

**Revealing the Comprehension Processes of Underprepared College Students:  
An Evaluation of the Reading Strategies Assessment Tool**

Joseph P. Magliano  
Georgia State University  
jmagliano@gsu.edu

Jodi P. Lampi  
Northern Illinois University  
jlampi@niu.edu

Melissa Ray  
University of Texas at Dallas  
Melissa.Ray@utdallas.edu

Greta Chan  
Northern Illinois University  
greta.c.chan@hotmail.com

To appear in: Journal of College Literacy and Learning, Vol. 46 (May, 2020)

Journal of College Literacy and Learning peer review process: Manuscripts submitted to this journal undergo editorial screening and peer review by anonymous reviewers. More details can be found at <https://j-cll.org/call-for-manuscripts>

**Contact Author:**

Joseph P. Magliano  
Professor  
Georgia State University  
33 Gilmer St. SE  
Atlanta, GA 30303

E: [jmagliano@gsu.edu](mailto:jmagliano@gsu.edu)

**Acknowledgement**

This research was supported by the Institute of Educational Sciences, US Department of Education (R305A150193, R305G040055, R305F100007) and the Center for the Interdisciplinary Study of Language and Literacy (CISLL) at Northern Illinois University. All opinions expressed in this manuscript are that of the authors and not of IES or the US Department of Education.

### Abstract

Coherent mental models for successful comprehension require inferences that establish semantic *bridges* between discourse constituents and *elaborations* that incorporate relevant background knowledge. While it is established that individual differences in the extent to which postsecondary students engage in these processes are correlated with reading outcomes, there is little research exploring if first-year students and first-year students enrolled in developmental education (DE) differ in the extent that they engage in these processes. In this manuscript, authors report on the implementation of the Reading Strategy Assessment Tool (RSAT) with first-year students and first-year students enrolled in DE employing a think aloud protocol. RSAT is a computer-based system that collects typed “think aloud” protocols and analyses of the protocols for inference processes that support mental model construction. In this study, RSAT scoring was compared to human coding of the protocols. There was convergence between both approaches for scoring the protocols. Controlling for comprehension proficiency, both groups bridged to a comparable extent. There was evidence of less elaborative processes in the first-year students enrolled in DE, suggesting differences in relevant background knowledge across both groups. This study illustrates the utility of RSAT in the study of underprepared college students.

Keywords: Inference, Comprehension, Computer-based assessment, natural language processing

## **Revealing the Comprehension Processes of Underprepared College Students: An Evaluation of the Reading Strategies Assessment Tool**

An important factor for success in college is a student's ability to successfully engage in academic reading, which we conceptualize as the ability to purposefully use text(s) to accomplish a variety of tasks associated with one's coursework (Britt, Rouet, & Durik, 2018; Rouet, 2006; Simpson, Stahl, & Francis, 2004; Snow, 2002). Students also need to be able to purposefully apply reading strategies that are appropriate both to the task (e.g., studying or writing a paper) and to the discipline (Shanahan & Shanahan, 2008; Shanahan, Shanahan, & Misischia, 2011). However, a foundational aspect of academic reading is basic comprehension, and in particular for a task that requires students to understand large segments of the texts (Graesser, Singer, & Trabasso, 1994; McNamara & Magliano, 2009b). Basic comprehension refers to a reader's ability to build a coherent mental model for a text or text segment (e.g., Kintsch, 1988; 1998). If students cannot read and construct a mental representation that accurately reflects the content of the texts they are asked to read, then obviously they work with a disadvantage using that context to take a test, write an essay, actively participate in class discussion, etc.

Unfortunately, a large number of students come to college not ready to meet the literacy challenges that they will face in their credit bearing courses (Bailey, 2009; Greene & Foster, 2003; Jenkins & Boswell, 2002). In fact, the Achieving the Dream Initiative indicated that more than 60% of community college students take at least one developmental education course to improve proficiencies in some dimension(s) associated with academic success (Bailey, 2009). Unfortunately, many DE programs are not reaching their intended goal of helping students successfully transition into credit bearing courses (e.g., Alliance for Excellent Education, 2006;

Bettinger, & Long, 2004; 2005; Calcagno & Long, 2008; Jenkins, Jaggars, & Roksa, 2009; Martorell & McFarlin, 2007; Vandal, 2010; Wirt, Choy, Rooney, Provasknik, Sen, & Tobin, 2004). For example, only 46% of community college students placed in DE reading courses finished their sequence (e.g., reading, writing, math) of DE courses (Bailey, 2009; Bailey, Jeong, & Cho, 2010), and a much lower percentage earned a degree.

Students are typically referred for enrollment in developmental reading courses based on their low performance on high-stakes standardized tests such as the ACT or SAT (Calcagno & Long, 2008), or an institution-specific, standardized placement exams (e.g., ACCUPLACER or Compass) that gauge students' reading comprehension skills. The utility of standardized tests of comprehension in placement protocols has been questioned for a variety of reasons, often around identifying appropriate cut-off scores (Barnett & Reddy, 2017) or their utility in a larger portfolio of student records (Burdman, 2012).

However, we contend that one problem is that the multiple-choice format is subject to test taking strategies that may be unique to that testing context (Rupp, Ferne, & Choi, 2006), some of which lead to subpar performance (e.g., skipping the texts, reading the questions, and then searching the text for segments that have the answers). Moreover, they generate a single score that provides little information about the strengths and challenges that students have in the cognitive processes that support academic literacy and basic comprehension (Magliano & Millis, 2003). Consistent with Stahl, Simpson, and Hayes (1992), we argue that research that is directed at the cognitive processes that support aspects of academic literacy, and in the case of the present study, basic comprehension, are needed to understand the strengths and challenges of students enrolled in developmental education courses. Standardized tests of comprehension will not likely provide useful insights to this end.

In the present study, we explored the extent that thinking aloud during reading can provide information about the extent to which the processes that support basic comprehension may be similar or different between first-year students and first-year students enrolled in developmental education courses. As will be discussed in the next section, thinking aloud is sensitive to basic comprehension processes delineated by theories of comprehension (e.g., Magliano & Millis, 2003; Millis, Magliano, & Todaro, 1996). A novel methodological feature of this study is that it employed the Reading Strategy Assessment Tool (RSAT; Magliano, Millis, The RSAT Assessment Team, Levenstein, & Boonthum, 2011). RSAT uses computational tools to analyze open ended answer to questions embedded in texts, including questions that are intended to evoke a think aloud response (but obviously typed). This study provided an assessment of the extent that the approach may have value in understanding struggling post-secondary readers that could be used in conjunction with more traditional standardized tools (e.g., McMaster, van den Broek, Epstein, White, Rapp, Kendeou, Bohn-Gettler, & Carlson, 2012).

### **Basic Comprehension**

Most theories of comprehension argue that comprehension arises from the construction of a coherent mental model (Graesser, Millis, & Zwaan, 1997). A coherent mental model can be conceptualized as a network of interconnected propositions, which reflect both ideas explicitly conveyed in the texts and inferences generated by the reader (e.g., Graesser et al., 1994; Kintsch, 1988; 1998). There are two broad categories of inferences that support mental model construction and specifically, bridging inferences and elaborative inferences (McNamara & Magliano, 2009b). Bridging inferences involve establishing connections between explicit discourse constituents, which can involve resolving anaphora (e.g., identifying the referents to pronouns) or inferring situational (causal) logical relationships. Elaborative inferences involve

readers drawing upon knowledge that is beyond the discourse context and relies on general knowledge structures (Graesser & Clark, 1985; Graesser et al., 1994; McNamara & Magliano, 2009b; Singer, 1988). Elaborative inferences can be generated based on existing generic knowledge of the world (e.g., Seifert, Dyer, & Black, 1986; Graesser & Nakamura, 1982), domain/text topic specific knowledge (e.g., McNamara & Kintsch, 2006), or based reasoning beyond the text content.

### **Think Aloud Methodology**

One approach for studying how bridging and elaborative inferences support comprehension and mental model construction is in the context of a think aloud methodology, which is of particular interest for the present study (Denton, Enos, York, Francis, Barnes, Kulesz, Fletcher, & Carter, 2015; Kendeou & van den Broek, 2007; Magliano & Millis, 2003; Trabasso & Magliano, 1996). In the context of a think aloud methodology, participants are asked to produce their thoughts at specific points in a text (e.g., Magliano & Millis, 2003; Trabasso & Magliano, 1996) or when they choose to do so (e.g., Coté & Goldman, 1999; Pressley & Afflerbach, 1995). In the present study we adopted the former methodology, as it is appropriate for assessing the processes that support mental model construction across different populations of readers (Magliano, 1999). That is, there is control over the locations where the protocols are produced, which affords a comparison of the processes that occurred at those locations.

Think aloud protocols reveal a broad range of processes that potentially support comprehension, such as inferences, paraphrases and restatements of text content, metacognitive statements and strategies, and affective evaluations of the texts (Denton et al., 2015; Magliano, 1999; Pressley & Afflerbach, 1995; Kendeou & van den Broek, 2007). Some processes revealed

in think aloud protocols are indicative of successful mental model construction, such as inferences (bridging and elaborative), paraphrases, and metacognitive processes (e.g., Pressley and Afflerbach, 1995). For example, there is extensive evidence suggesting that the extent that readers engage in bridging inferences is positively correlated with a variety of measure of comprehension for text for which the protocols were collected (Magliano & Millis, 2003; Millis, Magliano, & Todaro, 1996; Magliano et al., 2011; Magliano, Trabasso, & Graesser, 1999) and other texts (Magliano & Millis, 2003; Magliano et al., 2011). Additionally, the more students spontaneously “self-explain” a text as they think aloud is indicative of deep comprehension (Chi, Bassok, Lewis, Reimann, & Glaser, 1989; McNamara, 2004). When they self-explain, readers use a variety of strategies, such as bridging, elaborating, and paraphrasing to explain why content is being mentioned given the point of a text and how content is relevant to larger issues associated with the topic of the text (McNamara, 2004). Finally, successful comprehenders (i.e., demonstrate relatively high levels of performance in comprehension tests) tend to demonstrate higher levels of metacognitive awareness than less successful comprehenders (Helder, van Leijenhort, & van den Broek, 2016; Pressley & Afflerbach, 1995).

Other processes are less successful for supporting comprehension. For example, Todaro Magliano, Millis, McNamara, and Kurby (2008) had college students think aloud to science texts. They identified the extent that the words used by the students were indicative of strategies associated with self-explanation (e.g., bridging, elaboration, paraphrasing) or tangentially associative in nature, and in particular affective evaluations (e.g., “cancer is scary” when reading a text on how tumors develop) and recollective statements (e.g., “my grandfather had cancer”). The extent to which students engaged in processes associated with self-explanation was positively correlated with performance on a comprehension test for the text, whereas the

tangential processes were negatively correlated with performance. They reasoned that the extent to which students engaged in these processes consumed resources that could otherwise be devoted to mental model construction.

Performance on a think aloud task is also sensitive to individual differences in comprehension processes in a college population (Magliano & Millis, 2003; Magliano et al., 1999; 2011; Millis et al., 2006). For example, Magliano and Millis (2003) had college students read simple narratives and think aloud at predetermined sentences. They administered the Nelson-Denny test of reading and operationalized comprehension proficiency based on upper and lower quartile performance of the sample. They found that more proficient comprehenders (as determined by performance on an experimenter administered standardized test of comprehension) tended to bridge more than less proficient comprehenders when thinking aloud, whereas less proficient readers tended to paraphrase more than more proficient readers. As such, there is reason to believe that a think aloud procedure would be useful to explore the extent to which there are differences in how students enrolled in a developmental literacy course process texts differently than peers enrolled via traditional admissions criteria.

While a think aloud methodology has virtue as a tool for understanding strengths and challenges of developmental readers, it also has been proposed to have instructional virtues (Ebner & Ehri, 2016; Nist & Kirby, 1986). Think aloud externalizes the processes that would otherwise not be necessarily available to conscious awareness during silent reading (Trabasso & Magliano, 1996). Thinking aloud requires metacognitive processes (McNamara & Magliano 2009a), which is a cognitive process that is highly targeted and recommended in reading course curricula (Armstrong & Lampi, 2017; Mokhtari, 2017). As noted above, thinking aloud is sensitive to individual difference in comprehension proficiency. As such, the products of



thinking aloud give an externalized record of comprehension processes and strategies for students and instructors to evaluate (Nist & Kirby, 1986). Finally, instructors can use thinking aloud to model best practices in terms of how to actively engage in texts to support learning (Nist & Kirby, 1986). However, using thinking aloud as an instructional tool should ideally be informed by basic research directed at using thinking aloud to understand strengths and challenges in basic comprehension.

### **Computer-Based Assessment of Verbal Protocols**

Over the past two decades there have been dramatic advances in the development of computer-based systems that can automatically analyze verbal protocols, such as texts recall protocols, summaries, essays, answers to open ended questions, and think aloud protocols (Graesser & McNamara, 2011; Magliano & Graesser, 2012). These systems have been applied to intelligent tutoring systems that can support learning from texts, systems that teach reading strategies, and stand alone assessment systems (Magliano, & Graesser, 2012). In the present study, we used the RSAT (Magliano et al., 2011), which allows one to collect and analyze answers to open-ended questions. RSAT may have value in shedding insights into the comprehension processes of students enrolled in developmental literacy programs and how they may be different from students admitted through traditional means. This study can be construed as a proof of concept of the utility of RSAT in this research context.

In RSAT, students read texts one sentence at a time, presented on a computer. While not a naturalistic way to read, this decision was made to force students to use their mental model to answer questions that are periodically posed to them (Gilliam, Magliano, Millis, Levinstein, & Boonthum, 2007). Students received two types of questions that appear after preselected sentences: *direct* and *indirect questions*. Direct questions are adjunct why and how-questions about the sentence that was just read (e.g., “Why did the Confederate Government move the

capital to Richmond?” on text about the First Battle of Bull Run) that require students to access content in their mental model. The idea is that the more complete the answer, the better the test taker is demonstrating that they have built a coherent mental model for the texts. *Indirect questions* are intended to evoke a think aloud response and are always “What are you thinking now?” and students are given practice on how to answer these questions with responses akin to thinking aloud. The locations of these questions were pre-selected and empirically validated as locations that reveal individual differences in comprehension and inference processes (Magliano et al., 2011).

RSAT uses very simple computational algorithms to automatically analyze the answers to direct and indirect protocols, and specifically key word matching and Soundex (Birstwisle, 2002), which is an algorithm used to handle misspellings and word form changes. Answers to direct questions are compared to ideal answers and are scored by the number of content words (nouns, pronouns, verbs, adjectives, and adverbs) in the student answers that are in the ideal answers. Answers to indirect questions are analyzed to detect three processes; paraphrasing the prompt sentence, bridging to the prior discourse context, and elaborating upon the text. Paraphrase scores are computed by counting the content words (nouns, verb, adverbs, and adjectives) that were present in the prompt sentence. Bridging scores are computed by counting the content words in the protocols that appeared in the prior discourse content. Elaboration scores are computed by counting content words that have not appeared in the prior discourse context.

While surprisingly simple, RSAT automatic scoring shows good construct and predictive validity (Magliano et al., 2011). RSAT scores on direct questions correlated with performance on the Gate-MacGinitie test of comprehension ( $r = .53$ ) and the comprehension portion of the

ACT ( $r = .54$ ), which is comparable to correlations between those tests ( $r = .59$ ) (Magliano et al., 2011). Finally, RSAT has convergent validity in that the RSAT scores are highly correlated with human judgments of the verbal protocols with Pearson correlations ranging from .50 to .78, whereas the correlations between human judges ranged from .89-.92 (Magliano et al., 2011). RSAT has been primarily used as a research tool to learn about comprehension processes in a postsecondary population (Higgs, Magliano, Vidal-Abarca, Martinez & McNamara, 2017; Magliano, Durik, & Holt, 2011).

### **Study Goals and Research Questions**

The goal of the present study was to explore the extent that first-year students in a four-year institution process texts differently than first-year students enrolled in developmental education courses. RSAT was used to collect think aloud responses as participants processed texts, and both human judgments and RSAT scoring was used to analyze the protocols. Ultimately, the questions driving this present study were two-fold. Do first-year students process texts differently than first-year students enrolled in developmental education courses? And, does the use of RSAT automatic coding of the protocols reveal these differences in a similar manner as human coding of the protocols? If the answer to the second questions is yes, it provides a proof of concept of RSAT as research tool, and perhaps with additional research, one that could be beneficial to practitioners in monitoring students enrolled in developmental literacy programs designed to promote comprehension proficiencies. Given the exploratory nature of this study, we did not postulate predictions regarding differences in either the hand coding of the RSAT scoring of the indirect protocols across the two populations.

## Methods

### Data Sources

This particular study utilized two archival data sets from two separate studies conducted at the same institution, implementing RSAT. One data set involved first-year students enrolled in developmental courses at a large four-year institution in the Midwest. The second archival data set involved first-year students from the same institution. Admissions history was not collected for the study regarding first-year students, but ACT scores were obtained. These were used to ensure that the first-year students in the comparison group were never enrolled in developmental education courses at that institution, and specifically only students who had ACT scores above the cut-off for enrollment were used (i.e., composite scores above 19). Essentially, this study compares prior RSAT data from two groups of students from the same institution looking at comprehension processes between the groups as determined by computer and hand-coded analyses.

### Participants

Participants were classified into two enrollment groups: first-year students and first-year students enrolled in developmental education courses. A matched-pair design was used; first-year students in the comparison group were selected for inclusion based on their scores on RSAT comprehension questions. More specifically, each first-year student enrolled in developmental education courses was matched to a student in the comparison data set that had a similar RSAT comprehension score. The RSAT comprehension score is described below and is a measure of overall comprehension skill, and was used to create enrollment groups with similar levels of overall comprehension skill. Specifically, for each student enrolled in developmental education, we identified participants from the first archival data set (see Magliano et al., 2011) that had RSAT comprehension scores within .02 points of their scores. If there were more than one

student identified, we then randomly selected participants as the match. However, we did not treat the sample as a within-participants variable as is possible with matched samples because RSAT comprehension scores were the only basis for matching.

There were a total of 46 participants in the study: 23 first-year students enrolled in developmental education courses and 22 first-year students in the comparison group. One participant in the first-year student enrolled in developmental education courses sample had incomplete think aloud data, and therefore was dropped from the analyses. There was no additional student who met the matching criteria from the first archival data set.

Demographic information was not available from both archival sets for this study's participants, so described here are the demographics of first-year students and first-year students enrolled in developmental education courses at this institution. Participants for both archival data sets were selected from these umbrella demographics at a large Midwestern university.

Of the first-year students enrolled in developmental education courses, 60.2% identified as female and 39.8% identified as male. In addition, 67.7% were African-American, 13.8% were Caucasian/Non-Hispanic, 10.4% were Hispanic, 8.8% were Asian, 2.3% were Other, 1.0% were Multi-Ethnic, and .38% were American Indian/Native Alaskan. Of those first-year students enrolled in developmental education courses invited to complete the RSAT, 23 participants came forward.

Of the first-year students enrolled at this large Midwestern university, 57% were Caucasian/Non-Hispanic, 16% were African Americans, 15% were Hispanic, and 5% were Asian. The gender identity was roughly equal, with 49% female and 51% identifying as male (at the time the study was conducted, there were only two options for gender identity).

**Text Material**

Texts consisted of fiction and non-fiction texts. Non-fiction texts covered topics in science and history. The non-fiction texts contained an average of 25.5 sentences per passage, while the fiction texts had an average of 30.75 sentences per passage. Although students read the same number of texts from each genre (two per genre of science, history, and narrative), the actual topics of the texts varied across the participants.

**Instruments.**

RSAT is a computer-administered test that is designed to assess a student's level of comprehension and the processes that support it while reading (Gilliam et al., 2007; Magliano et al., 2011). Students read texts one sentence at a time and at pre-selected target sentences they are prompted to answer two different types of questions during reading. *Direct questions* are intended to provide an assessment of comprehension and involve why and how questions (Graesser & Clark, 1985). *Indirect questions* require readers to report thoughts regarding their understanding of the sentence in the context of the passage, akin to thinking aloud (Trabasso & Magliano, 1996). The texts are not available when they answer the questions, and so participants must consult their mental representation when producing responses to direct and indirect questions.

RSAT uses automatic natural language algorithms to automatically score responses to direct and indirect questions. Specifically, key word matching (i.e., letter to letter matching) and soundex algorithms (Birtwisle, 2002) were used to identify the number of content words (nouns, verb, adverbs, and adjectives) that were produced in a protocol that were present in *semantic benchmarks*, which vary depending on the type of question. Analyses of the answers to direct questions were used to determine a comprehension score and the semantic benchmarks were ideal answers that have been previously developed and tested (Gilliam et al., 2007). The

algorithms were used to count the number of content words in the participants' answers that were present in ideal answers that were constructed by the test designers. The answers to the indirect questions provided processing performance scores for three processes associated with building coherent mental models of texts: paraphrasing score, bridging score, and elaboration score. RSAT scoring is based on the assumption that these processes can be inferred from words produced in the protocols (Magliano & Graesser, 2012). Paraphrasing in this context refers to producing content from the sentences that were just read prior to the indirect prompts, and therefore semantic benchmarks for paraphrasing scores were the content words in the sentences immediately preceding indirect question prompts (i.e., the sentences that were just read prior to the indirect prompts). Bridging refers to establishing how the current sentence is related to prior discourse content, and therefore the semantic benchmark for bridging scores were the content words in the sentences prior discourse context (i.e., sentences that preceded the sentence occurring right before the indirect prompts). Elaboration refers to drawing upon knowledge not explicitly present in the discourse context. Elaboration scores are based on words produced by participants did not appear anywhere in the text prior to the indirect prompts. Mean for comprehension scores, paraphrasing scores, bridging scores, and elaboration scores were computed for each participant.

**Procedure**

Participants took RSAT administered on personal computers in a web-based environment (Gilliam, et al. 2007; Magliano et al., 2011). The texts were presented in black font in a gray field left justified near the top of the computer screen. The title of each text remained centered at the top of the screen while participants read the entire text. In the current study, only one sentence of a text was shown on the screen during reading because this presentation has been shown to be a good predictor of comprehension skill (Gilliam et al., 2007). Participants

navigated forward through the text by clicking on a “next” button, which is located near the bottom left portion of the computer screen. “NEW PARAGRAPH” markers appeared when there is a shift to a new paragraph. After participants clicked the “next” button, the next sentence appeared, provided it was a non-target sentence. The text sentences were not present on the screen when there was a question prompt, and the participants could not navigate back and reread the texts in response to the questions. For target sentences, a response box appeared to the right of the “next” button with a prompt above the box. The prompt for an indirect question was “*What are you thinking now?*” For direct questions, the target sentence was removed from the screen when the question and response box appeared. Participants typed their answers to the question in the response box. They clicked the next button when they were finished, after which the response box disappeared and the next sentence was presented. Responses were recorded on a computer server. The order of the texts was randomly presented to the participants.

### **Protocol Hand Coding**

The RSAT protocols were also coded by human raters, which were of primary interest for the present study. A coding system was developed to hand code the indirect protocols. Participants’ responses to the think-aloud questions were parsed into subject-verb clauses and then classified into one of 14 categories of comprehension strategies. Categories were based on previous comprehension research (Coté, Goldman, & Saul, 1998; Trabasso & Magliano, 1996, van den Broek, Lorch, Linderholm, & Gustafson, 2001). Table 1 contains the 14 coding categories and examples. Each verb clause in the protocols was coded as belonging to one of the 14 categories, and a verb clause could only belong to one category. Interrater reliability was conducted on 25% of the sample, with an inter-rater agreement of .78 (proportion of judgments that were in agreement). Discrepancies were resolved through discussion. The data were initially presented for all 14 categories. However, for the purposes of analysis of interest,



categories were collapsed. Given that Todaro et al (2008) found that associative elaboration, recollections, and evaluations were not supportive of comprehension and occurred at low frequencies, we also collapsed these into a category called *less helpful response*. The different questions types were also collapsed into one category. The comprehension strategies fell into one of six major groups: Paraphrase, Bridging inference, Text-relevant elaborative inferences, questions (combination of knowledge-based questions, text-based questions, and vague questions), metacognitive statements, and less supportive processes (combination of associative elaborations, recollections, evaluations, and vague statements). The proportion of verb clauses reflecting the different categories were computed for each protocol, and mean proportion scores were computed for each participant for the six categories.

## Results

Bayesian analyses were used in order to compare differences between the groups in, and specifically Bayes' factors were computed (Masson, 2011). Bayes factors have at least two advantages over traditional t-tests. First, Bayes factors are viewed as an appropriate approach for controlling for type 1 errors when there are a high number of comparisons being made and second, they may afford the interpretation of null results (Wetzell, Matzke, Lee, Rouder, Iverson, & Wagenmakers, 2011). Bayes' factors are an odds ration that ranges from 0 to infinity, but are centered on 1. Bayes' factors of 1 reveal that there is no difference between group means, but do not afford an interpretation that the lack of difference is meaningful. A Bayes' factors greater than 3 are seen as evidence that the differences between the two groups are meaningful and the

Table 1  
*Coding Categories and Examples*

| Category                 | Description   | Example   |
|--------------------------|---|---|
| Paraphrase               | Idea unit that is from the current sentence   | (1) That erosion destroys things on and in the earth<br>(2) That Don Juan was ashamed to go with his wife                             |
| Bridge                   | Idea unit that is based on a prior text sentence. It reflects an attempt to connect the current sentence to the prior text content.   | (1) Water speeds up the process of erosion.<br>(2) The Nationalists were quick, and/ move in from the West and southern part of Spain |
| Elaboration-Textual      | Idea unit based on prior knowledge that is related to the main points of the text. Can take the form of explanations, predictions, conclusions drawn from text content  | (1) Raindrops must be negatively charged then.<br>(2) Bribery is not the way to win another person s heart.                           |
| Elaboration-Associative  | Idea unit based on prior knowledge that is tangentially related to the main point(s) of the text.   | (1) Everyone should be aware of cancer<br>(2) Thunderstorms sometimes have tornadoes  |
| Recollection             | Idea unit based episodic information from reader's own life (events, people, places)  | (1) I've taken a course on the weather.<br>(2) I visit a cottage on Cape Cod every year   |
| Erroneous-Irrelevant     | Idea unit that is incorrect or that are irrelevant to the text.   | (1) Thunder is ions forced out or pulled apart<br>(2) We all have cancer cells in our body,   |
| Affective Evaluation     | Idea units that describe the general quality (good/bad, right/wrong) of the text, author, or reading task.  | (1) Cell division is interesting.<br>(2) I think that this is a good story so far   |
| Metacognitive            | Idea unit reflects (1) the degree to which the reader understands/knows something, (2) the ease or difficulty the reader is having in processing the texts (including the ease/difficulty of text), or (3) reader's thought processes | (1) I am a little lost<br>(2) Some of this information I already knew about<br>(3) I find this easy to read                           |
| Question-Knowledge-based | A wh-question (e.g. why, what, how) or yes/no question that asks about information that is not contained in the text up to that point.  | (1) Where the ghosts really people?<br>(2) If a family member has cancer, are you more at risk?                                       |

|                     |   |  |
|---------------------|---|--|
| Question-Text-based | A wh-question (e.g. who, why, how) or yes/no question that contains propositions from or the gist of the current sentence or prior texts sentences  | (1) What causes the thunder to develop or form?<br>(2) Will everything erode with enough time? |
| Vague Question      | A wh-question (e.g. who, why, how) or yes/no question that contains propositions from or the gist of a sentence occurring more than one sentence before the current one, or a macro-proposition | (1) How?<br>(2) What is going on?  |

---

strength of that interpretation increases as they increase; Bayes' factors less than .33 are considered evidence that the null effect is meaningful, with the strength of that evidence increasing as they approach 0. We also report the results of traditional t-tests for comparisons, but restricted our interpretations to the Bayes factors.

Preliminary analyses were conducted to demonstrate that the comprehension scores based on RSAT did not differ and that the two samples produced protocols at about the same length in terms of clauses. The average RSAT comprehension scores (i.e., direct questions scores) did not differ between the first-year students enrolled in developmental courses' sample ( $m = 1.49$ ,  $sd = .59$ ) and first-year students' samples ( $m = 1.46$ ,  $sd = .58$ ),  $t(44) = .02$ ,  $p = .98$ , *Bayes Factor* = .12. Participants in the two groups produced a similar number of clauses, with first-year students enrolled in developmental courses ( $m = 2.28$ ,  $sd = .97$ ) and first-year student comparison ( $m = 2.30$ ,  $sd = .77$ ) groups not being statistically different,  $t(44) = .10$ ,  $p = .92$ , *Bayes Factor* = 12. The Bayes Factors indicate that the null effects are meaningful. While that is to be expected given the matching procedure, it is important to demonstrate that there are no differences in the verbosity between the two samples.

Table 2 shows the mean proportion scores for the categories scored by the human coders and those computed in RSAT and the inferential statistics comparing the means between the two groups. First consider those strategies that were directly related to mental model construction (i.e., paraphrase, bridging inference, and elaborations. There was strong evidence that first-year students sample produced more elaborations than the first-year students in developmental courses sample for both the hand coding and RSAT. The results for paraphrasing and bridging indicated that there was no difference, but the analyses based on RSAT provided stronger evidence that the null effect was meaningful. Specifically, with respect to RSAT scores, the

Bayes' factor for both scores indicated that the null effect was meaningful.

Next consider metacognitive, questions, and less supportive strategies. There was strong evidence that the first-year students enrolled in developmental coursework sample produced more questions than the traditionally enrolled sample. This finding will be discussed in greater detail in the discussion section, but this might stem from the fact that they were in a support course that encouraged and emphasized strategic thinking and reading, which can involve generating questions. The Bayes' factor for metacognitive statements supported the interpretation that the null effect is meaningful and indicates that the two samples produced these statements at the same frequency. There was no difference between the samples for less supportive strategies, but the Bayes' factor did not support the interpretation of the null effect as meaningful.

Table 2

*Means (Standard Deviations) and inferential statistics for comprehension processes based on both human judgments and RSAT scoring*

| Processes                 | First-Year<br>Students in<br>DE | First-Year<br>Students | <i>t</i> | <i>p</i> | <i>Bayes'</i><br><i>Factors</i> |
|---------------------------|---------------------------------|------------------------|----------|----------|---------------------------------|
| <b>HUMAN JUDGMENTS</b>    |                                 |                        |          |          |                                 |
| Paraphrase                | .12 (.08)                       | .09 (.06)              | 1.76     | .085     | .50                             |
| Bridge                    | .20 (.14)                       | .14 (.10)              | 1.76     | .085     | .50                             |
| Elaboration-Textual       | .24 (.12)                       | .40 (.14)              | 4.16     | >.001    | 146.21                          |
| Questions                 | .16 (.16)                       | .05 (.06)              | 2.91     | .006     | 5.18                            |
| Metacognitive statements  | .15 (.13)                       | .14 (.12)              | 0.14     | .89      | 12                              |
| Less-supportive processes | .12 (.11)                       | .18 (.12)              | 1.58     | .12      | .38                             |
| <b>RSAT SCORING</b>       |                                 |                        |          |          |                                 |
| Paraphrase score          | .84 (.50)                       | .89 (.43)              | .32      | .75      | .12                             |
| Bridging score            | 1.63 (.71)                      | 1.49 (.90)             | .61      | .54      | .14                             |
| Elaboration Score         | 2.77 (.98)                      | 3.91 (1.49)            | 3.04     | .004     | 7.15                            |

*Note.* Less-supportive processes included associative elaborations, recollections, and evaluations

## Discussion

This study was conducted to explore the extent that first-year students enrolled in developmental education courses process texts differently than first-year students not enrolled in developmental education courses. RSAT was used to collect typed “think aloud” protocols in order to evaluate comprehension processes. We hand coded the protocols and relied on computational analyses to explore these differences. The most dramatic finding was that the participants enrolled in the developmental education courses had lower elaboration scores than those not enrolled in the program. Elaborations provide an important strategy for incorporating background knowledge into the mental model for expository texts (e.g., McNamara, 2004). The results indicated that there were no differences between the samples with respect to other strategies that support mental model construction, and in particular bridging and paraphrasing, but the results Bayes’ analysis only supported the interpretation that the lack of difference was meaningful for RSAT. Bridging is a critical skill for comprehension (McNamara & Magliano, 2009), and a lack of difference between the samples is encouraging because it suggests that the first-year student in developmental education have similar proficiency in this process as their traditionally enrolled cohort. However, given the lack of convergence with the gold standard of hand coded, we hesitate over interpret the null effects for RSAT. Additionally, we found that first-year students in the developmental courses generated more questions than those not in the courses.

The most notable difference between the two populations pertained to text relevant elaborations. One possible explanation for this difference is that lower text relevant elaboration scores may reflect a deficit in background knowledge, which is a potential barrier to comprehension (e.g., McKeown, Beck, & Sinatra, Loxterman, 1992; Pearson, Hansen, &

Gordon, 1979). If this explanation is accurate, then one culprit for the lower scores could be the nature of academic experiences prior to coming to college. The developmental program at Midwestern university does recruit from underperforming schools in urban and rural areas. As a result, prior to coming to college, these students may have read less than their first-year counterparts, and it is well documented that exposure to reading promotes knowledge growth (Alexander, 2000, 2003; Stanovich & Cunningham, 1992; West & Stanovich, 1991). Second, lower elaboration scores could be indicative of lower domain independent, critical thinking skills. Elaboration may reflect an ability and disposition to reason with and beyond the text. If this is the case, the weak elaboration scores may not be bound to reading, but reflect a general strategy towards learning contexts. Finally, the lower scores could reflect a different understanding of the task and a willingness to bring relevant background knowledge to bear. The present study was not designed to discriminate between these possibilities, but certainly points to a need to replicate this find and explore why it occurs.

The first-year students enrolled in developmental education coursework generated more questions than their first-year counterparts. At one level, this is encouraging because it has been shown that self-generated questions promote learning from texts and complex material (Graesser, McMahan, & Johnson, 1994), which is why it is often promoted in reading strategy interventions (e.g., Beck, McKeown, Hamilton, & Kugan, 1997; Palinscar & Brown, 1984). Participants in this study would have been exposed to reading strategy training by the time they participated in the study, and the differences in the groups could reflect that the first-year students in developmental education were using some of the strategies taught in the course. However, generating questions is a relatively easy approach to thinking aloud, and in isolation of other strategies, may not promote comprehension.

The convergence between RSAT and the hand coding is encouraging in the context of efforts to develop computer systems that can automatically code student-constructed responses (Graesser & McNamara, 2011; Magliano & Graesser, 2012). RSAT was primarily developed as a research tool (Magliano et al., 2011) for using constructed verbal protocols. While the sample in this study is relatively small, it affords using protocols on a larger scale. Certainly, there are aspects of processing (quality of the strategies) and types of processing (e.g., questions) that it cannot currently detect, but it does focus on strategies that are well documented in theory and research as being important for comprehension. It may have utility as a tool for practitioners, but more work is needed to realize that possibility (Magliano, Ray, & Millis, 2016). In the context of developmental education, it would have utility to the extent that courses are trying to teach strategies that support mental model construction, such as self-explanation training (McNamara, 2004; McNamara, Levenstein, & Boonthum, 2004). Self-explanation is a process of explaining challenging texts to oneself while reading, and involves a cluster of strategies, such as paraphrasing, bridging, and elaboration. The more students self explain naturally, the better they tend to learn from complex texts (Chi et al., 1989), and training promotes comprehension skills in struggling readers (e.g., McNamara et al., 2004). RSAT could play a role as a formative assessment of the development of this skill (Kurby, Magliano, Dandotkar, Woehrle, Gilliam, & McNamara, 2012).

### **Limitations**

There are a few limitations to this study that warrant consideration. It is important to acknowledge that this study is based on a relatively small sample to compare different populations, albeit it is a large sample in the context of think aloud studies. There have been recent arguments to ensure that findings of studies can be replicated and this is particularly the case for studies with small samples. Additionally, this study involved archival data, and



unfortunately demographic information is not available, which has implications on the generalizability of these results. Similarly, we did not have information on individual difference factors (e.g., working memory, reading proficiency, vocabulary knowledge) that might affect comprehension strategies (e.g., Kopatich, Magliano, Millis, Parker, & Ray, in press). Assessing a more comprehensive battery of assessments would improve upon the matching procedures and potentially afford the assessment of the factors that affect individual differences in the comprehension strategies within the different populations. These limitations lead to the conclusion that future research should assess the extent that these findings can be replicated and extended with a larger sample. The use of RSAT as a research tool affords collecting large samples in verbal protocols, which may not be feasible when one has to rely on hand coding the responses.

A final limitation is that the texts used in this study were not aligned with the participants' cultural backgrounds. While that is often the case in educational contexts, and in disciplinary specific college courses (e.g., psychology, biology, and chemistry), some educational researchers have argued that a misalignment in cultural background and texts inhibits students from engaging in complex thinking and reasoning (e.g., Gutierrez & Rogoff, 2003; Lee, 2006). Moreover, Lee (2006) has argued that culturally relevant texts can be used as an effective scaffold to teach students how to reason with and beyond disciplinary specific texts. As such, this study should be replicated with a manipulation of the cultural relevance of the texts to the participants. Such a study could gain insights in how to support struggling college readers in terms of learning how to use their background knowledge to support learning from unfamiliar and challenging texts.

## **Conclusions**

In conclusion, this study suggests that first-year students enrolled in developmental education coursework have some skills that require support and some that may require less support. The present study suggests that first-year students enrolled in developmental education courses engage in less elaborative strategies during reading. Importantly, reading promotes knowledge growth, which underscores the importance of addressing this issue. Given the goal of this study, RSAT is sensitive to these differences, which suggest that computer-based approaches for assessing reading strategies could have utility in developmental reading literacy programs. Strategies such as self-explanation training are intended to help low knowledge readers (McNamara et al., 2004) and may have a utility in developmental education instruction. RSAT is sensitive to changes in processing in response to iSTART, an intelligent tutoring system that teaches self-explanation (Kurby, Magliano, Dandotkar, Woehrle, Gilliam, McNamara, 2012). It specifically teaches students to rely on whatever background knowledge that they can bring to bear when reading unfamiliar and challenging texts. Specifically, these tools support reading processes that get students to employ domain-general knowledge to improve reading comprehension before domain-specific knowledge becomes a more nuanced factor of comprehension in content-specific courses. As such, there may be merit in exploring the use of these systems to support traditional programming and curriculum. However, there are classroom applications of teaching the process of self-explanation (McNamara, 2004). Interventions of this nature can potentially ameliorate the differences in elaboration observed in this study, but require sustained practice on the part of the students.

However, this potential use of iSTART underscores the importance of a study exploring if using culturally relevant texts would support elaborative processes in developing college readers relative to texts that are not culturally relevant. A study of this nature could be informative in terms of learning how to adapt iSTART and other systems to take advantage of the inherent strengths that developing college readers can leverage when reading challenging texts.

## References

- ACT (2006). *Reading between the lines: What the ACT reveals about college readiness in reading*. Retrieved June, 2008, from <http://www.act.org/path/policy/reports/reading.html>
- Alexander, P. A. (2000). Toward a model of academic development: Schooling and the acquisition of knowledge: The sequel. *Educational Researcher*, 29(2), 28-33, 44. Alliance for Excellent Education (2006). *Paying double: Inadequate high schools and community college remediation*. Retrieved from <http://www.all4ed.org/files/archive/publications/remediation.pdf>
- Armstrong, S. L., & Lampi, J. P. (2017). PILLAR: A reading strategy for a new era of strategy instruction at the college level. *Journal of College Literacy and Learning*, 43, 3-17.
- Bailey, T. (2009). Re-thinking developmental education in community college. *Community College Research Center Brief*, 40, 1-4.
- Bailey, T., Jeong, D. W., & Cho, S. W. (2010). Referral, enrollment, and completion in developmental education sequences in community colleges. *Economics of Education Review*, 29(2), 255–270.
- Barnett, E. A., & Reddy V. (2017, February). College placement strategies evolving considerations and practices. (CAPR Working Paper). New York, NY: Center for the Analysis of Postsecondary Readiness.
- Beck, I.L., & McKeown, M.G., Hamilton, R.L., & Kugan, L. (1997). *Questioning the author: An approach for enhancing student engagement with text*. Newark, DE: International Reading Association.

- Bettinger, E., & Long, B. T. (2004). *Shape up or ship out: The effects of remediation on students at four-year colleges* (NBER Working Paper No. 10369). Cambridge, MA: National Bureau of Economic Research.
- Bettinger, E., & Long, B. T. (2005). *Addressing the needs of under-prepared students in higher education: Does college remediation work?* (NBER Working Paper No. W11325). Retrieved May, 2009, from <http://ssrn.com/abstract=720411>
- Birtwisle M. (2002). *The Soundex Algorithm*. Retrieved from: [http://www.comp.leeds.ac.uk/matthewb/ar32/basic\\_soundex.htm](http://www.comp.leeds.ac.uk/matthewb/ar32/basic_soundex.htm).
- Britt, M.A., Rouet, J.F. & Durik, A.M. (2018). *Literacy Beyond Text Comprehension: A Theory of Purposeful Reading*. New York: Routledge.
- Burdman, P. (2012, February). Where to begin? The evolving role of placement exams for students starting college. Retrieved from [http://www.achievingthedream.org/sites/default/files/resources/Where\\_to\\_Begin.pdf](http://www.achievingthedream.org/sites/default/files/resources/Where_to_Begin.pdf)
- Calcagno, J. C. & Long, B. T. (2008). *The impact of postsecondary remediation using a regressions discontinuity approach: Addressing endogenous sorting and noncompliance*. A Report from The National Center for Postsecondary Research.
- Chi, M.T.H., Bassok, M., Lewis, M.W., Reimann, R., & Glaser, R. (1989). Self-explanation: How students study and use examples in learning to solve problems. *Cognitive Science*, 13, 145-182.
- Coté, N., & Goldman, S. R. (1999). Building representations of informational text: Evidence from children's think-aloud protocols. In H. van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 169–193). Mahwah: Erlbaum.

- Coté, N., Goldman, S. R., & Saul, E. U. (1998). Students making sense of informational text: Relations between processing and representation. *Discourse Processes, 25*, 1-53.
- Denton, C. A., Enos, M., York, M. J., Francis, D. J., Barnes, M. A., Kulesz, P.A., Fletcher, J. M., & Carter, S. (2015). Adequate and poor comprehenders reading accessible and challenging narrative and information text. *Reading Research Quarterly, 50*(4), 393-416.
- Ebner, R. J., & Ehri, L. C. (2016). Teaching students how to self-regulate their online vocabulary learning by using a structured think-to-yourself procedure. *Journal of College Reading and Learning, 46*(1), 62-73.
- 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, Instruments, & Computers, 39*, 199-204.
- Graesser, A. C., & Clark, L. F. (1985). *Structures and procedures of implicit knowledge*. Norwood, NJ: Ablex.
- Graesser, A. C., & McNamara, D. S. (2011). Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Sciences, 3*, 371–398.
- Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. In J. T. Spence, J. M. Darley, & D. J. Foss (Eds.), *Annual Review of Psychology, Vol. 48*. Palo Alto, CA: Annual Reviews Inc.
- Graesser & Nakamura, (1982). The impact of a schema on comprehension and memory. In G.H. Bower (Ed.), *The Psychology of learning and motivation* (pp 59-209). New York: Academic Press.

- Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*, 371-95.
- Gutierrez, K., & Rogoff, B. (2003). Cultural ways of learning: Individual traits or repertoires of practice. *Educational Researcher*, *32*, 19-25.
- Greene, J., & Foster, G. (2003, September). *Public high school graduation and college readiness rates in the United States* (Manhattan Institute, Center for Civic Information, Education Working Paper, No. 3). New York: Manhattan Institute.
- Helder, A., van Leijenhorst, L., & van den Broek, P. (2016). Coherence monitoring by good and poor comprehenders in elementary school: Comparing offline and online measures. *Learning and Individual Differences*, *48*, 17-23.
- Higgs, K., Magliano, J. P., Vidal-Abarca, E., Martínez, T., & McNamara, D. S. (2017). Bridging skill and task-oriented reading. *Discourse Processes*, *54*, 19-39.
- Jenkins, D., & Boswell, K. (2002). *State policies on community college remedial education: Findings from a national survey*. Denver, CO: Education Commission of the States.
- National Center for Education Statistics [NCES] (2004).
- Jenkins, D., Jaggars, S. S., & Roksa, J. (2009). *Promoting gatekeeper course success among community college students needing remediation*. Retrieved from <http://ccrc.tc.columbia.edu/Publication.asp?UID=714>
- 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*, 1567-1577.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.

- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction–integration model. *Psychological Review*, *95*(2), 163–182.
- Kopatch, R. D., Magliano, J. P., Millis, K. K., Parker, C. P., & Ray, M. (in press). Understanding how language specific and domain general resources support comprehension. *Discourse Processes*.
- Kurby, C. A., Magliano, J. P., Dandotkar, S., Woehrle, J., Gilliam, S., McNamara, D. S. (2012). Changing how student process and comprehend texts with computer-based self-explanation training. *Journal of Educational Computing Research*, *47*, 429-459.
- Lee, C. D., (2006). *Culture, literacy, and learning*. New York, NY: Teachers College Press.
- Magliano, J. P. (1999). Revealing inference processes during text comprehension. In S. R. Goldman, A. C. Graesser, & P. van den Broek (Eds.), *Narrative comprehension, causality, and coherence: Essays in honor of Tom Trabasso* (pp. 55–75). Mahwah: Erlbaum.
- Magliano, J. P., Durik, A. M., & Holt, J. K. (2011). Assessing the impact of topic interest on comprehension processes. *LSAC Research Report Series* (pp 1-78). Newtown, PA: Law School Admissions Council, Inc.
- Magliano, J. P., & Graesser, A. C. (2012). Computer-based assessment of student-constructed responses. *Behavioral Research Methods*, *44*, 608-621.
- Magliano, J.P., & Millis, K.K. (2003). Assessing reading skill with a think-aloud procedure. *Cognition and Instruction*, *21*, 251-283.
- Magliano, J. P., Millis, K.K., The RSAT Development Team1, Levinstein, I., & Boonthum, C. (2011). Assessing comprehension during reading with the reading strategy assessment Tool (RSAT). *Metacognition and Learning*, *6*, 131-154.

- Magliano, J. P., Trabasso, T., & Graesser, A. C. (1999). Strategic processes during comprehension. *Journal of Educational Psychology, 91*, 615-629.
- Martorell, P., & McFarlin, I. (2007). *Help or hindrance? The effects of college remediation on academic and labor market outcomes*. Unpublished manuscript.
- Masson, M. E. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behavior Research Methods, 43*(3), 679-690.  
<https://doi.org/10.3758/s13428-010-0049-5>
- 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. doi:10.2307/747834.
- McMaster, K. L., Van den Broek, P., Espin, C. A., White, M. J., Rapp, D. N., Kendeou, P., ... & Carlson, S. (2012). Making the right connections: Differential effects of reading intervention for subgroups of comprehenders. *Learning and Individual Differences, 22*(1), 100-111.
- McNamara, D.S. (2004). SERT: Self-explanation reading training. *Discourse Processes, 38*, 1-30.
- McNamara, D.S., & Kintsch, W. (1996a). Learning from text: Effects of prior knowledge and text coherence. *Discourse Processes, 22*, 247-288.
- McNamara, D. S., Levinstein, I. B., & Boonthum, C. (2004). iSTART: interactive strategy training for active reading and thinking. *Behavior Research Methods, Instruments & Computers, 36*, 222–233.
- 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. 60-81). Mahwah, NJ: Lawrence Erlbaum and Associates.
- McNamara, D. S., & Magliano, J. P. (2009b). Towards a comprehensive model of comprehension. In B. Ross (Ed), *The psychology of learning and motivation*, vol. 51, (pp. 297-384). New York, NY, US: Elsevier Science.
- 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*, 251-283.
- Mokhtari, K. (Ed.). (2017). Improving reading comprehension through metacognitive reading strategies instruction. New York, NY: Rowman & Littlefield.
- Nist, S. L., & Kirby, K. (1986). Teaching comprehension and study strategies through modeling and thinking aloud. *Reading Research and Instruction*, *25*(4), 254-264.
- Pearson, P.D., Hansen, J., & Gordon, C. (1979). The effect of back- ground knowledge on young children's comprehension of explicit and implicit information. *Journal of Reading Behavior*, *11*(3), 201–209.
- Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Hillsdale: Erlbaum.
- Rouet, J.F. (2006). *The skills of document use: From text comprehension to web-based learning*. Mahwah, NJ: Erlbaum.
- 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*(4), 441-474.

- Seifert, C. M., Dyer, M. G., & Black, John, B. (1986). Thematic knowledge in story understanding. *Text*, 6, 393-426.
- Shanahan, T. & Shanahan, C. (2008). Teaching disciplinary literacy to adolescents: Rethinking content-area literacy. *Harvard Educational Review*, 78(1), 40-59.
- Shanahan, C., Shanahan, T. & Misischia, C. (2008). Analysis of Expert Readers in Three Disciplines. *Journal of Literacy Research*, 43, 393-429.
- Simpson, M. L., Stahl, N. A., & Francis, M. A. (2004, Winter). Reading and Learning Strategies: Recommendations for the 21st Century. *Journal of Developmental Education*, 28(2), 2-4, 6, 8, 10-12, 14-15, 32.
- Singer, M. (1988) Inferences in reading comprehension. In M. Daneman, G. E., Mackinnon, & T. G. Waller (Eds.), *Reading research: Advances in theory and practice*, (pp 177-188). San Diego, CA: Academic Press.
- Snow, C. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Santa Monica, CA: RAND.
- Stanovich, K. E., & Cunningham, A. E. (1992). Studying the consequences of literacy within a literate society: The cognitive correlates of print exposure. *Memory & Cognition*, 20, 51-68.
- Stahl, N. A., Simpson, M. L., & Hayes, C. G. (1992). Ten recommendations from research for teaching high-risk college students. *Journal of Developmental Education*, 16(1), 2-4, 6, 8, 10.

- Todaro, S.A., Magliano, J.P., Millis, K.K., McNamara, D.S., & Kurby, C.A. (2008). Assessing the structure of verbal protocols. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 607-612). Austin, TX: Cognitive Science Society.
- Trabasso, T., & Magliano, J. P. (1996). Conscious understanding during text comprehension. *Discourse Processes, 21*, 255-288.
- Vandal, B. (2010). *Getting past go: Rebuilding the remedial education bridge to college success*. Retrieved from <http://www.gettingpastgo.org/docs/GPGpaper.pdf>
- van den Broek, P., Lorch, R.F., Jr., Linderholm, T., & Gustafson, M. (2001). The effects of readers' goals on the generation of inferences. *Memory & Cognition, 29* (8), 1081-1087.
- West, R. F., & Stanovich, K. E. (1991). The incidental acquisition of information from reading. *Psychological Science, 2*, 32-330.
- Wetzels, R. Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E-J. (2011). Statistical evidence in experimental psychology: An empirical comparison using 855 *t* Tests. *Perspectives on Psychological Science, 6*, 291-298.
- Wirt, J., Choy, S., Rooney, P., Provasnik, S., Sen, A., & Tobin, R. (2004). *The condition of education 2004 (NCES 2004-077)*. U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.



