledge. *Review of Educational Research, 61,* 315-

343.

Alexander, P. A., Sperl, C. T., Buehl, M. M., Fives, H., & Chiu, S. (2004). Modeling domain

learning: Profiles from the field of special education. *Journal of Educational Psychology,

96*(3), 545-557.

Alvermann, D. E., Smith, L. C., & Readence, J. E. (1985). Prior knowledge activation and the

comprehension of compatible and incompatible text. *Reading Research Quarterly*, 420-

436.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed

random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412.

Bigot, L. L., & Rouet, J. F. (2007). The impact of presentation format, task assignment, and prior

knowledge on students' comprehension of multiple online documents. *Journal of Literacy

Research,* 39(4), 445-470.

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some

investigations of comprehension and recall. *Journal of Verbal Learning and Verbal

Behavior*, *11*(6), 717-726.

Britt, M. A., Rouet, J. F., & Durik, A. M. (2017). *Literacy Beyond Text Comprehension: A Theory of Purposeful Reading*. New York, NY: Routledge.

Cromley, J., & Azevedo, R. (2007). Testing and refining the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology, 99*, 311-325.

Deane, P. (2012). NLP methods for supporting vocabulary analysis. In J. Sabatini, T., O'Reilly, & L. Albro (Eds.), *Reaching an understanding: Innovations in how we view reading assessment* (pp. 117-146). Lanham, MD: Rowman & Littlefield.

De Jong, T., & Ferguson-Hessler, M. (1996). Types and qualities of knowledge. *Educational Psychologist, 31*, 105-113.

Dochy, F., Segers, M., & Buehl, M. M. (1999). The relation between assessment practices and outcomes of studies: The case of research on prior knowledge. *Review of Educational Research, 69*(2), 145-186.

Earthman, E. (1992). Creating the virtual work: Readers' processes in understanding literary texts. *Research in the Teaching of English. 26,* 351-384.

Fancsali, C., Abe, Y., Pyatigorsky, M., Ortiz, L., Hunt, A., Chan, V., Saltares, E., Toby, M., Schellinger, A., & Jaciw, A. P. (2015). The impact of the Reading Apprenticeship Improving Secondary Education (RAISE) Project on academic literacy in high school: A report of a randomized experiment in Pennsylvania and California schools. (Empirical Education Rep. No. Empirical_ RAISE-7019-FR1-O.2). Palo Alto, CA: Empirical Education Inc.

Fincher-Kiefer, R., Post, T. A., Greene, T. R., & Voss, J. F. (1988). On the role of prior knowledge and task demands in the processing of text. *Journal of Memory and Language*, *27*(4), 416-428.

Gernsbacher, M. A. (1997). Two decades of structure building. *Discourse Processes*, *23*(3), 265-304.

Goldman, S. R., Britt, M. A., Brown, W., Cribb, G., George, M., Greenleaf, C., ... & Project READI. (2016). Disciplinary literacies and learning to read for understanding: A conceptual framework for disciplinary literacy. *Educational Psychologist*, *51*(2), 219-246.

Goldman, S.R., & Rakestraw, J.A. (2000). Structural aspects of constructing meaning from text. In M.L. Kamil, P.B. Mosenthal, P.D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 311–335). Mahwah, NJ: Erlbaum.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*(3), 371-395.

Graves, B. & Frederiksen, C. H. (1991). Literary expertise in the description of fictional narrative. *Poetics, 20,* 1-26.

Haenggi, D., & Perfetti, C. A. (1994). Processing components of college-level reading comprehension. *Discourse Processes, 17*(1)*,* 83-104.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness.* Cambridge, MA: Harvard University Press.

Kendeou, P., & van den Broek, P. (2007). The effects of prior knowledge and text structure on comprehension processes during reading of scientific texts. *Memory & Cognition*, *35*(7), 1567-1577.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review,* 95, 163-182.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition.* Cambridge, UK: Cambridge

    University Press.

LaRusso, M., Kim, H. Y., Selman, R., Uccelli, P., Dawson, T., Jones, S., ... & Snow, C. (2016).

    Contributions of academic language, perspective taking, and complex reasoning to deep

    reading comprehension. *Journal of Research on Educational Effectiveness*, *9*(2), 201-

    222.

Leu, D. J., Forzani, E., Burlingame, C., Kulikowich, J., Sedransk, N., Coiro, J., & Kennedy, C.

    (2013). The new literacies of online research and comprehension: Assessing and

    preparing students for the 21st century with Common Core State Standards. *Quality

    reading instruction in the age of Common Core Standards*, 219-236.

Mandler, G. (1984). *Mind and Emotion.* New York, NY: Wiley.

McCarthy, K. S., & Goldman, S. R. (in press). Constructing interpretive inferences about literary

    text: The role of domain-specific knowledge. *Learning and Instruction.*

McCarthy, K. S., Kopp, K. J., Allen, L. K., & McNamara, D. S. (in press). Methods of studying

    text: Memory, comprehension, and learning. In H. Otani & B. Schwarz (Eds.), *Research

    methods in human memory.* New York, NY: Routledge.

McCrudden, M. T., Magliano, J. P., & Schraw, G. (2010). Exploring how relevance instructions

    affect personal reading intentions, reading goals and text processing: A mixed methods

    study. *Contemporary Educational Psychology*, *35*(4), 229-241.

McKeown, M. G., Beck, I. L., Sinatra, G. M., & Loxterman, J. A. (1992). The contribution of

    prior knowledge and coherent text to comprehension. *Reading Research Quarterly*,

    27(1), 79-93.

McKeown, M. G., Deane, P. D., Scott, J. A., Krovetz, R., & Lawless, R. R. (2017). *Vocabulary Assessment to Support Instruction: Building Rich Word-Learning Experiences*. New York, NY: Guilford Press.

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., Graesser, A.C., & Louwerse, M.M. (2012). Sources of text difficulty: Across genres and grades. In J.P. Sabatini, E. Albro, & T. O'Reilly (Eds.), *Measuring up: Advances in how we assess reading ability* (pp. 89-116). Lanham, MD: R&L Education.

McNamara, D. S., & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes, 22,* 247-288.

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

McNamara, D.S., & McDaniel, M.A. (2004). Suppressing irrelevant information: Knowledge activation or inhibition? *Journal of Experimental Psychology: Learning, Memory, & Cognition, 30*, 465-482.

Metzger, M. J. (2007). Making sense of credibility on the Web: Models for evaluating online information and recommendations for future research. *Journal of the Association for Information Science and Technology*, *58*(13), 2078-2091.

Murphy, P. K., & Alexander, P. A. (2002). What counts? The predictive powers of subject-matter knowledge, strategic processing, and interest in domain-specific performance. *The Journal of Experimental Education*, 70(3), 197-214.

National Governors Association Center for Best Practices, Council of Chief State School

Officers (2010). *Common Core State Standards for English Language Arts*. National

Governors Association Center for Best Practices, Council of Chief State School Officers,

Washington, D.C.

O'Reilly, T., & McNamara, D. S. (2007a). The impact of science knowledge, reading skill, and

reading strategy knowledge on more traditional "high stakes" measures of high school

students' science achievement. *American Educational Research Journal, 44,* 161-196.

O'Reilly, T., & McNamara, D.S. (2007b). Reversing the reverse cohesion effect: good texts can

be better for strategic, high-knowledge readers. *Discourse Processes, 43*, 121-152.

O'Reilly, T., & Sabatini, J. (2013). Reading for understanding: How performance moderators and

scenarios impact assessment design. *ETS Research Report Series*, *2013*(2).

O'Reilly, T., Sabatini, J., & Wang, Z. (under revision). *The impact of topical knowledge and

knowledge monitoring on the ability comprehend and learn from text.*

O'Reilly, T., Wang, Z., Sabatini, J., Steinberg, J. & Weeks (in preparation). *GISA validity

argument: Claims and supporting evidence.*

O'Reilly, T., Weeks, J., Sabatini, J., Halderman, L., & Steinberg, J. (2014). Designing reading

comprehension assessments for reading interventions: How a theoretically motivated

assessment can serve as an outcome measure. *Educational Psychology Review*, *26*(3),

403-424.

Ozuru, Y., Best, R., Bell, C., Witherspoon, A., & McNamara, D.S. (2007). Influence of question

format and text availability on assessment of expository text comprehension. *Cognition &

Instruction, 25*, 399-438.

Peskin, J. (1998). Constructing meaning when reading poetry: An expert-novice study. *Cognition and Instruction, 16,* 135-263.

Rouet, J. F., Favart, M., Britt, M. A., & Perfetti, C. A. (1997). Studying and using multiple documents in history: Effects of discipline expertise. *Cognition and Instruction, 15*(1), 85-106.

Sabatini, J., Halderman, L., O'Reilly, T., & Weeks, J. (2016). Assessing Comprehension in Kindergarten through Third Grade. *Topics in Language Disorders, 36*, 4, 334-355.

Sabatini, J., O'Reilly, T., & Deane, P. (2013). Rationale for a new generation of reading comprehension assessments. In B. Miller, L. Cutting, & P. McCardle (Eds.), *Unraveling the behavioral, neurobiological, and genetic components of reading comprehension*, (pp. 100–111). Baltimore, MD: Brookes.

Sabatini, J. P., O'Reilly, T., Halderman, L. K., & Bruce, K. (2014). Integrating scenario-based and component reading skill measures to understand the reading behavior of struggling readers. *Learning Disabilities, 29,* 36-43.

Shapiro, A. M. (2004). How including prior knowledge as a subject variable may change outcomes of learning research. *American Educational Research Journal, 41*, 159-189.

Spilich, G., Vesonder, G., Chiesi, H., & Voss, J, (1979). Text processing of domain related information for individuals with high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior, 18*, 275-290.

Snow, C. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Santa Monica, CA: RAND.

Thompson, R. A., & Zamboanga, B. L. (2004). Academic aptitude and prior knowledge as predictors of student achievement in introduction to psychology. *Journal of Educational Psychology*, *96*(4), 778-784.

van den Broek, P., Lorch, R. F., Linderholm, T., & Gustafson, M. (2001). The effects of readers' goals on inference generation and memory for texts. *Memory & Cognition*, *29*(8), 1081-1087.

van den Broek, P., Young, M., Tzeng, Y., & Linderholm, T. (1999). *The landscape model of reading*. In. H. van Oostendorp & S. R. Goldman (Eds.), The construction of mental representations during reading (pp. 71–98). Mahwah, NJ: Erlbaum.

Voss, J. F., & Silfies, L. N. (1996). Learning from history text: The interaction of knowledge and comprehension skill with text structure. *Cognition and Instruction*, *14*(1), 45-68.

Walker, C. H. (1987). Relative importance of domain knowledge and overall aptitude on acquisition of domain-related information. *Cognition and Instruction*, *4*(1), 25-42.

Warren, J. E. (2011). "Generic" and "specific" expertise in English: An expert/expert study in poetry interpretation and academic argument. *Cognition and Instruction, 29*(4), 349-374.

Wiley, J., George, T., & Rayner, K. (2016). Baseball fans don't like lumpy batters: Influence of domain knowledge on the access of subordinate meanings. *The Quarterly Journal of Experimental Psychology,* 1-11.

Wineburg, S. S. (1991). Historical problem solving: A study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology, 83(1),* 73-87.

Wineburg, S. S. (1997). Beyond "breadth and depth": Subject matter knowledge and assessment. *Theory Into Practice, 36,* 255-261.

Wineburg, S. (1998). Reading Abraham Lincoln: An expert/expert study in the interpretation of

historical texts. *Cognitive Science*, *22*(3), 319-346.
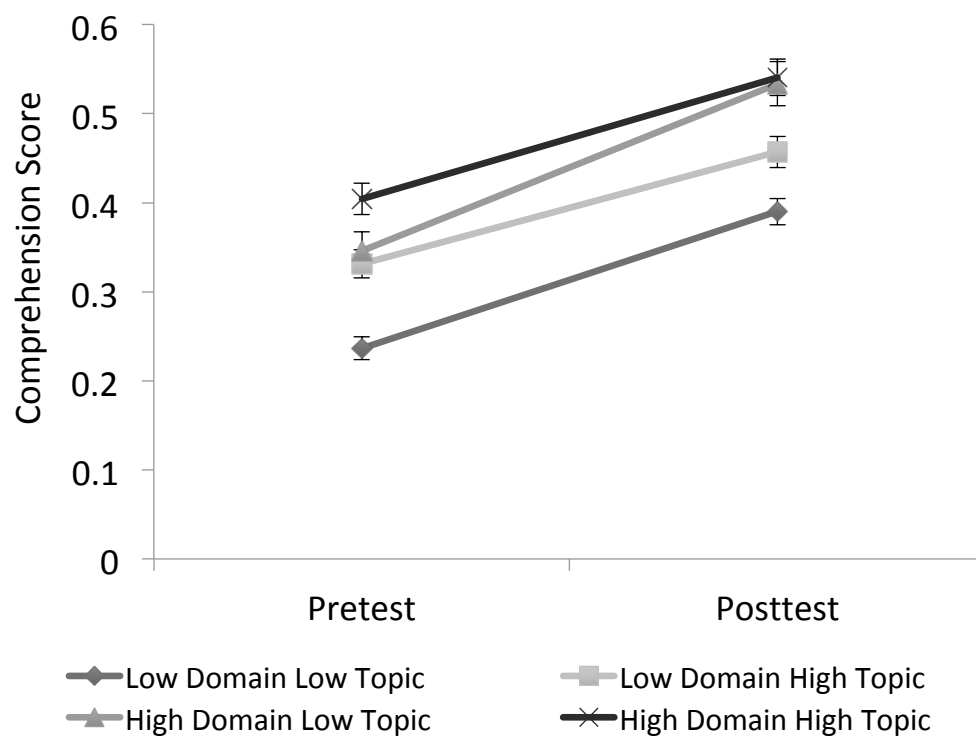
Figure 1. Pretest and posttest scores as a function of knowledge group determined by median-split scores. Error bars represent 95% confidence intervals.

Table 1.

Means and standard deviations of average scores as a function of grade

Grade	n	Domain BK	Topic Specific BK	Pretest	Sections 2-6	Posttest
		M(SD)	*M(SD)*	*M(SD)*	*M(SD)*	*M(SD)*
9	1183	.75(.20)	.41(.27)	.31(.25)	.48(.21)	.50(.30)
10	1806	.72(.21)	.43(.26)	.31(.25)	.41(.19)	.45(.29)
11	415	.71(.22)	.44(.26)	.31(.25)	.39(.19)	.43(.29)
12	156	.75(.21)	.47(.26)	.35(.27)	.43(.19)	.47(.31)
Total	**3560**	**.73(.21)**	**.43(.26)**	**.31(.25)**	**.43 (.20)**	**.47(.29)**

*scores are presented as proportion correct

Table 2.

Distribution of students as a function of high and low (median split) domain and topic-specific background knowledge

	Low Topic-Specific BK	High Topic-Specific BK
Low Domain BK	1412	501
High Domain BK	975	672

Table 3.

Correlation matrix for grade, background knowledge, and GISA scores

	1	2	3	4	5
1. Grade	--				
2. Domain Background Knowledge	-0.04*	--			
3. Topic-Specific Background Knowledge	0.05**	0.22**	--		
4. Pretest Score	0.01	0.24**	0.23**	--	
5. Comprehension Task Score	-0.13**	0.47**	0.15**	0.26**	--
6. Posttest Score	-0.07**	0.27**	0.09**	0.23**	0.43**

*$p < .05$, **$p < .01$

Table 4.

Coefficient table of multi-level model for score on Sections 2-6

	β	*SE*	*t*	*p*
Pretest	-0.04	0.00	-9.02	.00*
Domain background knowledge	0.11	0.04	2.54	.01*
Topic-Specific background knowledge	0.00	0.04	0.19	.85
Grade	0.02	0.00	6.90	.00*
Domain x Topic-Specific	-0.07	0.04	-1.81	.07
Domain x Grade	0.00	0.00	0.69	.49
Topic-Specific x Grade	0.00	0.00	-0.11	.91
Domain x Topic-Specific x Grade	0.01	0.00	2.03	.04*

Table 5.
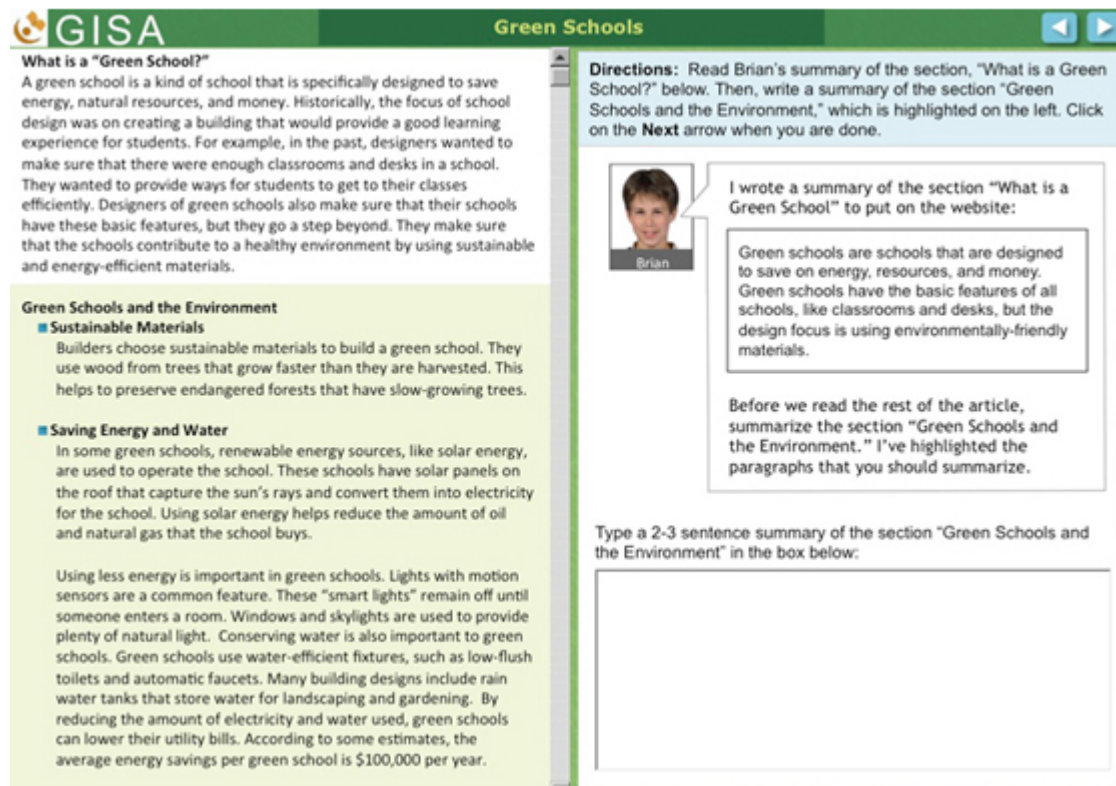
Coefficient table of mixed-effects model for posttest score

	β	SE	t	p
Pretest	-0.03	0.01	-4.22	.00*
Domain background knowledge	0.06	0.07	0.82	.41
Topic-Specific background knowledge	0.00	0.06	0.09	.93
Grade	0.04	0.01	8.84	.00*
Domain x Topic-Specific	-0.17	0.06	-2.99	.00*
Domain x Grade	0.00	0.01	0.05	.96
Topic-Specific x Grade	0.00	0.01	-0.07	.94
Domain x Topic-Specific x Grade	0.02	0.01	3.00	.00*

* $p < .05$

Appendix A

Sample GISA Interface with Text and Comprehension Items

GISA — Green Schools

Directions: Complete Diana's graphic organizer by dragging the correct choices into the boxes in the table. Not all of the choices will be used. Click on **View Article** to display the article. Click on the **Next** arrow when you are done.

Diana: There is a lot of information about the impact of green schools on the environment and learning, so I started a graphic organizer. Can you finish it for me?

Choices:
- Green schools are expensive to build.
- Reducing, Recycling and Reusing protect the environment
- Students experience real-world educational opportunities.
- Students share knowledge with families and members of the community.
- Research supports the idea that the setting plays an important role in learning.
- Renewable energy sources help reduce the amount of oil and natural gas used.

Impact on the Environment	Impact on Education
Green school builders select wood that comes from fast-growing trees to protect endangered forests.	
	Attendance rates for teachers and students are very high.
Energy-saving features are included in the design of green schools.	

View Article