

## Bridging Skill and Task-Oriented Reading

Karyn Higgs, Joseph P. Magliano, Eduardo Vidal-Abarca, Tomas Martínez & Danielle S. McNamara

**To cite this article:** Karyn Higgs, Joseph P. Magliano, Eduardo Vidal-Abarca, Tomas Martínez & Danielle S. McNamara (2015): Bridging Skill and Task-Oriented Reading, *Discourse Processes*, DOI: [10.1080/0163853X.2015.1100572](https://doi.org/10.1080/0163853X.2015.1100572)

**To link to this article:** <http://dx.doi.org/10.1080/0163853X.2015.1100572>



Accepted author version posted online: 22 Oct 2015.  
Published online: 22 Oct 2015.



Submit your article to this journal [↗](#)



Article views: 69



View related articles [↗](#)



View Crossmark data [↗](#)

# Bridging Skill and Task-Oriented Reading

Karyn Higgs and Joseph P. Magliano

*Department of Psychology  
Northern Illinois University*

Eduardo Vidal-Abarca and Tomas Martínez  
*Department of Educational & Developmental Psychology*

*University of Valencia, Valencia, Spain*

Danielle S. McNamara

*Department of Psychology  
Arizona State University*

Some individual difference factors are more strongly correlated with performance on postreading questions when the text is not available than when it is. The present study explores if similar interactions occur with bridging skill, which refers to a reader's propensity to establish connections between explicit text during reading. Undergraduates read science texts using two research tools. The Reading Strategy-Assessment Tool provided a measure of bridging skill. Texts and postreading questions were presented in Read&Answer, and the availability of the text while answering was manipulated. Contrary to prior research, bridging skill was comparably correlated with performance in both availability conditions. Although bridging skill was not correlated with search decisions, there was a trend toward a positive correlation with search time, suggesting that readers who tend to bridge more may also tend to persist longer in searching for answers. The results are discussed in terms of dynamic perspectives of task-oriented reading.

---

Correspondence concerning this article should be addressed to Karyn P. Higgs, Department of Psychology, Psychology-Computer Science Building, Northern Illinois University, Room 400, DeKalb, IL 60115, USA. E-mail: [khiggs@niu.edu](mailto:khiggs@niu.edu)

## INTRODUCTION

Reading in academic settings is typically grounded in a task or purpose, (i.e., task-oriented reading; Cerdán & Vidal-Abarca, 2008; Cerdán, Vidal-Abarca, Martínez, Gilabert, & Gil, 2009; Rouet, 2006; Rouet & Britt, 2011; Snow & the RAND Reading Study Group, 2002; Vidal-Abarca, Mañá, & Gil, 2010). Canonical theories of reading comprehension (e.g., Gernsbacher, 1990; Kintsch, 1988, 1998; Zwaan & Radvansky, 1998) explain mental model construction in terms of an interaction between the reader and a text but do not consider contextual factors (e.g., task/purpose, situational constraints) that also affect mental model construction (McNamara & Magliano, 2009b). Contemporary perspectives of academic literacy view reading comprehension as a dynamic interaction between the text, the reader, and the context (McCrudden & Schraw, 2007; McNamara & Magliano, 2009b; Rapp & van den Broek, 2005; Rouet, 2006; Snow & the RAND Reading Study Group, 2002). These dynamic perspectives have motivated a body of research to better understand how different characteristics of the reader and text interact with different task-oriented reading situations to influence comprehension.

One common task-oriented reading situation that students encounter is reading to answer questions. Postreading questions are of particular interest because they are frequently used to assess student learning and are often the basis for standardized tests of comprehension skill. One contextual factor that may affect question-answering performance is text availability. As one might expect, text availability has an impact on performance; readers perform better on questions when the text is available than when it is not (Ozuru, Best, Bell, Witherspoon, & McNamara, 2007).

Although many individual difference factors correlate with comprehension question performance (e.g., Hannon & Daneman, 2001; Kendeou, van den Broek, Helder, & Karlsson, 2014; McNamara & Kintsch, 1996), some studies have shown asymmetrical relationships between individual difference factors and performance as a function of text availability. Specifically, these studies found stronger positive correlations between individual difference measures (prior knowledge, fluency, etc.) and performance when the text was unavailable than when it was available (Ozuru et al., 2007; Schaffner & Schiefele, 2013; Schroeder, 2011). Given that the availability of texts varies across situations in which students are assessed for comprehension abilities (e.g., take-home tests, closed-book quizzes, high-stakes standardized tests), further investigation of interactions between individual difference factors and text availability is warranted.

The present study examined the impact of *bridging skill* on comprehension depending on text availability. Bridging skill refers to readers' propensity to establish relationships between discourse constituents (e.g., inferring causal relationships between sentences) during reading (Magliano & Millis, 2003; Magliano, Millis, The RSAT Development Team, Levinstein, & Boonthum, 2011). Unlike standardized measures of general comprehension skill, bridging

skill assesses individual differences in a process that theories of comprehension specify as critical for successful comprehension (e.g., Graesser, Singer, & Trabasso, 1994; McNamara & Magliano, 2009a, 2009b). Will asymmetric effects be seen between bridging skill and performance as a function of text availability or does this skill reflect a fundamental aspect of comprehension that equally predicts performance independent of text availability? The present study was conducted to answer this question.

### Dynamic Perspectives of Reading and Text Availability

One model of task-oriented reading, the Task Relevancy and Content Extraction (TRACE) model (Rouet, 2006; Rouet & Britt, 2011), outlines the processes involved in tasks such as question answering and provides a background for understanding why text availability might affect the relationship between individual difference factors and performance. TRACE assumes that task-oriented reading involves constructing multiple representations, including a task model, a text representation, and a response model. A task model is initially constructed before reading and guides processing decisions and the construction of subsequent representations (see also McCrudden & Schraw, 2007) and is dynamically updated during all phases of processing. Assessing the availability of information sources is an important part of constructing a task model and guiding processing decisions. These include internal resources (e.g., memory representations) and external resources (e.g., the text). In the case of postreading questions, readers who are aware that the text will be unavailable may choose to devote more resources during reading because they know they must rely on their memory for the text when answering. In contrast, readers who are aware that the text will be available may choose to devote fewer resources to initial reading because they can rely on searching the text for answers. For example, skim and search strategies have been observed in the context of standardized tests when the text was available (Farr, Pritchard, & Smitten, 1990; Salmerón et al., *in press*).

The different processing strategies afforded by these two task situations may affect the importance of text and reader characteristics for performance. In both situations question prompts serve as retrieval cues for relevant information in the reader's mental representation (Graesser & Clark, 1985; Graesser & Franklin, 1990). When the text is unavailable, the quality of the reader's representation will directly impact question-answering performance. When the text is available, the quality of the representation may not be as important because readers can rely on searching the text. Thus, reader and text factors that affect the quality of the mental representation may be more important when the text is unavailable.

However, factors that affect the quality of the mental representation may still influence performance even when readers can search for answers. TRACE assumes complex relationships between various aspects of task performance and individual

differences (Rouet, 2006), which is borne out in the literature (Cerdán, Gilabert, & Vidal-Abarca, 2011; Mañá, Vidal-Abarca, Dominguez, Gil, & Cerdán, 2009; Vidal-Abarca et al., 2010). When readers answer questions with the text available, a key distinction must be made between factors closely associated with reading comprehension skills and those relatively independent of these skills. During search, the ability to determine what information is relevant to answer a question is correlated with general comprehension skills (Cerdán et al., 2011; Vidal-Abarca et al., 2010). However, Vidal-Abarca and colleagues have shown that the decision to search requires metacognitive processes. This involves an evaluation of the need to search, which, surprisingly, appears to be uncorrelated with general comprehension skills (Cerdán et al., 2011; Vidal-Abarca et al., 2010). Similarly, other researchers have found that verbal skills are independent of comprehension monitoring accuracy (Maki, 1998; Maki, Shields, Wheeler, & Zachilli, 2005). Moreover, Mañá et al. (2009) have found weak correlations between comprehension performance and search decisions, indicating that both skilled and less-skilled readers are equally *inaccurate* in determining when it is important to search. Thus, some aspects of search behavior may be associated with individual differences in comprehension processes, whereas other aspects may be correlated with factors that are not the target of the present study (e.g., metacognitive skills).

### Interaction Between Bridging Skills and Text Availability

The present study focuses on *bridging skill*, an individual factor closely associated with comprehension skill. Bridging skill refers to one's propensity to construct bridging inferences that establish relationships between explicit discourse constituents. These inferences play a critical role in constructing a coherent and durable representation of text (e.g., Cain, Oakhill, Barnes, & Bryant, 2001; Graesser et al., 1994; Hannon & Daneman, 2001; Magliano & Millis, 2003; Magliano et al., 2011; McNamara & Magliano, 2009a, 2009b), and the extent to which readers establish these relationships is consistently correlated with comprehension and memory for text (e.g., Magliano & Millis, 2003; Magliano et al., 2011). However, the extent to which readers generate these inferences varies both between individuals and within individuals across different reading situations (Cain et al., 2001; Rapp & van den Broek, 2005; van den Broek, Fletcher, & Risdén, 1993; van den Broek, Lorch, Linderholm, & Gustafson, 2001).

Given the importance of bridging inferences for constructing coherent mental models, it is reasonable to predict that bridging skill would be positively correlated with performance when the text is *unavailable*. Discourse constituents in a mental model are more accessible to the extent that they are elaborated (e.g., Myers & O'Brien, 1998), and bridging inferences elaborate on relationships between discourse constituents (McNamara & McDaniel, 2004). The more one generates bridging inferences, the higher the likelihood that the knowledge

necessary to answer questions will be available in the mental model and accessible upon reading question prompts.

However, bridging skill may also be positively correlated with performance when the text is available, even though readers can search the text. First, the quality of the memory representation should affect the likelihood that relevant text content is activated when reading the question prompt (Graesser & Clark, 1985), regardless of text availability. Further, bridging skill may influence how effectively readers search when they decide to do so. A reader's mental representation of a text can provide information about where content is located in the texts (e.g., Rothkopf, 1971). The more coherent the mental model one constructs, the stronger the memory for where content is located (e.g., Cataldo & Oakhill, 2000). Therefore, students who tend to bridge more may be better at locating relevant answer information when searching than those who bridge less, similar to the skilled and less skilled comprehenders in Cerdán et al. (2011) and Vidal-Abarca et al. (2010). Finally, readers who tend to bridge more may conduct a more complete and systematic search, particularly for questions requiring integration. First, they may have a better understanding that complex questions may require answers from different parts of the text (e.g., bridging inference questions). Moreover, a more coherent and enriched representation may lead high bridgers to persist longer in searching than low bridgers, because they would be more likely to perceive that the effort to search would yield a positive outcome. It has long been argued that much of human behavior is driven by the perceived probability of success or the expected utility of one's actions (e.g., Von Neumann & Morgenstern, 1953; Van Eerde & Theiry, 1996).

Another factor that affects question-answering performance is question difficulty (Cerdán & Vidal-Abarca, 2008; Cerdán et al., 2009; Martínez, Vidal-Abarca, Gil, & Gilabert, 2009; McNamara & Kintsch, 1996; Rouet, Vidal-Abarca, Bert-Erboul, & Millogo, 2001). Questions can vary in difficulty based on the extent that they require the reader to locate, integrate, or reason about the discourse constituents (Organisation for Economic Cooperation and Development, 2001). It is reasonable to hypothesize that bridging skill influences the relationship between performance and the difficulty of answering questions that require integration. The more readers generate bridging inferences during reading, the greater the likelihood they will establish the relationships between text constituents required to answer integration questions (McNamara, O'Reilly, Best, & Ozuru, 2006). Therefore, although performance should decline as integration questions increase in difficulty, that decline should be less steep for high bridgers than for low bridgers. However, it is unclear whether this interaction also depends on text availability.

Given the literature reviewed thus far, there are two possible patterns of relationship between bridging skill and performance on comprehension questions as a function of availability. First, there could be asymmetrical relationships similar to those found in prior research (Ozuru et al., 2007; Schaffner & Schiefele, 2013; Schroeder, 2011). Specifically, there may be a stronger correlation when the

text is not available than when it is available. When a text is available readers can rely on different processes (i.e., searching) to answer questions. However, when the text is unavailable, readers must rely entirely on the mental representation they construct during initial reading. On the other hand, bridging skill could be equally correlated with performance across availability conditions due to its importance for constructing coherent mental models.

Regarding question difficulty, question–answering accuracy was expected to decline as a function of two difficulty factors: (1) the number of sentences needed to answer a question and (2) the distance between the sentences. Moreover, bridging skill was also predicted to interact with question difficulty, such that a steeper negative slope was expected for low bridgers than for high bridgers as difficulty increased.

Finally, we assessed how bridging influenced two dimensions of search behavior when the text was available, the decision to search and search time. Regarding search decisions, it was predicted that participants would decide to search more often as question difficulty increased and that there would be no difference between high and low bridgers due to the metacognitive character of this decision and its independence from comprehension and verbal skills (Cerdán et al., 2011; Maki, 1998; Maki et al., 2005; Vidal-Abarca et al., 2010). Regarding search time, one potential pattern is that high bridgers might spend less time than low bridgers because their search would be facilitated by a more coherent representation. An interesting alternative is that high bridgers might spend *more* time searching than low bridgers.

## METHODS

### Participants

Participants were 117 students at a large U.S. Midwestern university who received course credit for an introductory psychology class; 94 participants (50 men, 44 women) were used in the analyses. Twelve participants were not included in the analyses because data files were missing from one of the two computer-based tools (e.g., computer crashed or data file was not created). Additionally, six participants were excluded because of noncompliance with instructions (e.g., did not read one or more texts or answered no questions for at least one text), and five participants were excluded because they received incorrect texts according to the counterbalancing scheme.

### Materials and Assessments

*Texts and questions.* Participants read eight science texts adapted from those used by Magliano et al. (2011). Information about the texts can be viewed

TABLE 1  
Description of the Texts

<i>Title</i>	<i>No. of Words</i>	<i>No. of Sentences</i>	<i>No. of Paragraphs</i>	<i>FKGL</i>
Thunderstorm Development	251	20	4	8.6
Understanding Life	493	35	5	8.2
How Does Lightning Happen	348	27	4	6.1
How Cancer Develops	359	27	5	9.6
The Power of Erosion	312	21	5	9.9
How Radar Works	285	24	4	9.3
West Nile Virus	356	21	5	9.5
The Pancreas and Diabetes	411	24	6	11.1

FKGL, Flesh-Kincaid grade level.

in Table 1. Participants read four texts in the Reading Strategy-Assessment Tool (RSAT) and four in Read&Answer. Texts were grouped in four pairs, and the order of presentation was counterbalanced across the two assessment tools. Text availability was also counterbalanced so that each participant read a pair of texts in each availability condition (available or unavailable) in both assessment tools.

There were four comprehension questions for each text. Each participant answered 16 questions related to the four texts they read in Read&Answer. The questions were adapted from those developed by Magliano et al. (2011) for presentation in RSAT during reading and were reworded when necessary for presentation after reading the texts. The questions were originally constructed to require integration of information from multiple sentences in the text. However, it was realized after the study was conducted that one of the reworded questions could be answered based on information in a single sentence. The mean number of sentences needed to answer questions was 3.63. An example text with associated questions is shown in Appendix A.

Participants' answers were scored by human judges using a scoring scheme derived from ideal answers developed by Magliano et al. (2011). Each ideal answer was parsed into idea units with a mean of 3.13 idea units per question. Scores represented the proportion of idea units produced in the answer. Scoring reliability was assessed by randomly selecting a minimum of 20% of responses for each question. The intraclass correlation coefficient for absolute agreement between the two judges was high (.90). The remainder of the responses was scored by a single judge.

Question difficulty was represented by two factors, the number of text sentences containing idea units relevant to answering the question and the distance between these relevant sentences in the text. Distance was intended to capture differences in difficulty between questions that required bridging between adjacent sentences and bridging between sentences where any number of nonrelevant sentences could intervene. Distance was obtained by counting the



number of intervening sentences between relevant sentences and adding a value of 1. Adjacent relevant sentences received a distance value of 1, and relevant sentences with one intervening sentence received a value of 2 and so on. For the question where there was only one relevant sentence, the distance value was zero. The average of the distances between text sentences needed to answer a given question was used in the analysis, with a mean distance value of 1.10. It is important to note that one question was dropped because virtually no participants produced one part of the ideal answer.<sup>1</sup>

### Computer-Based Assessment Tools

*RSAT.* The measure of bridging skill was obtained using RSAT (Magliano et al., 2011). RSAT derives this measure by having participants produce open-ended verbal protocols using a variant of think-aloud instructions. Think-aloud measures have been shown to be sensitive to individual differences in the propensity to engage in bridging inferences (Magliano & Millis, 2003; Millis, Magliano, & Todaro, 2006; Rapp, van den Broek, McMaster, Kendeou, & Espin, 2007; van den Broek et al., 2001; Trabasso & Magliano, 1996).

In RSAT text sentences are presented incrementally. Participants advanced to the next sentence by pressing a button marked “next.” After target sentences participants saw the prompt, “What are you thinking now?,” appear on the screen and typed their responses into a text box beneath the prompt. Two types of sentence presentation format were used in the current study to implement the two text availability conditions. In the single-sentence presentation format, participants could see only the current sentence, and transitions to new paragraphs were marked by the phrase, “NEW PARAGRAPH.” In the “whole text” condition, all prior sentences were visible and new sentences were added following a paragraph structure. Although prior research has shown no differences in bridging scores as a function of these presentation formats (Gilliam, Magliano, Millis, Levinstein, & Boonthum, 2007), we had participants read under both presentation conditions to ensure that the availability of prior text did not affect bridging scores.

RSAT computes bridging scores using an automated scoring procedure that counts the number of content words from the prior text that are produced in the participant’s response (for a detailed description of the algorithms see Magliano et al., 2011). RSAT process measures have been shown to have respectable validity and reliability. The bridging score is highly correlated with human judgments of bridging (Pearson  $r$  ranging from .64 to .70; Magliano et al., 2011). Additionally, correlations between test–retest bridging scores for different forms

---

<sup>1</sup>For the question that was dropped from the analyses, only 5% of the participants produced one of the four idea units that constituted the answer in the scoring scheme.

of RSAT are also respectable given the open-ended nature of the task of “thinking aloud” (Pearson  $r = .79$ ; Millis & Magliano, 2012).

In the current study participants read four texts in RSAT and produced responses at 16 sentences (4 per text). These prompt locations were selected based on sentences identified by Magliano et al. (2011) that afforded bridging between the target sentence and prior discourse (for an example text with prompt locations see Appendix A). Participants showed reasonably high consistency in bridging scores across locations. For the four counterbalance groups, intraclass correlation coefficients ranged from .76 to .90.

*Read&Answer.* Measures of reading and question–answering behavior were obtained using a computer-based research tool, Read&Answer (Vidal-Abarca et al., 2011). Using a masking technique, Read&Answer collects a log of virtually every action taken by a user, akin to the type of continuous behavioral data collected by an eye tracker. Texts are presented in masked segments defined by the researcher (for screen shots of the text presentation and question windows see Appendix B). The layout of a text is visible to the reader, but the content is masked until a segment is revealed by clicking on it. When the reader clicks on a new segment, the previously viewed segment is masked again. Read&Answer captures reading time (in milliseconds) for each text segment exposure and the sequence in which segments are accessed, similar to the moving window method (e.g., Haberlandt & Graesser, 1985).

To access the question screen readers click an icon on the text presentation screen. Readers click to reveal masked question prompts and type answers into a text box below the prompt. Read&Answer allows the researcher to specify whether or not participants can return to the text screen once they have accessed the questions. We used this functionality to implement the text availability manipulation. When the text was available, readers could click on an icon to return to the text screen to search. Read&Answer captured these clicks for use in the search decision analysis.

Text segments were defined so participants would be required to click on multiple segments when searching for the answers to questions (for an example text showing segmentation see Appendix A). However, for 6 of the 32 comprehension questions it was deemed detrimental to the flow of the text to break the idea units into separate segments. The average number of segments per text was 13.63 (range, 11–19), with an average number of sentences per segment of 1.81 (range, 1–4). Exposure times on these segments during search were aggregated to obtain a measure of search time for each question.

*Background knowledge.* The background knowledge assessment consisted of 20 four-alternative multiple-choice questions that targeted general science knowledge, including questions related to biology, chemistry, earth science, math, and research methods. This assessment was originally developed

by analyzing the first two chapters of four high school science books to identify inferences needed to understand the text. Variations of the test have been validated with over 4,000 high school and college students (McNamara et al., 2006; O'Reilly, Best, & McNamara, 2004; O'Reilly & McNamara, 2007) and showing high reliability and validity, with split-half reliability ranging between .72 and .81, correlations with reading comprehension between .37 and .68, and correlations with science comprehension between .30 and .60.

## Procedure

Sessions were conducted with groups of up to six participants and took approximately 1.5 to 2 hours to complete. Participants were randomly assigned to a counterbalancing condition, determining the assignment of the texts and availability conditions across the two assessment tools.

In phase 1 participants were told they would read texts presented one sentence at a time in RSAT and would occasionally see the question "What are you thinking now?" appear on the screen. Participants were instructed to respond to the prompt by reporting whatever thoughts immediately came to mind about the sentence they had just read and how it related to the text (Trabasso & Magliano, 1996). Participants then practiced responding to the prompt using a paper and pencil packet. The experimenter reviewed the practice packets and in some circumstances provided feedback to encourage more substantive responses. Specifically, feedback was provided when participant responses contained no semantic content related to the reading, such as vague responses (e.g., "ok," or "I don't know") or responses that involved only simple metacognitive statements (e.g., "makes sense"). Feedback consisted of telling participants, "We are interested in your thoughts about the text. In your responses to the prompt, please tell us more about your understanding about what you are reading." Participants then read the texts in RSAT. Two texts were read with all prior sentences visible, and two texts were read with only the current sentence visible. Before reading each text pair, participants were informed whether or not they would be able to see prior sentences as they read.

Instructions for phase 2 were provided after all participants had completed phase 1 and were offered the opportunity to take a short break. In phase 2 participants read four texts using Read&Answer and answered four comprehension questions for each text. Two texts were read with the text available when answering questions and two were read with the text unavailable. Participants were first guided through practice texts for both availability conditions to familiarize them with unmasking text sentences and navigating through the text and question screens. Next, participants read the two pairs of experimental texts and before reading were informed of whether or not they would be able to access the text while answering questions.

After phase 2 was completed, participants completed the science background knowledge assessment. This was presented after the texts to ensure participants were not differentially primed as a function of individual differences in background knowledge (McNamara & Kintsch, 1996; O'Reilly & McNamara, 2007). None of the questions on the background knowledge assessment could be answered based on the text content.

## Data Analysis

Hierarchical linear modeling was used to conduct analyses on the comprehension and search outcomes. The primary goal of the study was to assess how bridging skill influenced comprehension and search behavior under different task conditions. A multilevel modeling approach was appropriate because question and task-related factors varied with each observation of the outcome variable (i.e., at the item level), whereas reader characteristics applied across all observations for an individual (i.e., the person level). With nested data structures such as this, a multilevel modeling approach is needed to account for variability at the different hierarchical levels as well as dependencies between the levels (Bryk & Raudenbush, 1992).

Three hierarchical linear modeling (HLM) models were analyzed: one predicting comprehension scores, one predicting search decisions, and one predicting search times when participants made the decision to search. The outcome measures and the item level predictors that varied with these observations (e.g., question difficulty and text availability) were entered at Level 1. These observations were nested within Level 2 groups (i.e., persons). At Level 2 variables associated with the individual (e.g., bridging scores) were entered into the model. In HLM, Level 2 variables can be entered both as a predictor of the outcome (i.e., main effect) and as a cross-level predictor of variation in the relationship between the outcome and a Level 1 variable (i.e., slopes).

*Model variables.* There were 16 observations for the comprehension score outcome with scores ranging from .00 to 1.0. The search decision outcome reflects whether or not a participant opted to search the text when answering the eight questions in the text available condition. The search time outcome refers to the amount of time spent searching when the participant chose to search. Because participants did not always choose to search, the number of observations of search time varied between individuals.

The three models contained both common and unique predictor variables. All predictor variables were grand mean centered with the exception of text availability, which was contrast coded (Unavailable = -1, Available = 1). The predictor variables common to all models included the question difficulty variables (Relevant Sentences and Distance) at the item level (Level 1) and

Bridging and Background Knowledge at the person level (Level 2). For Bridging both the main effect and cross-level interactions with each Level 1 predictor were assessed. Background Knowledge was included as a control variable, and cross-level interactions were not assessed.

Each model also contained unique predictor variables. At Level 1 the comprehension model included text availability and interactions between availability and the question difficulty variables (i.e., Available  $\times$  Relevant Sentences, Available  $\times$  Distance). These variables were not included in the two search outcome models because searching was only permitted in the available condition. The model predicting search time also included the number of questions for which participants made the decision to search. This variable was entered at Level 2, but cross-level interactions were not assessed. See Appendix C for the model equations.

## RESULTS

Means and standard deviations (SDs) for the comprehension and search decision models are shown in Table 2 and for the search time model in Table 3. Preliminary analyses were conducted on the RSAT bridging scores and Read&Answer reading time data to assess differences as a function of text availability. The first analysis confirmed that there was no significant difference in bridging scores between the whole text presentation condition ( $M = 3.23$ ,  $SD = 2.37$ ) and the single sentence presentation condition ( $M = 3.01$ ,  $SD = 1.91$ );  $t(92) = 1.11$ ,  $p = .27$ . Thus, we computed an aggregated bridging score for each individual by averaging over the two presentation conditions.

Second, we assessed whether text availability affected time spent reading before viewing the questions in Read&Answer, which would be indicative of

TABLE 2  
Means and SDs for Variables in Comprehension and Search Decision Models

	<i>n</i>	<i>Min</i>	<i>Max</i>	<i>M</i>	<i>SD</i>
Comprehension	1,457	.00	1.00	.39	.32
Search decisions	727	.00	1.00	.62	.49
Text availability	1,457	- 1.00	1.00	.00	1.00
Relevant sentences	1,457	1.00	6.00	3.65	1.21
Distance	1,457	.00	2.00	1.04	.29
Bridging	94	.30	10.60	3.13	1.92
Background knowledge	94	7.00	20.00	15.73	3.00

The Search Decision model included only items in the available condition ( $n = 727$ ). The means for the relevant sentences and distance predictors in that model are comparable with the comprehension model and therefore are not reported here.

TABLE 3  
Means and SD for Variables in Search Time Model

	<i>n</i>	<i>Min</i>	<i>Max</i>	<i>M</i>	<i>SD</i>
Search time	451	2.00	135.00	26.22	20.03
Relevant sentences	451	1.00	6.00	3.69	1.12
Distance	451	0.00	2.00	1.06	0.26
Bridging	451	0.30	10.60	3.16	1.89
Background knowledge	85	7.00	20.00	15.68	2.96
Number of Searches	85	1.00	8.00	5.31	2.20

Model includes only items for which participants searched while answering a question.

readers adopting different reading strategies as a function of knowing whether or not they could search the text when answering. Readers spent significantly more time reading (in seconds) before viewing the questions in the unavailable condition ( $M = 130.41, SD = 36.09$ ) than in available condition ( $M = 89.29, SD = 50.71$ ),  $t(93) = 7.45, p < .001$ . Thus, readers adjusted their reading strategies in response to the different task demands in the two availability conditions.

### Model Predicting Comprehension Outcomes

The outcome measures for this analysis were scores from 16 comprehension questions from four texts read under two different task conditions (text available/not available). Intraclass correlations based on an unconditional model (i.e., with no predictors) revealed that the proportion of variance between individuals (i.e., Level 2 units) was .16.

Model results related to Level 1 item variables are reported first, followed by the Level 2 reader factors and cross-level interactions. Additionally, the proportion of variance explained by the addition of predictors at each level was obtained using methods recommended by Snijders and Bosker (1994, 2011). These formulas provided pseudo- $R^2$  statistics measuring the proportional reduction of error in prediction at each level based on the residual variances from a fitted model (including the variables of interest) and a baseline model (without these variables). Results of the full HLM model are provided in Table 4.

*Level 1 question and task variables.* As expected, text availability, number of relevant sentences, and distance all significantly influenced comprehension scores. Comprehension performance was significantly better when the text was available ( $M = .43, SD = .31$ ) than when the text was unavailable ( $M = .35, SD = .32$ ) (Text Availability  $\beta_{10} = .04, p < .001$ ; see Table 4). The number of relevant sentences was negatively correlated with comprehension score (Sentences  $\beta_{20} = -.02, p = .001$ ); as the number of

TABLE 4  
HLM Results for Full Model Predicting Comprehension Score Outcomes

<i>Fixed Effects</i>	$\beta$	<i>SE</i>	<i>t</i> -ratio	<i>p</i>
Intercept1 ( $\pi_0$ )				
Intercept2 ( $\beta_{00}$ )	.39	.01	31.62	<.001
Bridging ( $\beta_{01}$ )	.02	.01	2.36	.020
Background knowledge ( $\beta_{02}$ )	.02	.00	5.44	<.001
Text availability slope ( $\pi_1$ )				
Intercept2 ( $\beta_{10}$ )	.04	.01	5.29	<.001
Bridging ( $\beta_{11}$ )	.00	.00	0.44	.658
Relevant sentences slope ( $\pi_2$ )				
Intercept2 ( $\beta_{20}$ )	-.02	.01	-3.33	.001
Bridging ( $\beta_{21}$ )	.01	.00	2.05	.044
Distance slope ( $\pi_3$ )				
Intercept2 ( $\beta_{30}$ )	-.11	.02	-5.09	<.001
Bridging ( $\beta_{31}$ )	-.01	.01	-1.14	.258
Available $\times$ relevant sentences slope ( $\pi_4$ )				
Intercept2 ( $\beta_{40}$ )	-.01	.01	-1.49	.139
Bridging ( $\beta_{41}$ )	.00	.01	.52	.608
Available $\times$ distance slope ( $\pi_5$ )				
Intercept2 ( $\beta_{50}$ )	-.00	.02	-.07	.941
Bridging ( $\beta_{51}$ )	-.01	.01	-.70	.487
Random Effects	Variance	<i>df</i>	$\chi^2$	<i>p</i>
Intercept1 ( $r_0$ )	.01	91	202.42	<.001
Relevant sentences slope ( $r_2$ )	.00	92	80.03	>.500
Distance slope ( $r_3$ )	.00	92	60.80	>.500
Available $\times$ relevant sentences Slope ( $r_4$ )	.00	92	101.48	>.500
Available $\times$ distance slope ( $r_5$ )	.00	92	70.59	.234
Level 1, <i>e</i>	.08			

sentences relevant to answering a question increased, comprehension performance decreased. Increases in distance between relevant sentences were also associated with decreases in comprehension score, (Distance  $\beta_{30} = -.11$ ,  $p < .001$ ). The interactions between text availability and the relevant sentences and distance measures were not significant. The addition of Level 1 predictors reduced prediction error of comprehension by 4.19% at the item level and .00% at the participant level.

*Level 2 reader variables.* Bridging scores were positively correlated with comprehension score (Bridging  $\beta_{01} = .02$ ,  $p = .020$ ; see Table 4); high bridgers performed better on comprehension questions than did low bridgers. Background knowledge also had a significant positive relationship with comprehension score (Background Knowledge  $\beta_{02} = .02$ ,  $p < .001$ ). There was also a significant

cross-level interaction between bridging score and the number of relevant sentences (Bridging  $\beta_{21} = .01, p = .044$ ), which is shown in Figure 1. Although performance decreased for both high and low bridgers as relevant sentences increased, the slope for high bridgers was less steep, indicating they were less negatively affected by the number of relevant sentences. No other cross-level interactions involving bridging scores were significant. The addition of the Level 2 predictors reduced prediction error by 7.34 % at the item level and 32.57% at the participant level.

### Model Predicting Search Decisions

Search decision data for the eight questions in the text available condition were obtained from Read&Answer. Because the outcome was binary (Searched = 1, No Search = 0), a Bernoulli distribution using a logit link function was used. The intraclass correlation revealed that the proportion of variance between individuals was .46. The coefficients reported for Level 1 fixed effects regression represent changes in log odds and are interpreted in a manner similar to logistic regression.

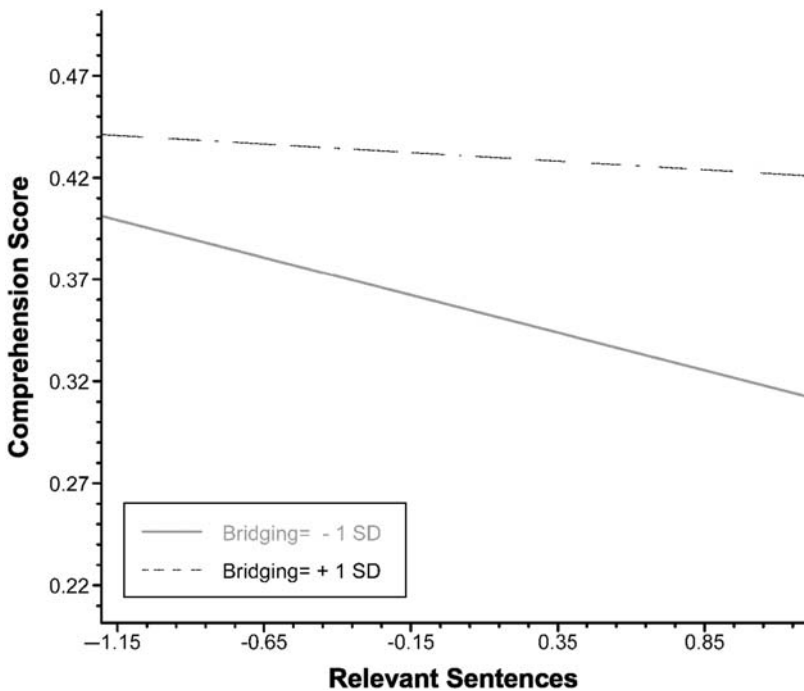


FIGURE 1 Significant cross-level interaction of bridging skill on the relationship between relevant sentences and comprehension score.



TABLE 5  
HLM Results for Full Model Predicting Search Decision Outcomes

<i>Fixed Effects</i>	<i>B</i>	<i>Search Decision</i>		
		<i>SE</i>	<i>t-ratio</i>	<i>p</i>
Intercept1 ( $\pi_0$ )				
Intercept2 ( $\beta_{00}$ )	.72	.21	3.50	<.001
Bridging ( $\beta_{02}$ )	.07	.11	.62	.537
Background knowledge ( $\beta_{03}$ )	-.06	.08	-.81	.422
Distance slope ( $\pi_1$ )				
Intercept2 ( $\beta_{10}$ )	.28	.39	.73	.469
Bridging ( $\beta_{11}$ )	-.09	.25	-.35	.726
Relevant sentences slope ( $\pi_2$ )				
Intercept2 ( $\beta_{20}$ )	.17	.08	2.10	.038
Bridging ( $\beta_{21}$ )	-.01	.03	-.40	.687
Random Effects	Variance	<i>df</i>	$\chi^2$	<i>p</i>
Intercept1 ( $r_0$ )	2.94	67	193.64	<.001
Distance slope ( $r_1$ )	.01	68	49.46	>.500
Relevant sentences slope ( $r_2$ )	.01	68	45.15	>.500

Chi-square statistics are based on only 70 of 94 units with sufficient data for computation. Fixed effects and variance components are based on all data.

To facilitate interpretation, odds ratios are provided when discussing significant Level 1 effects. Results of the full HLM model are provided in [Table 5](#).

With dichotomous outcome models it is difficult to obtain measures of variance explained related to the addition of predictors (Bauer, 2009; Snijders & Bosker, 2011). Thus, a single proportion of variance explained measure for the full model is provided using the method recommended by Snijders and Bosker (2011) for dichotomous outcome models. The total proportion of variance explained by the full search decision model was 27.04%.

*Level 1 question variables.* The number of sentences relevant to answering a question (Sentences  $\beta_{20} = .17, p = .038$ ) had a significant positive relationship with search decisions; as the number of relevant sentences increased, search decisions increased. The odds ratio for relevant sentences was 1.18, indicating that for each one unit increase in relevant sentences there was an 18% increase in the odds of a search decision. Distance between relevant sentences (Distance  $\beta_{10}$ ) did not significantly influence search decisions.

*Level 2 reader variables.* There were no significant effects of Level 2 predictors or any significant cross-level interactions.

## Model Predicting Search Time

Search times were obtained from Read&Answer and represent total time spent searching the text (in seconds) while answering a specific question. Search time data are based only on questions for which participants decided to search. Data were log transformed to obtain a normal distribution. Because participants did not search for every question, the number of questions for which participants searched was included as a Level 2 predictor. Intraclass correlations, calculated using an unconditional model, revealed that the proportion of variance between individuals was .23. Results of the full HLM model are provided in Table 6.

*Level 1 question variables.* Neither the number of sentences relevant to answering a question (Sentences  $\beta_{10}$ ) nor the distance between relevant sentences (Distance  $\beta_{20}$ ) significantly affected search time. The addition of Level 1 predictors reduced prediction error by 1.82% at the item level and 3.47% at participant level.

TABLE 6  
HLM Results for Full Model Predicting Search Time Outcomes

<i>Fixed Effects</i>	$\beta$	<i>Search Time</i>		
		<i>SE</i>	<i>t-ratio</i>	<i>p</i>
Intercept1 ( $\pi_0$ )				
Intercept2 ( $\beta_{00}$ )	2.91	.05	59.66	< .001
Number Searches ( $\beta_{01}$ )	.06	.02	2.81	.006
Bridging ( $\beta_{02}$ )	.04	.02	1.91	.060
Background knowledge ( $\beta_{03}$ )	-.05	.02	-3.40	.001
Relevant sentences slope ( $\pi_1$ )				
Intercept2 ( $\beta_{10}$ )	.02	.03	.69	.494
Bridging ( $\beta_{11}$ )	-.03	.02	-1.42	.158
Distance slope ( $\pi_2$ )				
Intercept2 ( $\beta_{20}$ )	.04	.14	.26	.798
Bridging ( $\beta_{21}$ )	.07	.07	.98	.333
Random Effects	Variance	<i>df</i>	$\chi^2$	<i>p</i>
Intercept1 ( <i>r0</i> )	.08	36	66.92	.002
Relevant Sentences ( <i>r1</i> )	.01	38	53.65	.047
Distance ( <i>r2</i> )	.17	38	48.03	.128
Level 1, <i>e</i>	.39			

Chi-square statistics are based on only 40 of 85 units with sufficient data for computation. Fixed effects and variance components are based on all data.

*Level 2 reader variables.* Background knowledge had a significant negative relationship with search time (Background Knowledge  $\beta_{03} = -.05$ ,  $p = .001$ ); higher background knowledge was associated with less time spent searching the text when answering questions. Predictably, the number of questions for which readers made search decisions had a significant positive relationship with search time (Searches  $\beta_{01} = .06$ ,  $p = .006$ ). The main effect of bridging on search time showed a nonsignificant trend, suggesting that the more individuals bridged, the longer they spent searching the text (Bridging  $\beta_{02} = .04$ ,  $p = .060$ ). No cross-level interactions between bridging and the two question difficulty variables (Bridging  $\beta_{11}$ , Bridging  $\beta_{21}$ ) were significant. The Level 2 predictors reduced prediction error of search time by 8.43% at the item level and by 18.17% at the participant level.

## DISCUSSION

The goal of the present study was to explore relationships between bridging skill, question difficulty, and text availability when answering comprehension questions. Consistent with prior research, readers spent significantly more time reading the text before viewing the questions in the unavailable condition than in the available condition (Ferrer, Vidal-Abarca, Avila, Mañá, & Llorens, 2010), and question-answering performance was better when the text was available than when it was unavailable (e.g., Ozuru et al., 2007). The most important findings of the present study pertain to the relationship between bridging skill and question performance and search behaviors. Contrary to prior research showing asymmetrical relationships between individual factors and performance as a function of text availability (Ozuru et al., 2007; Schaffner & Schiefele, 2013; Schroeder, 2011), bridging skill was equally correlated with performance in both availability conditions. Moreover, the negative correlation between relevant sentences and performance was qualified by an interaction with bridging skill (Figure 1). Low bridgers were more negatively affected by the number of relevant sentences than high bridgers. As expected, the decision to search was not correlated with bridging skill (Cerdán et al. 2011; Mañá et al., 2009; Vidal-Abarca et al., 2010). However, when the text was available, there was a nonsignificant trend suggesting that high bridgers spent *more* time searching for answers than low bridgers.

How do we explain these effects and what are their implications for theory? First, consider the most provocative finding that bridging skill was equally correlated with performance regardless of text availability. There are at least two possible explanations, neither of which are mutually exclusive. The first is that bridging skill reflects a fundamental aspect of comprehension skill (Cain et al., 2001; Magliano & Millis, 2003; Magliano et al., 2011; Millis et al., 2006). The more one bridges, the more coherent and elaborated the representation, which

should affect the accessibility of relevant text content after reading a question prompt, regardless of whether the text is available to search (see also Cerdán et al., 2011; Vidal-Abarca et al., 2010).

A second explanation is that the nature of the comprehension questions may have reinforced the role played by bridging skills in the available condition. Successfully answering the open-ended questions in this study was difficult and required generating bridging inferences to connect the ideas across the text. The difficulty of answering these integration questions is evidenced by the mean score of 39% and the high likelihood of searching the text (62% of the time). Under these conditions bridging skills play a determinant role for performance. It is an open question whether bridging skill would be equally correlated across availability conditions for questions that require other processes, such as locating a key concept or a complex evaluation (Organisation for Economic Cooperation and Development, 2001). Nonetheless, bridging skill has been found to be positively correlated with performance for question formats that do not require integration (e.g., Magliano & Millis, 2003).

With respect to search behavior, the lack of correlation between bridging skill and search decisions is consistent with arguments that the decision to search is a metacognitive process that does not directly depend on reading comprehension skills (Maki, 1998; Maki et al., 2005; Mañá et al., 2009; Vidal-Abarca et al., 2010). Contrary to expectations that bridging skill would facilitate search (i.e., lead to faster search times), a provocative nonsignificant trend ( $p = .06$ ) suggested that high bridgers search longer than low bridgers for answers. These results are consistent with theories that assume that persistence in a behavior is correlated with the expectations of obtaining a positive outcome as a result (e.g., Van Eerde & Thierry, 1996). A more coherent representation may have led high bridgers to perceive that their efforts to search a text would lead to a more complete answer. Moreover, high bridgers likely had a better sense of what constituted complete answers and may have chosen to continue searching until they found the necessary information.

The results of the present study cannot be readily explained by standard theories of comprehension, which were never intended to account for interactions with contextual factors (see also McNamara & Magliano, 2009b; McNamara, Jacobina, & Allen, 2015). This study lends credence to the need for cognitively based theories of task-oriented reading, such as the TRACE framework (Rouet, 2006). We originally appealed to TRACE to provide an explanation for asymmetrical relationships between individual difference factors and comprehension performance as a function of text availability, which were not found in the present study. Nonetheless, some results of the study can be readily interpreted in terms of the framework. First, readers spent more time reading in the unavailable condition than the available condition, suggesting they developed task models that were sensitive to whether or not they could search for answers.

Importantly, TRACE assumes that individual difference factors can affect the development of the task, text, and product models. It is possible that the persistence in search was in part due to skilled bridgers having a more developed model of the task (i.e., what is required to answer the questions) than less skilled bridgers as well as a more coherent mental representation for the text. One important implication for TRACE is that the present results suggest that constructing a coherent text model is important for task performance, even when one has the opportunity to reread texts, particularly for tasks that require deep comprehension, such as answering open-ended why questions (e.g., Graesser & Franklin, 1990).

However, TRACE is best conceptualized as a theoretically motivated framework rather than a formal model of task-oriented reading. It specifies mental representations and processes that support task-oriented reading but is not a process model. As such, the framework can provide insights into the present data but did not provide a basis for generating principled processing assumptions that lead to hypotheses and predictions about the relations between the reader and task explored in the present study. A growing number of studies have explored the complex and dynamic interactions between reader, texts, and task (e.g., Bråten, Britt, Strømsø, & Rouet, 2011; Gil, Martinez, & Vidal-Abarca, 2015; Griffin, Wiley, & Thiede, 2008; Kaakinen, Hyona, & Keenan, 2003; Linderholm & van den Broek, 2002; Martinez et al., 2009; Ozuru et al., 2007) that similarly lack a formal theory to guide research questions, hypotheses, and predictions. The results of the present study underscore the need for such a theory, which ultimately should have the same level of specificity as formal theories of comprehension, such as the Construction-Integration (Kintsch, 1988, 1998) or Landscape (van den Broek, Rapp, & Kendeou, 2005; van den Broek, Young, Tzeng, & Linderholm, 1999) models.

Although bridging skill is a relatively new construct, a growing body of evidence suggests that computerized measures based on verbal protocols are correlated with measures of comprehension (e.g., Magliano & Millis, 2003; Magliano et al., 2011; Millis et al., 2006). However, it is important to note that RSAT provides a relatively gross measure of bridging based on the number of words the test-taker produces from the prior discourse context. RSAT does not differentiate whether test-takers are only restating explicit text content or generating a true bridging inference that establishes how the current sentence is semantically related (e.g., causal, logical) to the prior discourse context. However, correlations between RSAT scores and human judgments of bridging inferences are fairly robust ( $r = .64-.70$ ; Magliano et al., 2011). Certainly, improving the diagnostic features of RSAT to discriminate between coherence building and noncoherence building processes would improve the utility of the tool.

Finally, the results of this study also have important practical implications. It is a common practice to instruct students that they do not need to read a text to

perform well on standardized, high-stakes tests that make the text available concurrently with the test items (Rupp, Ferne, & Choi, 2006). This is particularly the case for standardized tests that adopt a multiple-choice format. Apparently, some practitioners assume it is not important to deeply comprehend the texts that are used in standardized texts and thereby recommend adopting a low standard of comprehension (i.e., skim the text or even skip it and go straight to the questions). The underlying assumptions are that (1) students will save time by not comprehending the text and therefore be able to devote time to reasoning about the questions, and (2) students will readily be able to locate question-relevant segments in the text based on semantic overlap with the questions and answers choices. The present results call into question the merits of this advice and suggest that further research with different types of test questions is warranted.

## FUNDING

This research was supported by funding from the Center for the Interdisciplinary Study of Language and Literacy at Northern Illinois University, The Institute for Educational Sciences in the U.S. Department of Education (grant no. R305F100007, R305A130124), the Spanish Ministry of Science and Innovation (grant EDU2011-27091), and Generalitat Valenciana (grant PROM-ETEO/2013/081). The opinions expressed are those of the authors and do not represent the views of the funding agencies.

## REFERENCES

- Bauer, D. J. (2009). A note on comparing the estimates of models for cluster-correlated or longitudinal data with binary or ordinal outcomes. *Psychometrika*, *74*, 97–105.
- Bråten, I., Britt, M. A., Strømsø, H. I., & Rouet, J. F. (2011). The role of epistemic beliefs in the comprehension of multiple expository texts: Toward an integrated model. *Educational Psychologist*, *46*, 48–70.
- Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.
- Cain, K., Oakhill, J. V., Barnes, M. A., & Bryant, P. E. (2001). Comprehension skill, inference-making ability, and their relation to knowledge. *Memory & Cognition*, *29*, 850–859.
- Cataldo, M. G., & Oakhill, J. (2000). Why are poor comprehenders inefficient searchers? An investigation into the effects of text representation and spatial memory on the ability to locate information in text. *Journal of Educational Psychology*, *92*, 791–799.
- Cerdán, R., Gilabert, R., & Vidal-Abarca, E. (2011). Selecting information to answer questions: Strategic individual differences when searching texts. *Learning and Individual Differences*, *21*, 201–205.
- Cerdán, R., & Vidal-Abarca, E. (2008). The effects of tasks on integrating information from multiple documents. *Journal of Educational Psychology*, *100*, 209–222.

- Cerdán, R., Vidal-Abarca, E., Martínez, T., Gilabert, R., & Gil, L. (2009). Impact of question-answering tasks on search processes and reading comprehension. *Learning and Instruction, 19*, 13–27.
- Farr, R., Pritchard, R., & Smitten, B. (1990). A description of what happens when an examinee takes a multiple-choice reading comprehension test. *Journal of Educational Measurement, 27*, 209–226.
- Ferrer, A., Vidal-Abarca, E., Ávila, V., Mañá, A., & Llorens, A. C. (2010). The role of question format and text availability on Reading Comprehension Assessment. Presented at the 1st joint conference of the Experimental Psychology Society (EPS) and the Spanish Experimental Psychological Society. Granada, Spain, April 15–17.
- Gernsbacher, M. A., Varner, K. R., & Faust, M. E. (1990). Investigating differences in general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 430–445.
- Gil, L., Martínez, T., & Vidal-Abarca, E. (2015). Online assessment of strategic reading literacy skills. *Computers & Education, 82*, 50–59.
- Gilliam, S., Magliano, J. P., Millis, K. K., Levinstein, I., & Boonthum, C. (2007). Assessing the format of the presentation of text in developing a Reading Strategy Assessment Tool (R-SAT). *Behavior Research Methods, 39*, 199–204.
- Graesser, A. C., & Clark, L. F. (1985). *Structures and procedures of implicit knowledge*. Norwood, NJ: Ablex.
- Graesser, A. C., & Franklin, S. P. (1990). QUEST: A cognitive model of question answering. *Discourse Processes, 13*, 279–303.
- Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101*, 371–395.
- Griffin, T. D., Wiley, J., & Thiede, K. W. (2008). Individual differences, rereading, and self-explanation: Concurrent processing and cue validity as constraints on metacomprehension accuracy. *Memory & Cognition, 36*, 93–103.
- Haberlandt, K. F., & Graesser, A. C. (1985). Component processes in text comprehension and some of their interactions. *Journal of Experimental Psychology: General, 114*, 357–374.
- Hannon, B., & Daneman, M. (2001). A new tool for measuring and understanding individual differences in the component processes of reading comprehension. *Journal of Educational Psychology, 93*, 103–128.
- Kaakinen, J. K., Hyönä, J., & Keenan, J. M. (2003). How prior knowledge, WMC, and relevance of information affect eye fixations in expository text. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 447–457.
- Kendeou, P., van den Broek, P., Helder, A., & Karlsson, A. K. (2014). A cognitive view of reading comprehension: Implications for reading difficulties. *Learning Disabilities Research & Practice, 29*, 10–16.
- 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.
- Linderholm, T., & van den Broek, P. (2002). The effects of reading purpose and working memory capacity on the processing of expository text. *Journal of Educational Psychology, 94*, 778–784.
- Magliano, J. P., & Millis, K. K. (2003). Assessing reading skill with a think-aloud procedure and latent semantic analysis. *Cognition and Instruction, 21*, 251–283.
- Magliano, J. P., Millis, K. K., the RSAT Development Team, Levinstein, I., & Boonthum, C. (2011). Assessing comprehension during reading with the Reading Strategy Assessment Tool (RSAT). *Metacognition and Learning, 6*, 131–154.
- Maki, R. H. (1998). Test prediction over text material. In D. J. Hacker, A. C. Graesser, & J. Dunlosky (Eds.), *Metacognition in educational theory and practice* (pp. 117–144). Mahwah, NJ: Erlbaum.

- Maki, R. H., Shields, M., Wheeler, A. E., & Zachilli, T. L. (2005). Individual differences in absolute and relative metacomprehension accuracy. *Journal of Educational Psychology, 97*, 723–731.
- Mañá, A., Vidal-Abarca, E., Domínguez, C., Gil, L., & Cerdán, R. (2009). Papel de los procesos metacognitivos en una tarea de pregunta-respuesta contextos escritos. *Infancia y Aprendizaje, 32*, 553–565.
- Martínez, T., Vidal-Abarca, E., Gil, L., & Gilabert, R. (2009). On-line assessment of comprehension processes. *Spanish Journal of Psychology, 12*, 308–319.
- McCrudden, M. T., & Schraw, G. (2007). Relevance and goal-focusing in text processing. *Educational Psychology Review, 19*, 113–139.
- McNamara, D. S., Jacovina, M. E., & Allen, L. K. (2015). Higher order thinking in comprehension. In P. Afflerbach (Ed.), *Handbook of individual differences in reading: Text and context* (pp. 164–176). New York, NY: Routledge.
- McNamara, D. S., & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes, 22*, 247–288.
- McNamara, D. S., & Magliano, J. P. (2009a). Self-explanation and metacognition: The dynamics of reading. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 60–82). New York, NY: Routledge.
- McNamara, D. S., & Magliano, J. P. (2009b). Toward a comprehensive model of comprehension. *Psychology of Learning and Motivation, 51*, 297–384.
- McNamara, D. S., & McDaniel, M. A. (2004). Suppressing irrelevant information: Knowledge activation or inhibition? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 465–482.
- McNamara, D. S., O'Reilly, T., Best, R., & Ozuru, Y. (2006). Improving adolescent students' reading comprehension with iSTART. *Journal of Educational Computing Research, 34*, 147–171.
- Millis, K. K., Magliano, J. P., & Todaro, S. (2006). Measuring discourse-level processes with verbal protocols and latent semantic analysis. *Scientific Studies of Reading, 10*, 225–240.
- Myers, J. L., & O'Brien, E. J. (1998). Accessing the discourse representation during reading. *Discourse Processes, 26*, 131–157.
- O'Reilly, T., Best, R., & McNamara, D. S. (2004). Self-explanation reading training: Effects for low-knowledge readers. In *Proceedings of the 26th annual meeting of the Cognitive Science Society* (pp. 1053–1058). Mahwah, NJ: Lawrence Erlbaum.
- O'Reilly, T., & McNamara, D. S. (2007). 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.
- Organisation for Economic Cooperation and Development. (2001). *Knowledge and skills for life: First results from PISA 2000*. Paris, France: OECD.
- Ozuru, Y., Best, R., Bell, C., Witherspoon, A., & McNamara, D. S. (2007). Influence of question format and text availability on the assessment of expository text comprehension. *Cognition and Instruction, 25*, 399–438.
- Rapp, D. N., & van den Broek, P. (2005). Dynamic text comprehension an integrative view of reading. *Current Directions in Psychological Science, 14*, 276–279.
- Rapp, D. N., van den Broek, P., McMaster, K. L., Kendeou, P., & Espin, C. A. (2007). Higher-order comprehension processes in struggling readers: A perspective for research and intervention. *Scientific Studies of Reading, 11*, 289–312.
- Rothkopf, E. Z. (1971). Incidental memory for location of information in text. *Journal of Verbal Learning and Verbal Behavior, 10*, 608–613.
- Rouet, J. F. (2006). *The skills of document use: From text comprehension to web-based learning*. Mahwah, NJ: Lawrence Erlbaum.



- Rouet, J. F., & Britt, M. A. (2011). Relevance processes in multiple document comprehension. In M. T. McCrudden, J. P. Magliano, & G. Schraw (Eds.), *Text relevance and learning from text* (pp. 19–52). Greenwich, CT: Information Age Publishing.
- Rouet, J.-F., Vidal-Abarca, E., Bert-Erboul, A., & Millogo, V. (2001). Effects of information search tasks on the comprehension of instructional text. *Discourse Processes, 31*, 163–186.
- Rupp, A. A., Ferne, T., & Choi, H. (2006). How assessing reading comprehension with multiple-choice questions shapes the construct: a cognitive processing perspective. *Language Testing, 23*, 441–474.
- Salmerón, L., Vidal-Abarca, E., Mañá, A., Martínez, T., Gil, L., & Naumann, J. (in press). Decisions in task-oriented reading. *The Spanish Journal of Psychology*.
- Schaffner, E., & Schiefele, U. (2013). The prediction of reading comprehension by cognitive and motivational factors: Does text accessibility during comprehension testing make a difference? *Learning and Individual Differences, 26*, 42–54.
- Schroeder, S. (2011). What readers have and do: Effects of students' verbal ability and reading time components on comprehension with and without text availability. *Journal of Educational Psychology, 103*, 877–896.
- Snijders, T. A., & Bosker, R. J. (1994). Modeled variance in two-level models. *Sociological Methods & Research, 22*, 342–363.
- Snijders, T. A., & Bosker, R. J. (2011). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London, UK: Sage.
- Snow, C., & the RAND Reading Study Group (2002). *Reading for understanding. Towards a R&D program for reading comprehension*. Santa Monica, CA: RAND.
- Trabasso, T., & Magliano, J. P. (1996). Conscious understanding during comprehension. *Discourse Processes, 21*, 255–287.
- van den Broek, P., Fletcher, C. R., & Risdien, K. (1993). Investigations of inferential processes in reading: A theoretical and methodological integration. *Discourse Processes, 16*, 169–180.
- 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*, 1081–1087.
- van den Broek, P., Rapp, D. N., & Kendeou, P. (2005). Integrating memory-based and constructionist processes in accounts of reading comprehension. *Discourse Processes, 39*, 299–316.
- van den Broek, P., Young, M., Tzeng, Y., & Linderholm, T. (1999). The landscape model of reading: Inferences and the online construction of a memory representation. In H. van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 71–98). Mahwah, NJ: Lawrence Erlbaum.
- Van Eerde, W., & Thierry, H. (1996). Vroom's expectancy models and work-related criteria: A meta-analysis. *Journal of applied psychology, 81*, 575–586.
- Vidal-Abarca, E., Mañá, A., & Gil, L. (2010). Individual differences for self-regulating task-oriented reading activities. *Journal of Educational Psychology, 102*, 817–826.
- Vidal-Abarca, E., Martínez, T., Salmerón, L., Cerdán, R., Gilabert, R., Gil, L., Mañá, A., Lloréns, A., & Ferris, R. (2011). Recording on-line processes in task-oriented reading with Read&Answer. *Behavior Research Methods, 43*, 179–192.
- Von Neumann, J., & Morgenstern, O. (1953). *Theory of Games and economic behavior*. Princeton, NJ: Princeton University Press.
- Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123*, 162–185.

## APPENDIX A

<i>R&amp;A Segment</i>	<i>RSAT Prompt</i>	<i>Paragraph</i>	<i>Sentence</i>	<i>Text</i>	<i>Relevant to Question</i>
1	-	0	0	West Nile virus	
2	-	1	1	The West Nile virus is an infectious organism that can cause fatal neurological disease in birds, horses, and humans and has been a serious health concern in the past several years.	Q1
2	-	1	2	There is no cure for the West Nile virus infection, and no vaccine has been developed to prevent infection in humans.	Q1
3	-	1	3	West Nile virus is named for a district in Uganda.	
3	-	1	4	The virus first emerged in North America in the late 1990s.	
4	-	1	5	It has quickly spread to over 40 states of the United States, north to Canada, and south to the Caribbean islands.	Q1
4	Yes	1	6	Thus, the West Nile virus has caused a great deal of alarm.	Q1
5	-	2	7	The virus is transmitted by the bite of an infected mosquito.	Q2, Q3
5	-	2	8	The virus multiplies in the body of the mosquito and eventually collects in the mosquito's salivary glands 7 to 14 days after infection.	Q2, Q3
6	-	2	9	After the mosquito bites a host, the genetic material within a cell is made to manufacture new copies of the virus.	Q2, Q3
6	Yes	2	10	The West Nile virus then invades cells throughout the target's body.	Q2, Q3
7	-	3	11	West Nile virus was initially only found in birds.	Q3
7	-	3	12	Mosquitoes become infected with the West Nile virus by feeding on the blood of a bird that carries the virus.	Q3
8	Yes	3	13	It was first identified in humans in 1937 in Uganda.	

*(Continued)*

(Continued)

<i>R&amp;A Segment</i>	<i>RSAT Prompt</i>	<i>Paragraph</i>	<i>Sentence</i>	<i>Text</i>	<i>Relevant to Question</i>
9	-	3	14	Not every human infected by the virus will experience symptoms, with about 1 out of 200 people infected with West Nile virus becoming ill.	
10	-	3	15	Symptoms of the virus range from mild, flu-like symptoms, to more serious symptoms.	Q4
11	-	4	16	In some, however, the virus reaches the brain.	Q4
11	-	4	17	Encephalitis can then develop, which can cause high fever, a stiff neck, disorientation, paralysis, convulsions, and coma.	Q4
12	Yes	4	18	This can cause the brain tissues to become inflamed and in about 10 percent of the cases, death.	Q4
13	-	5	19	Humans most at risk for severe illness include those over the age of 50 and those with a weakened immune system.	
13	-	5	20	Many states have initiated or improved community mosquito-control programs, using pesticides to reduce mosquito populations and draining standing water where mosquitoes breed.	
14	-	5	21	People can lower the risk of exposure to West Nile virus by remaining indoors from dusk to dawn.	

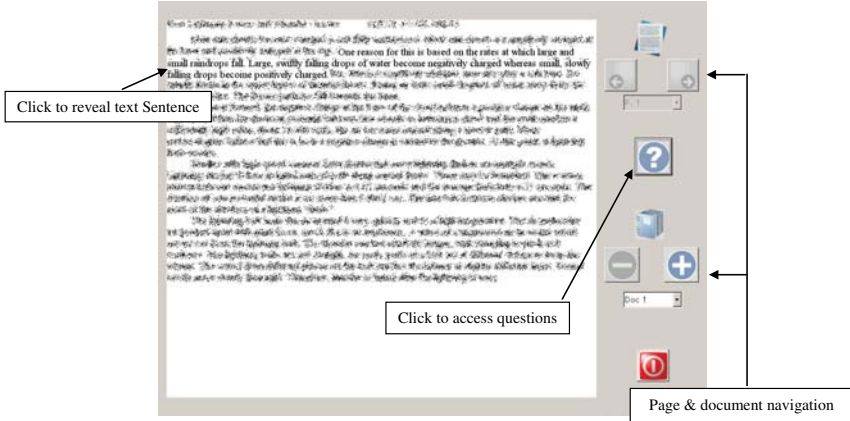
## Comprehension Questions for “West Nile Virus”

<i>Question</i>	<i>Comprehension Question Presented in Read&amp;Answer</i>	<i>No. of Relevant Sentences</i>	<i>Distance Average</i>
Q1	According to the text, why has West Nile virus caused great alarm?	4	1.667
Q2	How does the West Nile virus spread throughout the body?	4	1
Q3	How did the West Nile virus spread to humans?	6	1
Q4	What happens when the West Nile virus reaches the brain?	3	1

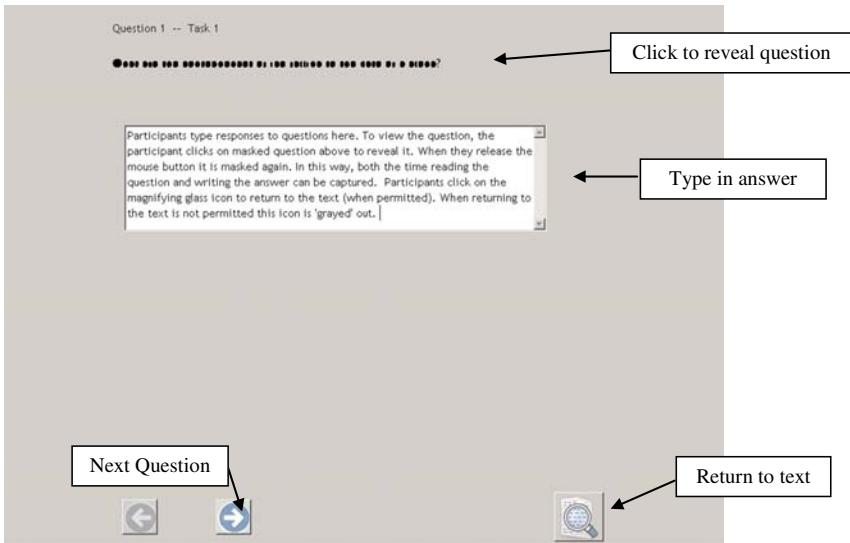
### APPENDIX B

## Example of Text and Question Answering Screens in Read&Answer

### Text Presentation Screen



### Question Screen



## APPENDIX C: HLM MODEL EQUATIONS

## Equation for the HLM Model Predicting Comprehension Score

Note: For all models, all predictors were grand mean centered before entry in model except Availability, which was contrast coded (Unavailable = -1, Available = 1).

*Level 1 Model*

$$\begin{aligned} \text{COMPREHENSIONSCORE}_{ti} = & \pi_{0i} + \pi_{1i} \times (\text{AVAILABILITY}_{ti}) + \pi_{2i} \times (\text{RELEVANT} \\ & \text{SENTENCES}_{ti}) + \pi_{3i} \times (\text{DISTANCE}_{ti}) + \pi_{4i} \times (\text{AVAILABILITY} \times \text{DISTANCE}_{ti}) \\ & + \pi_{5i} \times (\text{AVAILABILITY} \times \text{RELEVANTSENTENCES}_{ti}) + e_{ti} \end{aligned}$$

*Level 2 Model*

$$\pi_{0i} = \beta_{00} + \beta_{01} \times (\text{BRIDGE}_i) + \beta_{02} \times (\text{BCK\_KNOWLEDGE}_i) + r_{0i}$$

$$\pi_{1i} = \beta_{10} + \beta_{11} \times (\text{BRIDGE}_i)$$

$$\pi_{2i} = \beta_{20} + \beta_{21} \times (\text{BRIDGE}_i) + r_{2i}$$

$$\pi_{3i} = \beta_{30} + \beta_{31} \times (\text{BRIDGE}_i) + r_{3i}$$

$$\pi_{4i} = \beta_{40} + \beta_{41} \times (\text{BRIDGE}_i) + r_{4i}$$

$$\pi_{5i} = \beta_{50} + \beta_{51} \times (\text{BRIDGE}_i) + r_{5i}$$

*Mixed Model*

$$\begin{aligned} \text{COMPREHENSIONSCORE}_{ti} = & \beta_{00} + \beta_{01} \times \text{BRIDGE}_i + \beta_{02} \times \text{BCK\_KNOWLEDGE}_i \\ & + \beta_{10} \times \text{AVAILABILITY}_{ti} + \beta_{11} \times \text{BRIDGE}_i \times \text{AVAILABILITY}_{ti} \\ & + \beta_{20} \times \text{RELEVANT SENTENCES}_{ti} + \beta_{21} \times \text{BRIDGE}_i \times \text{RELEVANT SENTENCES}_{ti} \\ & + \beta_{30} \times \text{DISTANCE}_{ti} + \beta_{31} \times \text{BRIDGE}_i \times \text{DISTANCE}_{ti} \\ & + \beta_{40} \times \text{AVAILABILITY} \times \text{DISTANCE}_{ti} + \beta_{41} \times \text{BRIDGE}_i \times \text{AVAILABILITY} \times \text{DISTANCE}_{ti} \\ & + \beta_{50} \times \text{AVAILABILITY} \times \text{RELEVANT SENTENCES}_{ti} + \beta_{51} \times \text{BRIDGE}_i \\ & \times \text{AVAILABILITY} \times \text{RELEVANT SENTENCES}_{ti} \\ & + r_{0i} + r_{1i} \times \text{AVAILABILITY}_{ti} + r_{2i} \times \text{RELEVANT SENTENCES}_{ti} + r_{3i} \times \text{DISTANCE}_{ti} \\ & + r_{4i} \times \text{AVAILABILITY} \times \text{DISTANCE}_{ti} + r_{5i} \times \text{AVAILABILITY} \\ & \times \text{RELEVANT SENTENCES}_{ti} + e_{ti} \end{aligned}$$

## Equation for Full HLM Model Predicting Search Decisions

*Level 1 Model*

$$\text{Prob}(Y=1|\beta) = P$$

$$\log[P/(1 - P)] = P_0 + P_1 \times (\text{DISTANCE}) + P_2 \times (\text{RELEVANT SENTENCES})$$

*Level 2 Model*

$$P_0 = \beta_{00} + \beta_{01} \times (\text{BRIDGE}) + \beta_{02} \times (\text{BCK\_KNOWLEDGE}) + r_0$$

$$P_1 = \beta_{10} + \beta_{11} \times (\text{BRIDGE}) + r_1$$

$$P_2 = \beta_{20} + \beta_{21} \times (\text{BRIDGE}) + r_2$$

$$\text{Level 1 variance} = 1/[P(1 - P)]$$

*Mixed Model*

$$\begin{aligned} \eta = & \beta_{00} + \beta_{01} \times \text{BRIDGE} + \beta_{02} \times \text{BCK\_KNOWLEDGE} \\ & + \beta_{10} \times \text{DISTANCE} + \beta_{11} \times \text{BRIDGE} \times \text{DISTANCE} \\ & + \beta_{20} \times \text{RELEVANT SENTENCES} + \beta_{21} \times \text{BRIDGE} \times \text{RELEVANT SENTENCES} \\ & + r_0 + r_1 \times \text{DISTANCE} + r_2 \times \text{RELEVANT SENTENCES} \end{aligned}$$

**Equation for Full HLM Model Predicting Search Time***Level 1 Model*

$$\text{SEARCHTIME}_{i_i} = \pi_{0i} + \pi_{1i} \times (\text{RELEVANT SENTENCES}_{i_i}) + \pi_{2i} \times (\text{DISTANCE}_{i_i}) + e_{i_i}$$

*Level 2 Model*

$$\pi_{0i} = \beta_{00} + \beta_{01} \times (\#\text{SEARCHES}_{i_i}) + \beta_{02} \times (\text{BRIDGE}_{i_i}) + \beta_{03} \times (\text{BCK\_KNOWLEDGE}_{i_i}) + r_{0i}$$

$$\pi_{1i} = \beta_{10} + \beta_{11} \times (\text{BRIDGE}_{i_i}) + r_{1i}$$

$$\pi_{2i} = \beta_{20} + \beta_{21} \times (\text{BRIDGE}_{i_i}) + r_{2i}$$

**Mixed Model**

$$\begin{aligned} \text{SEARCHTIME}_{i_i} = & \beta_{00} + \beta_{01} \times \#\text{SEARCHES}_{i_i} + \beta_{02} \times \text{BRIDGE}_{i_i} + \beta_{03} \times \text{BCK\_KNOWLEDGE}_{i_i} \\ & + \beta_{10} \times \text{RELEVANT SENTENCES}_{i_i} + \beta_{11} \times \text{BRIDGE}_{i_i} \times \text{RELEVANT SENTENCES}_{i_i} \\ & + \beta_{20} \times \text{DISTANCE}_{i_i} + \beta_{21} \times \text{BRIDGE}_{i_i} \times \text{DISTANCE}_{i_i} \\ & + r_{0i} + r_{1i} \times \text{RELEVANT SENTENCES}_{i_i} + r_{2i} \times \text{DISTANCE}_{i_i} + e_{i_i} \end{aligned}$$