A Multidimensional Investigation of Deep-level and Surface-level Processing

Daniel L. Dinsmore & Patricia A. Alexander

To cite this article: Daniel L. Dinsmore & Patricia A. Alexander (2016) A Multidimensional Investigation of Deep-level and Surface-level Processing, The Journal of Experimental Education, 84:2, 213-244, DOI: 10.1080/00220973.2014.979126

To link to this article: https://doi.org/10.1080/00220973.2014.979126

Published online: 12 Mar 2015.

Article views: 574

View related articles

View Crossmark data

Citing articles: 11 View citing articles
A Multidimensional Investigation of Deep-level and Surface-level Processing

Daniel L. Dinsmore
University of North Florida

Patricia A. Alexander
University of Maryland, College Park

This study examines the moderating effects of a situational factor (i.e., text type) and an individual factor (i.e., subject-matter knowledge) on the relation between depth of processing and performance. One-hundred and fifty-one undergraduates completed measures of subject-matter knowledge, read either an expository or persuasive text about the existence of extraterrestrials while thinking aloud, and then completed a passage recall task and an open-ended task. Results indicated that the relation between depth of processing and the open-ended tasks was moderated by the type of text participants read (i.e., expository or persuasive). Moreover, there was a significant interaction between the passage recall measure and open-ended task for depth of processing and type of text.

Keywords: deep- and surface-level processing, expository text, open-ended task, persuasive text, recall, strategies

IN THE PSYCHOLOGICAL LITERATURE, there has been a contention that deep processing (i.e., intentional and meaningful learning of the material; Marton & Saljo, 1976a) such as making connections to one’s prior knowledge while reading, should lead to more desirable outcomes, whereas surface or shallow processing (i.e., cognitive processes aimed at rote memorization; Marton & Saljo, 1976b), such as underlining text, should lead to less desirable outcomes (e.g., Chamorro-Premuzic & Furnham, 2008; Phan, 2008). Furthermore, there have been implicit assumptions within the literature (e.g., Biggs, 1978; Marton & Saljo, 1976a) that individuals’ levels of processing (i.e., deep versus surface) remain relatively stable across situations. However,
these relations have often been weakly correlated in empirical studies (e.g., Cano, 2007; Dinsmore & Alexander, 2012).

Such assumptions—that deeper levels of processing should lead to better academic outcomes and that an individual’s depth of processing is stable across tasks—are problematic for several reasons. For one, they do not take into account differences the characteristics of tasks that may instigate more deep- or surface-level processing. For another, these assumptions do not acknowledge variability in learner characteristics (e.g., prior knowledge) that might affect processing during task performance. As far back as 1988, Alexander and Judy demonstrated through the empirical literature that knowledge and strategic processing were inextricably intertwined and that the occasion for and nature of strategic processing cannot be addressed in isolation.

Moreover, certain models of expertise development (e.g., the Model of Domain Learning; Alexander, 1997) contend that the level of processing shifts as one moves from acclimation (i.e., the beginning stages of domain expertise) to proficiency (i.e., the end stages of domain expertise). The Model of Domain Learning predicts that surface-level strategies (i.e., those that pertain to initially encoding the problem at hand) are necessarily invoked more often in acclimation and diminish as one moves through competence and toward proficiency. The opposite pattern is predicted for deep-level processing strategies (i.e., those that entail probing or transforming a given problem) that are increasingly evidenced in competence and expertise, but relatively rare when individuals are confronted with an unfamiliar task in an unfamiliar domain.

However, this interplay between deep-processing and surface-level strategies is not meant to privilege one form over the other. Deep-processing and surface-level strategies are evident at all stages of domain expertise and their interplay depends in part of the specific nature of the task and the purpose for which that task is engaged (Alexander, 2003). In other words, the most advantageous level of processing for two individuals may not be or perhaps should not be the same but should depend on their stage of development (i.e., acclimation, competence, or expertise) and their performance goals. This view directly counters the notion that deep processing will always be advantageous and that the use of surface strategies will inevitably be disadvantageous.

The purpose of the present study, therefore, was to put common assumptions about and deep-processing and surface-level strategies to the test by investigating how levels of processing relate to performance outcomes by taking into account the nature of the learner (i.e., prior knowledge), the task (i.e., expository vs. persuasive text), and the learning outcome (i.e., passage recall vs. open-ended task). To guide this investigation, two theoretical frames were used: the Model of Domain Learning, and the Construction-Integration Model.

The Model of Domain Learning

The Model of Domain Learning (Alexander, 1997, 2004) serves as a useful framework to guide selection of variables that constitute important individual characteristics and ensuing hypotheses, specifically the role of knowledge and strategic processing. In terms of knowledge, the Model of Domain Learning describes the role of domain and topic knowledge. Domain knowledge refers to the breadth and scope of subject-matter knowledge, whereas topic knowledge refers to the depth of knowledge about specific concepts related to a domain (Alexander, Murphy, Woods, Duhon, & Parker, 1997). Strategic processing is likewise composed of two types, surface- and deep-level strategies. Surface-level processing is the use of strategic and monitoring behavior related to the basic encoding of textual content (i.e., ascertaining the nature of the problem at
hand). By contrast, deep-level processing is the use of strategic and monitoring behavior that involve a more extensive manipulation or transformation of a task or text (Alexander et al., 2004; Nolen & Haladyna, 1990).

Knowledge and strategic processes are hypothesized to interact with each other as individuals develop from acclimation to expertise. As one develops in an academic domain, one is likely to encounter increasingly complex tasks (Alexander, 2004). These more complex problems and tasks require the coordination of knowledge and strategies that go beyond simple heuristics (Chi, Glaser, & Rees, 1981). For example, when learning from text, simpler tasks may consist of answering a set of recall questions (i.e., the ability to reproduce the text verbatim, in paraphrase, or by summarizing it; Kintsch, 1998). This task would require limited coordination of prior knowledge and may use simple heuristics, such as a search and destroy tactic to match words in the question with sentences in a passage to find the correct answer (Pearson, 1978). However, more open-ended outcomes (i.e., outcomes in which there may be many possible goal states and many possible solution path; Frederickson, 1990; Simon, 1978) would require a greater coordination of prior knowledge in an effort to interpret and make inferences from text, rather than simply recall it (McNamara & Kintsch, 1996).

The Construction-Integration Model

Since learning in the present study uses a reading task, we need to position hypotheses regarding levels of strategic processing within a text comprehension framework. The Construction-Integration Model (Kintsch, 1988, 2004; Kintsch & Welsch, 1991), one of the most accepted models of reading comprehension, is helpful in differentiating different processes that occur while reading.

In particular, the Construction-Integration Model describes processes aimed at the text base and those aimed at the situation model. The text base, defined as the semantic representation of a text, includes the microstructure (i.e., propositions in the text that represent meaning) and the macrostructure (i.e., global organization of the microstructure in the text; Kintsch, 2004). Kintsch referred to the mental representations a reader forms from the text base as “surface-level memory” (2004, p. 1273). By comparison, the situation model represents the integration of the text with a reader’s prior knowledge. By virtue of the fact that the construction of the situation models requires the coordination of a readers’ prior knowledge, it can be presumed these tasks would require deeper strategic processing, particularly when asked to complete more complex tasks after reading (e.g., evaluating an argument in text) versus tasks that only accessed surface-level memory (i.e., recall questions).

Furthermore, models of comprehension like the Construction-Integration Model acknowledge the role of text genre and complexity in the nature of comprehension processing that ensue. The construction-integration model (and models like it) has been used to examine expository (e.g., Graesser & Bertus, 1998; Wolfe, 2005) and persuasive (i.e., refutational) texts (e.g., Kendeou & van den Broek, 2007; Mikkilä-Erdmann, 2001). Persuasive text is defined as text in which an author argues a point of view in order to change a reader’s knowledge, beliefs, or interest (Kamalski, Sanders, & Lent, 2008; Murphy, Long, Holleran, & Esterly, 2003). Expository text is defined as nonfiction reading material in which the intent is to inform or explain (Williams, Stafford, Lauer, Hall, & Pollini, 2009).
Measuring Deep and Surface Processing

Deep processing and surface processing have been measured in a variety of ways (Dinsmore & Alexander, 2012), including self-reports and retrospective interviews in the extant literature. For this investigation, processing was measured using think-aloud protocol. Think-aloud protocols (e.g., Ericsson, 2006) and more specifically verbal protocols of reading (e.g., Pressley & Afflerbach, 1995) elicit cognitive and metacognitive processes as reading transpires. Researchers concerned with strategic processing must find techniques to externalize individuals’ thoughts, given that such processing generally occurs covertly. Verbal protocols of reading, one such technique, elicits utterances from readers about what they are thinking or doing in the moment—utterances that can be quite revealing about strategic processing (Ericsson, 2006).

Thus, a review of this issue was helpful in selecting the most appropriate measure given the theoretical frameworks used in this study. Most of the studies reviewed by Dinsmore and Alexander (2012), particularly those that aligned with approaches to learning, used questionnaires about the intention to use deep or surface processing before the task. However, because online processing was of interest here, concurrent verbal protocols of reading align more closely to the definition proffered in that depth of processing represents deep- and surface-level strategies used during a task, providing some construct validity to this investigation. In this instance, concurrent verbal protocols of reading are also more desirable than strategic checklists, questionnaires, or interviews conducted subsequent to a task because these measures and measurements rely on individuals’ memory of their task performance, which may not reflect what they actually did (Ericsson, 2006). In Pressley and Afflerbach’s (1995) view, the advantages of verbal protocols—real-time indicators of cognitive processes, access to reasoning processes, and analysis of affective processes of reading—more than compensate for any disadvantages.

Research Questions and Hypotheses

Three research questions guided the present investigation. First, does the type of text (i.e., expository vs. two-sided refutation) moderate the relation between levels of processing (deep and surface) and reading outcomes (i.e., recall and open-ended responses) during the reading of science-related texts? It was hypothesized that participants reading the two-sided refutational text would engage in more deep-level strategies than those reading the expository text and that a higher proportion of deep-level strategies would result in higher scores on the open-ended task. These differences were expected due to previous interaction effects in the extant literature (Kamalski et al., 2008), different amounts of working memory requirements (Kellogg, 2001), and differences on multiple-choice outcome scores (Carrell & Connor, 1991). Specifically, two-sided refutational text has been found to be more effective at changing knowledge and beliefs (Buehl, Alexander, Murphy, & Sperl, 2001; Murphy et al., 2003), as well as changes in metacognitive processing (Dinsmore et al., in press).

Second, do individual characteristics (i.e., subject-matter knowledge) moderate the relation between levels of processing (deep and surface) and reading outcomes (i.e., recall and open-ended responses) during the reading of science-related texts? It was hypothesized that increased levels of subject-matter knowledge would moderate the relation between levels of processing and reading outcomes. Specifically, the relation between the strategies reported (i.e., deep or surface) and reading outcomes was predicted to be greater for those with higher levels of subject-matter
knowledge than those participants with lower levels of subject-matter knowledge. This hypothesis is supported by previous investigations of the Model of Domain Learning (e.g., Alexander et al., 2004; Murphy & Alexander, 2002).

Third, does the relation between depth of processing and learning outcomes differ between a passage-recall task and an open-ended task? It was hypothesized that participants demonstrating a greater use of deep processing (i.e., increased use of deep-processing strategies) would have higher scores on passage recall and open-ended tasks, but that this difference in performance would be greater for the open-ended task. The larger disparity for the open-ended task was hypothesized since differences between forced choice and open-ended task performance have been found in regards to search strategy, where open-ended problems require complex, expansive, and multidisciplinary knowledge of a field (Laxman, 2010). In addition, it has been found that students engage in more peer learning strategies and critical thinking during open-ended tasks (Lodewyk, Winne, & Jamieson-Noel, 2009).

**METHOD**

**Participants**

The participants for this study consisted of 151 (109 female) undergraduates from a large mid-Atlantic university. The sample consisted of 27 freshmen, 14 sophomores, 59 juniors, 45 seniors, and 6 participants who reported more than four years of undergraduate education with an average age of 20.55 ($SD = 2.38$). These undergraduates were ethnically diverse (61.6% White, 18.5% Asian, 12.6% Black, 2% Hispanic, .7% American Indian, and 4.6% other) and reported a wide variety of academic majors. These majors consisted of primarily physical and life sciences and the social sciences (34.4% and 57.0%, respectively). Students reported their grade point average, with a mean of 3.31 ($SD = .42$) for this sample. Participants were recruited from numerous courses on campus. All participants consented to participate in the study and all participants but one consented to be audiotaped. Participants were either offered extra course credit or compensated US$10 for their participation.

**Materials**

The materials for this study consisted of two text passages (Appendix A); an expository text meant to inform and a two-sided refutational text meant to persuade. The passages were adapted from Stephen Webb’s (2002) book, *Where is Everybody? Fifty Solutions to the Fermi Paradox and the Problem of Extraterrestrial Life*. The topic of extraterrestrial life was selected because it was likely that participants would at the very least have some folk, if not scientific, knowledge about the topic. Moreover, it seems probable that participants would demonstrate at least a moderate level of situational interest for the topic, where situational interest can be described a momentary affective reaction to conditions within the immediate context (Hidi & Renninger, 2006).

Both texts addressed the possibility of the existence of intelligent extraterrestrial civilizations (ETCs) in the galaxy. An initial segment of text was adapted by the first author from Webb’s text and adjusted by the first author, in consultation with experts in the field of astrobiology to create one expository and one refutational text. The main difference between the expository and
refutational texts was the inclusion of the persuasive content in the two-sided refutational text. Specifically, the two-sided refutational text contained 10 sentences that we identified as intending to persuade the reader into adopting the view that ETCs do not exist and refute the other two views (i.e., that ETCs exist and are here, or that ETCs exist but have not yet communicated). For example, in support of the notion that species might extinguish themselves through nuclear war or another catastrophe, the refutational passage stated that “the rate at which human activity is wiping out other species supports the idea that we may indeed extinguish ourselves before we are able to communicate with other ETCs.”

To create a more purely expository version of this passage, while retaining comparability in length and difficulty, we deleted those 10 particularly persuasive sentences and replaced them with text that elaborated stated ideas in nonargumentative ways. For example, the basic passage included the sentence, “Perhaps ETCs are using other types of signals such as gravitational signals, particle signals like neutrinos, or hypothetical tachyon signals that we may not be able to detect or interpret.” To this we added the following sentence: “Tachyons are any theoretical particles that have an imaginary mass.” The added sentence gave an additional detail about tachyons that did not seek to refute or persuade. Table 1 presents data that demonstrate that passage equivalence.

Reading experts (n = 3) and content experts (astronomy and physics; n = 2) were asked to judge the equivalence and suitability of the passages and their contents. These external evaluators determined that (a) the texts were equivalent in regard to length and difficulty; (b) the premise in the expository text was a description of three views about the Fermi paradox; (c) there was an argument in the two-sided refutational text that the view that ETCs do not exist is the most likely view; and (d) the passages were accurate with a minor modification about the number of planets in other solar systems discovered due to the continuing new discoveries by the Kepler spacecraft.

Measures

The measures for the study consisted of a demographics measure; a subject-matter knowledge test; a prior beliefs about intelligent ETC measure (which is not included in the present study; a postpassage measure beliefs about intelligent ETC measure (also not included in this study), and a passage recall measure.

Demographics

The demographics measure consisted of questions pertaining to participants’ gender, age, ethnicity, year in school (e.g., freshman), grade point average, and academic major.
**Subject-Matter Knowledge**

The subject-matter knowledge measure was designed to quantify participants’ domain and topic knowledge in astronomy and astrobiology, respectively. This measure consisted of 16 total multiple-choice items, 10 multiple-choice items measuring their domain knowledge of astronomy (i.e., their breadth of knowledge about astronomy), and 6 multiple-choice items measuring their topic knowledge of extraterrestrial life (i.e., their depth of knowledge about the study of the origin, evolution, distribution, and future of life in the universe; NASA, 2010).

This measure was constructed by sampling concepts from textbooks about astronomy and astrobiology (Jones, 2004; Lunine, 2005; Shostak & Barnett, 2003). Domain knowledge items developed for inclusion in the measure were concepts about astronomy common across the astronomy textbooks. These items measured the breadth of individuals’ knowledge about astronomy. Topic knowledge items developed for inclusion in the measure were concepts common across chapters in the astronomy textbooks about astrobiology and concepts found in astrobiology textbooks. These items measured the depth of individuals’ knowledge about astrobiology (a topic within the domain of astronomy).

The response model for the multiple-choice items in this measure was a graduated response model. Graduated response models have the advantage of being able to uncover patterns in individuals’ incorrect responses (e.g., Alexander, Murphy, & Kulikowich, 1998; Kelderman, 1996). Because growth in domain and topic knowledge might be fragmented, resulting in partial knowledge about topics (Fox, Dinsmore, Maggioni, & Alexander, 2008), measurements that are sensitive to this fragmentary or partial knowledge are valuable. A graduated response model was used in order to discriminate between individuals’ levels of knowledge about the targeted domain or topic (i.e., astronomy and astrobiology). The answer choices and scoring system allowed an examination of individuals’ full or partial knowledge about each item. Furthermore, the specific level designation was confirmed by the same one expert in the content domain (i.e., astronomy) and two assessment experts. A sample domain knowledge item with the associated points presented in parentheses was:

The lowest-energy photons of all electromagnetic radiation are:
   a. Radio waves (4)—domain correct (astronomy)
   b. Gamma rays (2)—domain incorrect (astronomy)
   c. Tidal waves (1)—nonastronomy science incorrect (oceanography)
   d. Stun rays (0)—misconception or popular lore (science fiction)

Answer choice (a) was the correct response and was the response choice most likely to be selected by someone with high levels of domain knowledge. Answer choice (b) was an incorrect response but was still within the domain of astronomy (they are higher energy photons than radio waves) and was likely to be selected by someone with at least some domain knowledge of astronomy. Answer choice (c) was an incorrect response from a different domain of science, in this case from oceanography. Answer choice (d) was an incorrect response and was either a common misconception held by someone who was a layperson in the domain or popular lore. In the example, answer choice (d) was from science fiction. All items on this measure followed the same distracter model.

The six topic knowledge items were in a similar graduated-response format. A topic correct response was scored a four, a topic incorrect response was scored a two, a nonastronomy science
incorrect response was scored a one, and a common misconception or popular lore response was scored a zero. All item responses for both the domain knowledge and topic knowledge questions were randomly ordered.

Content validity was established for this measure using the four steps suggested by Crocker and Algina (1986). First, the performance domains (i.e., astronomy and astrobiology) were defined. Second, items were sent to experts ($n = 3$) in the domains of astronomy, physics, and biochemistry. Third, these experts were asked to (a) make sure each item matched the intended domain or topic (i.e., astronomy or astrobiology), and (b) judge each response in terms of the response model (i.e., domain correct, domain incorrect, nonastronomy science incorrect, or common misconception/popular lore). Fourth, these data were collected and either items that did not match the performance domain or responses that did not match the response model were revised or deleted. Specifically, three items were reworded and six answer choices were reworded or changed.

To correct for measurement error in these 16 items ($\alpha = .73$), we subjected items to an exploratory factor analysis and a factor score was calculated for each individual. The exploratory factor analysis indicated a one-factor solution that explained 22.00% of the total variance. These factor scores have a mean of 0 and a standard deviation of 1.

**Passage Recall Measure**

The passage recall measure was a 10-item multiple-choice measure assessing participants’ recall of the text passages. Like the domain and topic knowledge measures, the passage recall measure also used a graduated response model. The levels of the graduated response model for the passage recall questions were as follows: correct response from the passage (scored 4); incorrect response from the passage (scored 2); incorrect response from the domain of astronomy but not from the passage (scored 1); and, an incorrect response not from the domain of astronomy (scored 0). The following is a sample item from this measure:

The Voyager probe took 21 years to reach:
- a. Pluto (4)—correct response from the passage
- b. Proxima Centauri (2)—incorrect response from the passage
- c. Saturn (1)—incorrect response from the domain of astronomy
- d. Cato Neimoidia (0)—incorrect response not from the domain of astronomy

The passage recall measure followed the same validity procedure (content validity) as the domain and topic knowledge measures. All items were judged by the experts ($n = 3$) to match the topic of the passage. The alpha coefficient was low for the passage recall measure ($\alpha = .53$). However, because these items were used as indicators for a latent variable in the resulting model, the structural relations in the model were corrected for any measurement error in the original indicators (Byrne, 1994).

**Measurements**

In addition to the aforementioned measures, think-aloud data and open-ended passage outcome data were collected then quantified using the described coding schemes.
Think-Aloud Data

While reading the text passage, participants were asked to think aloud. These think-aloud verbalizations were recorded using an Olympus WS-110 digital voice recorder. The recordings were then transcribed into text documents by the first author.

To code participants’ utterances, we used an initial set of 30 codes developed in preceding studies (e.g., Fox, Dinsmore, & Alexander, 2007) and based on Pressley and Afflerbach’s (1995) summative overview of reading behaviors. Specifically, utterances were coded as either a surface-processing strategy (S; i.e., a strategy aimed at encoding information), a deep-processing strategy (D; i.e., a strategy aimed at integrating prior knowledge), or as an other type of strategy (O; e.g., a regulatory strategy). Furthermore, if the surface-processing or deep-processing strategies did not appear to be appropriate for comprehension of the passage, a minus sign was included with the code (i.e., an S– or D–). Because the prevailing assumption in the literature is that deep processing is more helpful than surface processing is, we wanted to understand when deep and surface processing may or may not be helpful. Investigating how often strategies were used that were not appropriate was another way to examine this relation between processing and outcomes. Our previous experiences with think-aloud data presented many cases where deep-level strategies were used, but they were unrelated to the topic of the text.

Given that the appropriateness of each strategy was exploratory, care was taken to be conservative in assigning a minus to a strategy only when it was readily apparent it was not an appropriate strategic move in comprehending the text. For example, if a deep-processing strategy such as an elaboration was completely irrelevant to the text, it was coded as a D–. If a surface-processing strategy, such as a local restatement was incorrect, it was coded an S–. The list of codes and examples of coded utterances from this study appear in Table 2.

To build inter-rater agreement, the think-alouds were coded independently by the first author and another trained rater until acceptable agreement was reached. For 25 of the 150 participants (16.67%) the kappa coefficient for the codes consisting (409 total utterances) was .85. This agreement was considered substantial and the remaining transcripts were coded by the first author. All differences in codes were resolved by discussion before the remaining transcripts were coded.

Level of processing was calculated by dividing an individual’s total number of deep-processing strategies by the sum of his or her deep- and surface-processing strategies. The resulting proportion for depth of processing scores ranged from .00 (no deep-processing strategies) to 1.00 (all deep processing strategies; \(M = .61, SD = .28\)). The distribution for these scores by percentage is presented in Figure 1. The average frequency was 15.58 (\(SD = 12.79\)) for surface-level and 11.14 (\(SD = 12.80\)) for deep-level strategies. The appropriateness of strategies was calculated by dividing the number of inappropriate strategies (surface and deep) over the total number of strategies (surface and deep). The average frequency for appropriately used deep-level strategies was 12.50 (\(SD = 11.41\)); appropriately used deep-level strategies was 3.08 (\(SD = 3.63\)); appropriately used surface-level strategies was 9.92 (\(SD = 12.28\)); and, appropriately used surface-level strategies was 1.22 (\(SD = 1.88\)).
### TABLE 2
Coding Scheme for the Verbal Protocol Analysis

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Surface level</strong></td>
<td></td>
</tr>
<tr>
<td>Reading aloud</td>
<td>Overpopulation destroys intelligent life before it becomes able to colonize intelligent planets.</td>
</tr>
<tr>
<td>Rereading</td>
<td>Okay, rereading about the Fermi paradox.</td>
</tr>
<tr>
<td>Skimming (skipping portions)</td>
<td>Now I’m skimming ahead.</td>
</tr>
<tr>
<td>Guessing the meaning of a word in context</td>
<td>I’m guessing from this sentence that interstellar means between stars.</td>
</tr>
<tr>
<td>Predicting (about the micro- or macrostructure)</td>
<td>Okay, here we go, this paragraph addresses why extraterrestrials would be observing Earth.</td>
</tr>
<tr>
<td>Underlining or marking text</td>
<td>And I’m underlining everything that seems important so it’s better retained.</td>
</tr>
<tr>
<td>Using a text feature</td>
<td>Um, looks like an equation.</td>
</tr>
<tr>
<td>Rehearsing (repeating)</td>
<td>I’m going to say this a couple times in my head.</td>
</tr>
<tr>
<td>Local restatement (paraphrasing text at the microstructure level)</td>
<td>Okay, looks like we’re looking at different views of extraterrestrials.</td>
</tr>
<tr>
<td>Global restatement (paraphrasing text at the macrostructure level)</td>
<td>Okay, so this is suggesting that there are extraterrestrials and they are trying to communicate to us but we either aren’t listening or we haven’t figured out how to listen.</td>
</tr>
<tr>
<td>Making connections to prior text</td>
<td>Oh, huh, they’re explaining it [the Drake equation].</td>
</tr>
<tr>
<td>Making connections to the research task (a prior or subsequent measure)</td>
<td>I think Panspermia was a response for a multiple choice question in the study.</td>
</tr>
<tr>
<td>Evaluating text quality of the micro- or macrostructure</td>
<td>Huh, they did not consistently put quotes around this phrase.</td>
</tr>
<tr>
<td>Evaluation of the importance of text (micro-or macrostructure)</td>
<td>Um, I don’t know the point of this equation.</td>
</tr>
<tr>
<td>Evaluation of task difficulty</td>
<td>This equation looks crazy.</td>
</tr>
<tr>
<td><strong>Deep level</strong></td>
<td></td>
</tr>
<tr>
<td>Predicting (about the argument)</td>
<td>Who knows if there is any unintelligent life on other planets.</td>
</tr>
<tr>
<td>Questioning</td>
<td>Okay, just because numerous studies of genetics demonstrate that all life on Earth is highly related, some use as evidence that our DNA came from a single origin outside of Earth. Why couldn’t that single origin be within Earth? It seems plausible.</td>
</tr>
<tr>
<td>Arguing with text</td>
<td>I feel the zoo scenario is the same thing we are doing, by looking for extraterrestrial communication.</td>
</tr>
<tr>
<td><strong>Global restatement (aimed at argument)</strong></td>
<td>So this is saying that it is more likely that we are alone in the galaxy.</td>
</tr>
<tr>
<td>Making connections to background knowledge</td>
<td>I remember hearing about the Doppler effect before.</td>
</tr>
<tr>
<td>Making connections to personal experience</td>
<td>I’ve never seen evidence of ETCs anywhere on this Earth.</td>
</tr>
<tr>
<td>Interpreting (reasoning beyond information in the text base to integrate into situation model)</td>
<td>I feel the zoo scenario is the same thing we are doing, by looking for extraterrestrial communication.</td>
</tr>
<tr>
<td>Elaborating (reasoning beyond information in the text base to build meaning tangential to the text)</td>
<td>The Earth is precariously placed as far as distance from the sun so it makes sense that would be really rare, extremely rare, that there would be another planet like Earth that could sustain life. At another star.</td>
</tr>
<tr>
<td>Evaluating agreement with the text</td>
<td>I also agree that there aren’t ETCs.</td>
</tr>
<tr>
<td>Evaluating text quality about the argument</td>
<td>I wish these people would provide proof of where they’ve seen these extraterrestrial sightings.</td>
</tr>
</tbody>
</table>

(Continued on next page)
## Table 2
Coding Scheme for the Verbal Protocol Analysis (Continued)

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regulatory and other</td>
<td></td>
</tr>
<tr>
<td>Evaluating comprehension</td>
<td>Not sure what they mean by the great silence.</td>
</tr>
<tr>
<td>Moving on</td>
<td>Okay, moving on.</td>
</tr>
<tr>
<td>Evaluation of interest</td>
<td>Interesting the 98.4% of human DNA is similar to that of ape DNA.</td>
</tr>
<tr>
<td>Expression of empathy</td>
<td>I feel sorry for any aliens that try to come here.</td>
</tr>
<tr>
<td>Expression of surprise</td>
<td>Wow, we share 50% of our DNA with bananas!</td>
</tr>
<tr>
<td>Expression of amusement</td>
<td>We share half our genes with a banana. That’s funny.</td>
</tr>
<tr>
<td>Adjusting reading rate</td>
<td>Okay, I’m going to go faster.</td>
</tr>
<tr>
<td>No code (not enough information to categorize)</td>
<td>Turning the page.</td>
</tr>
</tbody>
</table>

### Open-Ended Response

Participants were asked to respond to an open-ended question about the text passage. This question immediately followed a posttest administration of their beliefs about ETCs in which question one asked about their beliefs about the existence of ETCs. The open-ended question asked participants to “justify your answer to Question 1 based on evidence from the passage or from your background knowledge.”

Written responses were typed and saved as an electronic document. Participants’ responses to the open-ended question were quantified using the Structured Outcome of the Learning Observation taxonomy (Biggs & Collis, 1982). The taxonomy evaluates the structure of the participant responses on the basis of capacity (working memory and attention span), relating operations (how

![FIGURE 1](chart.png)

**FIGURE 1** Distribution of level of processing percentages.
### TABLE 3
SOLO Taxonomy for the Problem-Solving Task Outcome

<table>
<thead>
<tr>
<th>Taxonomy level</th>
<th>Score</th>
<th>Response characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prestructural</td>
<td>0</td>
<td>Cue and response undifferentiated&lt;br&gt;No logical interrelation for cue and response&lt;br&gt;High closure or low consistency&lt;br&gt;Cue linked with irrelevant feature(s)</td>
</tr>
<tr>
<td>Unidimensional</td>
<td>1</td>
<td>Relate question with one piece of relevant data with a logical operation&lt;br&gt;Drawing a conclusion from a particular instance&lt;br&gt;Responses equally correct but inconsistent with each other&lt;br&gt;One relevant feature to link question and response</td>
</tr>
<tr>
<td>Multistructural</td>
<td>2</td>
<td>Two or more relevant concepts or data&lt;br&gt;Uses several features but does not link them&lt;br&gt;Closure but lack of consistency&lt;br&gt;Several relevant features link question and response</td>
</tr>
<tr>
<td>Relational</td>
<td>3</td>
<td>Response which interrelates multiple concepts&lt;br&gt;Overall concept or principle accounting for data presented&lt;br&gt;Waits for all aspects before interrelating to make coherent whole&lt;br&gt;Definite overgeneralized answer tied to concrete experience&lt;br&gt;Uses relevant data in a conceptual scheme</td>
</tr>
<tr>
<td>Extended abstract</td>
<td>4</td>
<td>Give information comprehended in relevance to an overriding abstract principle&lt;br&gt;True logical deduction&lt;br&gt;Heavily qualifies set out principle to application in given situations&lt;br&gt;Question left relatively open&lt;br&gt;Relevant data with interrelations under hypothetical abstract structure with alternative outcomes and no definite closure</td>
</tr>
<tr>
<td>Transitional responses</td>
<td>.5; 1.5; 2.5; or 3.5</td>
<td>At a level of the taxonomy but marked by confusion or inconsistency&lt;br&gt;Handles more information than able to cope with&lt;br&gt;Loses track of the argument&lt;br&gt;Forced to give up before reaching next SOLO level</td>
</tr>
</tbody>
</table>

*Note. Data from Bigg and Collis (1982). SOLO = Structured Outcome of the Learning Observation.*

The scoring rubric also allowed for transitional responses (i.e., responses that fall between levels of the taxonomy), resulting in a scale that ranged from zero (prestructural) to four (extended abstract) with the possibility of a response being between two of the levels (e.g., a 3.5 if the response fell between relational and extended abstract). The resulting distribution of scores on the open-ended response ($M = 1.57, SD = 1.06$) is presented in Figure 2.

To build interrater agreement for the open-ended responses, the following procedure was undertaken. First, the first author and another rater scored five random open-ended responses together. A score and rationale for the score was tabled. Next, each rater independently scored another five random open-ended responses marking a score and a rationale for the score. The two raters compared the scores and rationale for the scores. Differences were rectified in conference. Raters then independently scored another ten responses. For the 15 responses independently coded...
(10.0% of the total responses), the kappa coefficient for the exact ratings was .69. Furthermore, the average deviation of ratings for both raters for these 15 responses was .03 points. This kappa coefficient was considered good according to Fleiss’s (1981) guidelines and the average deviation of the rater score were considered well within tolerable limits. All differences in codes were resolved by discussion before the remaining transcripts were coded. The remaining transcripts were coded by the first author.

Procedure

Participants completed the measures and measurements in a research laboratory at the university. Participants first read and signed the consent form. Participants then completed the demographics measure, the subject-matter knowledge measure, and beliefs about ETCs measure.

Next, participants were given a practice passage to read while thinking aloud. The practice passage was about mosquitoes and was adapted from a popularly written science article by Marston Bates (1975). Once participants were comfortable with the think-aloud protocol, they read either the expository or persuasive text. The text condition (i.e., expository or persuasive) was randomly assigned to obtain roughly the same number of participants in each condition ($n_{expository} = 76$, $n_{persuasive} = 75$). The specific directions for the think-aloud protocol were as follows:

In this investigation, we are interested in what you think and do while you read a text. What we want you to do is say what you are thinking and doing out loud. You can decide for yourself whether you would like to read the text silently or out loud, or do some of both. Do whatever feels most natural to you. We are only interested in what you are thinking and doing as you read. For example, if you are going back to reread, please say that’s what you are doing. If something in the text reminds you of prior experiences or things you already know, please say that. If you are thinking that you don’t understand something, please say that, too. There are no right or wrong things to say here, just
whatever is going through your head as you read. If you are quiet for a period of time, I’ll ask you to say what you’re thinking. Do you have any questions?

After reading the passage and thinking aloud participants again completed the passage recall measure, the beliefs about ETCs questionnaire, and the open-ended response. During passage recall, participants did not have access to the text, but did have access to the text during the open-ended response. The average participant took approximately 30 min to complete these tasks.

Overview of Analyses

Structural equation modeling was used to provide the statistical tests for these three questions. Question 1 was a test for the moderation of type of text on the relations between depth of processing and the two outcome measures (i.e., the passage recall task and the open-ended task; \( H_0: b_{F1V3} = 0 \) and \( H_0: b_{V7V3} = 0 \); see Figure 3). The parameters of the model tested for this question were the direct effects of the interaction of depth of processing and type of text on the passage recall task scores and open-ended task scores. Specifically, the null hypothesis was that the direct effects of this interaction on each of the outcomes in the model were 0.

Question 2 was a test for the moderation of subject-matter knowledge on the relations between depth of processing and the two outcome measures (i.e., the passage recall task and the open-ended task; \( H_0: b_{F1V5} = 0 \) and \( H_0: b_{V7V5} = 0 \); see Figure 3). The parameters of model tested for this question were the direct effects of the interaction of depth of processing and subject-matter
knowledge on the passage recall task scores and open-ended task scores. Specifically, the null hypothesis was that the direct effects of this interaction on each of the outcomes in the model were 0.

Question 3 was a test of the difference between three pairs of paths. The first of these pairs were the paths between depth of processing on the passage recall task and depth of processing on the open-ended task (coefficients $b_{F1V1}$ and $b_{V7V1}$ in Figure 3, respectively). The second of these pairs were the paths between the interaction of depth of processing and type of text on the passage recall task and open-ended task, respectively (coefficients $b_{F1V3}$ and $b_{V7V3}$ in Figure 3, respectively). The third of these pairs were the paths between the interaction of depth of processing and subject-matter knowledge on the passage recall task and open-ended task, respectively (coefficients $b_{F1V5}$ and $b_{V7V5}$ in Figure 3). Specifically, the null hypotheses for these tests was that the estimated parameters for these paths were equal ($H_0: b_{F1V1} = b_{V7V1}$, $H_0: b_{F1V3} = b_{V7V3}$). The only difference in the analysis for Question 3 was that the passage recall task and open-ended task were standardized (i.e., transformed to have a mean of zero and standard deviation of 1) to allow for a direct comparison of these paths using a t test.

Before conducting the study, an a priori power analysis (i.e., sample size determination) was conducted for testing parameters within a model, according to the procedure outlined by Hancock (2006). The first step was to decide on the desired level of power ($\pi = .80$) and alpha level ($\alpha = .05$) for the focal parameter tests (i.e., $H_0: b_{F1V1} \neq b_{V7V1}$, $H_0: b_{F1V2} \neq b_{V7V2}$; $H_0: b_{F1V5} \neq b_{V7V5}$; $H_0: b_{F1V3} \neq 0$, $H_0: b_{V7V3} \neq 0$, $H_0: b_{F1V5} \neq 0$, and $H_0: b_{V7V5} \neq 0$), which corresponded to a noncentrality parameter of $\lambda = 7.85$. Next, numerical values were selected for all parameters (standardized) of the full model as well as for the standardized focal parameters (Table 4).

Using EQS 6.1 (Bentler, 2006) moment information was determined for population. This information was used to determine population data-model fit for a reduced model. An arbitrary sample size was set and EQS was again used to determine a model fit function value for the reduced model. Necessary sample sizes were derived for each of the focal parameters of interest. The largest of the sample sizes needed for testing $H_0: b_{F1V1} = b_{V7V1}$ with $\pi = .80$ was $n = 150$. The sample size for testing $H_0: b_{F1V3} = b_{V7V3}$ and $H_0: b_{F1V5} = b_{V7V5}$ was deemed too large for the scope of this study, so while it will still be tested, it should be noted that these were not tested with $\pi = .80$.

To evaluate the hypotheses in questions one and two, the measurement equations and standard errors were evaluated as part of the EQS output. For these questions, the $z$ value reported in EQS was evaluated at $\alpha = .05$. For Question 3, the differences in the path values were divided by the standard error of the difference (i.e., the square root of the sums of each parameters standard error minus two times the covariance of each pairs parameter estimate) to obtain a t value that was also evaluated at $\alpha = .05$. To make the paths comparable to the outcome variables, each outcome variable was standardized before making these comparisons (i.e., they both had means of zero and standard deviations of 1). Factor scores were computed using principle components analysis.

Following data collection and coding, a structural equation model was fitted using EQS 6.1 (Bentler, 2006). According to the procedure suggested by Aiken and West (1991), exogenous variables were first centered then multiplied to derive each of the product terms using the centered exogenous variables (i.e., depth of processing, type of text, and subject-matter knowledge). The measurement model was fit first with error terms for the passage recall measures left uncorrelated. The measurement model demonstrated satisfactory fit and any attempts to constrain error terms in
TABLE 4
Standardized Focal and Peripheral Parameters Values for the A Priori Power Analysis

<table>
<thead>
<tr>
<th>Parameter estimate</th>
<th>Standardized coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Focal parameters</strong></td>
<td></td>
</tr>
<tr>
<td>$c_1$</td>
<td>.20</td>
</tr>
<tr>
<td>$c_2$</td>
<td>.20</td>
</tr>
<tr>
<td>$e_1$</td>
<td>.20</td>
</tr>
<tr>
<td>$e_2$</td>
<td>.20</td>
</tr>
<tr>
<td>$a_2-a_1$</td>
<td>.15</td>
</tr>
<tr>
<td>$c_2-c_1$</td>
<td>.15</td>
</tr>
<tr>
<td>$e_2-e_1$</td>
<td>.15</td>
</tr>
<tr>
<td><strong>Peripheral parameters</strong></td>
<td></td>
</tr>
<tr>
<td>$a_1$</td>
<td>.20</td>
</tr>
<tr>
<td>$a_2$</td>
<td>.20</td>
</tr>
<tr>
<td>$b_1$</td>
<td>.30</td>
</tr>
<tr>
<td>$b_2$</td>
<td>.30</td>
</tr>
<tr>
<td>$d_1$</td>
<td>.40</td>
</tr>
<tr>
<td>$d_2$</td>
<td>.40</td>
</tr>
<tr>
<td>$f_1$</td>
<td>.10</td>
</tr>
<tr>
<td>$f_2$</td>
<td>.10</td>
</tr>
<tr>
<td>$V_1V_2$</td>
<td>.20</td>
</tr>
<tr>
<td>$V_1V_3$</td>
<td>.20</td>
</tr>
<tr>
<td>$V_2V$</td>
<td>.00</td>
</tr>
</tbody>
</table>

*Note. All other peripheral parameters were estimated to be .01.*

the measurement model lead to convergence issues. Standardized loadings for the passage recall factor in the measurement model ranged from .17 to .52. Construct reliability for the factor was calculated using coefficient H (Hancock & Mueller, 2001). Coefficient H for this factor was .62 which indicates moderate construct reliability.

The structural model was then fitted using full-information maximum likelihood (FIML; five participants’ depth of processing and one participant’s open-ended response were missing). Correlations between exogenous variables were estimated along with the covariance between the error and disturbance of the endogenous variables.

To evaluate the hypotheses in questions one and two, the measurement equations and standard errors were evaluated as part of the EQS output. For these questions, the z value reported in EQS was evaluated at $\alpha = .05$. For Question 3, the differences in the path values were divided by the standard error of the difference (i.e., the square root of the sums of each parameters standard error) to obtain a t value that was also evaluated at $\alpha = .05$.

RESULTS AND DISCUSSION

Table 5 presents the means, standard deviations, skewness, and kurtosis for each of the variables before factor analysis. Table 6 presents the correlation matrix for the means of the measured variables before factor analysis.
TABLE 5
Correlations, Means, Standard Deviations, Skewness, and Kurtosis for the Measured Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>SMK</th>
<th>DoP</th>
<th>PR</th>
<th>OER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject-matter knowledge</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Depth of processing</td>
<td>.17*</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Passage recall</td>
<td>.39**</td>
<td>.09</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Open-ended response</td>
<td>.47**</td>
<td>.17*</td>
<td>.28**</td>
<td>—</td>
</tr>
<tr>
<td>Mean</td>
<td>39.69</td>
<td>0.61</td>
<td>31.80</td>
<td>1.57</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>9.69</td>
<td>0.28</td>
<td>4.95</td>
<td>1.06</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.21</td>
<td>1.00</td>
<td>−0.60</td>
<td>0.27</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>−0.67</td>
<td>0.08</td>
<td>0.36</td>
<td>−0.72</td>
</tr>
</tbody>
</table>

Note. * = significant at $\alpha < .05$; ** = significant at $\alpha < .01$.

Preliminary Data Analyses

The purpose of the preliminary data analyses was to ensure that the resulting data were appropriate for the structural equation modeling analysis. In regards to normality, these data met Finney and DiStefano’s (2006) suggestion that multivariate kurtosis (Mardia-based kappa) should be less than three ($\kappa = 0.16$). In addition, raw data of the cases with the largest contribution to normalized multivariate kurtosis were examined and not found to problematic. As suggested by Bentler (2006), the distribution of standardized residuals were symmetric and centered closely around zero. Table 7 presents the model covariance matrix for the measured and latent variables along with the means for each of the measured variables. In addition, overall model fit was good according to the joint criteria suggested by Hu and Bentler (1999): $\text{CFI} = .99$, $\text{SRMR} = .065$, and $\text{RMSEA} = .048$ (90% CI [.023, .068]). These values indicate that the suggested model is one plausible model for these data. The standardized path coefficients are displayed in Figure 4 with significant paths bolded.

Interaction of Depth of Processing and Type of Text on Performance

The first research question addressed how the relation between depth of processing and performance (a passage recall task and an open-ended task) would be moderated by type of text (i.e.,
### TABLE 7
Model Covariance and Mean Matrix for Measured and Latent Variables

<table>
<thead>
<tr>
<th></th>
<th>DoP</th>
<th>ToT</th>
<th>DoPX Tot</th>
<th>SMK</th>
<th>DoPXMK</th>
<th>TW</th>
<th>OET</th>
<th>PR1</th>
<th>PR2</th>
<th>PR3</th>
<th>PR4</th>
<th>PR5</th>
<th>PR6</th>
<th>PR7</th>
<th>PR8</th>
<th>PR9</th>
<th>PR10</th>
<th>PRT</th>
</tr>
</thead>
<tbody>
<tr>
<td>DoP</td>
<td>.076</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tot</td>
<td>.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DoPX Tot</td>
<td>.002</td>
<td>.000</td>
<td>.019</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMK</td>
<td>.047</td>
<td>−.027</td>
<td>.009</td>
<td>.993</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DoPXMK</td>
<td>.000</td>
<td>.008</td>
<td>−.005</td>
<td>.034</td>
<td>.076</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TW</td>
<td>−.004</td>
<td>.011</td>
<td>.000</td>
<td>−.009</td>
<td>.002</td>
<td>.019</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OET</td>
<td>.047</td>
<td>.012</td>
<td>−.018</td>
<td>.512</td>
<td>.062</td>
<td>.003</td>
<td>1.117</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PR1</td>
<td>.007</td>
<td>−.005</td>
<td>.003</td>
<td>.102</td>
<td>.000</td>
<td>−.004</td>
<td>.066</td>
<td>.400</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PR2</td>
<td>.013</td>
<td>−.011</td>
<td>.006</td>
<td>.205</td>
<td>.000</td>
<td>−.008</td>
<td>.134</td>
<td>.068</td>
<td>1.057</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PR3</td>
<td>.009</td>
<td>−.007</td>
<td>.004</td>
<td>.141</td>
<td>.000</td>
<td>−.005</td>
<td>.092</td>
<td>.047</td>
<td>.094</td>
<td>.293</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PR4</td>
<td>.012</td>
<td>−.010</td>
<td>.006</td>
<td>.190</td>
<td>.000</td>
<td>−.007</td>
<td>.124</td>
<td>.063</td>
<td>.127</td>
<td>.087</td>
<td>1.361</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PR5</td>
<td>.018</td>
<td>−.014</td>
<td>.008</td>
<td>.276</td>
<td>.000</td>
<td>−.011</td>
<td>.180</td>
<td>.091</td>
<td>.184</td>
<td>.127</td>
<td>.171</td>
<td>1.824</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PR6</td>
<td>.025</td>
<td>−.020</td>
<td>.011</td>
<td>.383</td>
<td>.000</td>
<td>−.015</td>
<td>.250</td>
<td>.127</td>
<td>.256</td>
<td>.176</td>
<td>.237</td>
<td>.345</td>
<td>1.805</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PR7</td>
<td>.008</td>
<td>−.006</td>
<td>.004</td>
<td>.121</td>
<td>.000</td>
<td>−.005</td>
<td>.079</td>
<td>.040</td>
<td>.081</td>
<td>.056</td>
<td>.075</td>
<td>.109</td>
<td>.151</td>
<td>1.683</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PR8</td>
<td>.012</td>
<td>−.009</td>
<td>.005</td>
<td>.183</td>
<td>.000</td>
<td>−.007</td>
<td>.119</td>
<td>.061</td>
<td>.122</td>
<td>.084</td>
<td>.113</td>
<td>.165</td>
<td>.229</td>
<td>.072</td>
<td>1.262</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PR9</td>
<td>.015</td>
<td>−.012</td>
<td>.007</td>
<td>.237</td>
<td>.000</td>
<td>−.009</td>
<td>.155</td>
<td>.079</td>
<td>.159</td>
<td>.109</td>
<td>.147</td>
<td>.214</td>
<td>.297</td>
<td>.094</td>
<td>.142</td>
<td>.848</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PR10</td>
<td>.013</td>
<td>−.010</td>
<td>.006</td>
<td>.196</td>
<td>.000</td>
<td>−.008</td>
<td>.128</td>
<td>.065</td>
<td>.131</td>
<td>.090</td>
<td>.122</td>
<td>.177</td>
<td>.246</td>
<td>.078</td>
<td>.117</td>
<td>.152</td>
<td>2.151</td>
<td></td>
</tr>
<tr>
<td>PRT</td>
<td>.007</td>
<td>−.005</td>
<td>.003</td>
<td>.102</td>
<td>.000</td>
<td>−.004</td>
<td>.066</td>
<td>.034</td>
<td>.068</td>
<td>.047</td>
<td>.063</td>
<td>.091</td>
<td>.127</td>
<td>.040</td>
<td>.061</td>
<td>.079</td>
<td>.065</td>
<td>.034</td>
</tr>
<tr>
<td>Means</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.044</td>
<td>.009</td>
<td>.000</td>
<td>3.77</td>
<td>3.52</td>
<td>3.84</td>
<td>3.17</td>
<td>2.87</td>
<td>3.13</td>
<td>2.14</td>
<td>3.13</td>
<td>3.66</td>
<td>2.58</td>
</tr>
</tbody>
</table>

*Note.* DoP = Depth of Processing; Tot = type of text; DoP X Tot = interaction of depth of processing and type of text; SMK = subject-matter knowledge; DoP X SMK = interaction of depth of processing and subject-matter knowledge; TW = three way interaction of depth of processing, type of text, and subject-matter knowledge; OET = open-ended task; PR = passage recall items; PRT = passage recall task factor.
expository vs. persuasive). For the passage recall task and open-ended task, this was done by testing the null hypotheses $H_0: b_{F1V3} = 0$ and $H_0: b_{V7V3} = 0$ (respectively) from Figure 3. This was a test to determine whether the direct effects of the interaction of depth of processing and type of text on each of the outcomes were significantly different than zero. Results from the path estimates in the model (all of which are included in Figure 4) revealed that there was a significant interaction of type of text on the relation between depth of processing and the open-ended task ($b_{V7V3} = -1.13, p = .039$), but not for the passage recall task ($b_{F1V3} = .11, p = .42$). In other words, the direct effect of the interaction between depth of processing and type of text on the open-ended task was significantly different than zero (in the negative direction), while the direct effect of the interaction between depth of processing and type of text on the passage recall tasks was not significantly different than zero.

The significant path between the interaction of depth of processing and type of text with the open-ended responses indicates that those individuals with shallower processing (lower depth of processing percentages; DoP) had higher scores on the open-ended responses in the persuasive condition than those in the expository condition. However, individuals with deeper processing (higher DoP percentages) had higher open-ended responses in the expository condition than the persuasive condition. This interaction is displayed graphically in Figure 5. The regression equations for the simple slopes were calculated (Cohen, Cohen, West, & Aiken, 2003) for the two text conditions: $z_{exp}: \hat{Y} = .92x - .035$ and $z_{per}: \hat{Y} = -.21x + .035$. Furthermore, the actual slopes (i.e., those calculated using the scores from both the expository and persuasive conditions)
was significantly different than zero for the expository text condition ($b = 1.17$, $t = 2.60$, $p = .011$), but not for the persuasive condition ($b = .13$, $t = .29$, $p = .77$). In other words, the effect was significant for depth of processing on the expository text, but not for the persuasive text. This may indicate that overall depth of processing appears to be more influential (as indicated by the larger slope) for the expository text than the persuasive text.

One possible explanation for this interaction effect deals with the explicitness of the argument. For the expository condition, there was no explicit argument as to which of the three explanations for ETCs was most scientifically testable. Therefore, for some of the participants deeper processing was needed to formulate a position. Thus, it may have prompted the engagement in deeper levels of processing as a means to justify their position in the open-ended task. However, for some in the persuasive condition, deeper levels of processing were not necessary since the argument was given. However, it is likely that these relations would also be influenced by characteristics of the individuals discussed later.

A second explanation centers on the appropriateness of the strategies used (i.e., deep or surface). As part of the coding scheme inappropriate strategies were coded with an S– (inappropriate surface-level strategy) or a D– (inappropriate deep-level strategy). Although, there was good interrater agreement for these codes, how appropriate a strategy was for a participant was difficult to determine upon examining the transcripts. To this end, both raters were conservative in this coding and caution should be observed when interpreting these data. However, the correlation between the appropriateness of the strategies used (proportion of D– and S– to the total number of strategies) and the open-ended responses was significant ($r = .27$, $t = 2.36$, $df = 144$, $p = .020$). It is possible that the explicit argument in the persuasive text prompted participants to engage in less facilitative deep-processing strategies—a situation not evidenced for the expository text.
Interaction of Depth of Processing and Subject-Matter Knowledge on Performance

The second research question addressed how the relation between depth of processing and performance (a passage recall task and an open-ended task) would be moderated by subject-matter knowledge. For the passage recall task and open-ended tasks, this was done by testing the hypotheses $H_0: \beta_1 = 0$ and $H_0: \beta_2 = 0$ (respectively) from Figure 3. This was a test if the direct effects of the interaction of depth of processing and subject-matter knowledge on each of the outcomes were significantly different than zero. Results from the path estimates in the model revealed no significant interactions of subject-matter knowledge on the relation between depth of processing and the open ended tasks ($p_{TV5} = .503$, $p = .067$) or the passage recall task ($p_{FIV5} = -.032$, $p = .62$). In other words, the direct effect of the interaction between depth of processing and type of text on both outcomes were not significantly different than zero.

The lack of significance between the path testing the interaction of depth of processing and subject-matter knowledge to the open-ended response was surprising. This could potentially be a power issue, because the parameter estimates used in the a priori power analyses were somewhat different the actual parameter estimates in the model. For example, the peripheral parameter estimates for depth of processing to the passage recall task and for depth of processing to the open-ended response were thought to be about .20. However, the actual model parameters fell far below those at .019 and .093, respectively. It may be the case that a larger sample size would be necessary for path significance, rather than there not being a significant relation in the population.

Another possible explanation for the nonsignificant finding may be attributed to error in the measured and latent variables used in the analysis. The relatively low variance explained by the one factor, prior topic knowledge (22%), may indicate that the items did not represent the content well. However, given that item development entailed expert validation of item content, another more plausible explanation could be related to the fragmentary knowledge of the participants. Lower reliabilities of prior knowledge measures (and in this case the lower amount of variance accounted for in the factor) are more typical in studies examining individuals in acclimation. Such as condition is ascribed to the generally limited and fragmented knowledge acclimating learners possess (Alexander, 1997, 2003). This condition stands in contrast to the knowledge base of more competent or expert participants whose domain or topic knowledge tends to be quantitatively richer and qualitatively more cohesive (Alexander 2003; Alexander, Kulikowich, & Schulze, 1994).

The failure to reject the null hypothesis (i.e., that there was no interaction effect between depth of processing and subject-matter knowledge on the open-ended task) in the typical null hypothesis testing framework precludes generalization of these findings. However, the close proximity of the test statistic to the region of rejection warrants further examination. In other words, retaining the null hypothesis does not rule out that there are interaction effects, rather it signifies an inability to confidently state that the effect is significantly different than zero in a population according to a chosen decision rule ($\alpha = .05$). In the ensuing discussion of the depth of processing and subject-matter interaction, caution should be exercised in interpreting these effects beyond this particular sample.

To meaningfully interpret the interactions of two continuous variables, Cohen and colleagues (2003) suggested examining the regression of the criterion (scores on the open-ended task) on
one predictor (depth of processing) at each of several values of the other predictor (subject-matter knowledge). The regression line of the criterion on one of the predictors (depth of processing) at one value of the other predictor (subject-matter knowledge) is termed a simple regression line. In accordance with Cohen and colleagues’ (2003) suggestion, the three values chosen for subject-matter knowledge in calculating the simple slopes were a subject-matter knowledge value at the mean (labeled \( z_{\text{mean}} \)), a subject-matter knowledge value one standard deviation above the mean (labeled \( z_{\text{high}} \)), and a subject-matter knowledge value one standard deviation below the mean (labeled \( z_{\text{low}} \)). Because the factor scores for subject-matter knowledge had a mean of zero and a standard deviation of 1, these three values for subject-matter knowledge were 0, 1, and -1, respectively. The equations for these simple slopes were as follows: \( z_{\text{low}}: \hat{Y} = -.15x - .070; \) \( z_{\text{mean}}: \hat{Y} = .36x; \) and, \( z_{\text{high}}: \hat{Y} = .86x + .070 \) and are graphically displayed in Figure 6.

While parallel lines would indicate no interaction, the pattern of the slopes in these simple regression lines indicates an interaction effect for this sample (again, this should not be generalized beyond this sample). The positive interaction term in this sample indicates that higher proportions of deep-processing strategies more often lead to higher scores on the open-ended response for those with higher and mean amounts of subject-matter knowledge. However for those with lower values of subject-matter knowledge, a higher proportion of deeper-processing strategies more often lead to lower scores on the open-ended response.

Similar to the proposed interaction for type of text, the appropriateness of the strategies used may be one of the reasons for the presence of the subject-matter knowledge and depth of processing interaction for this sample. The correlation between the proportion of inappropriate strategies (i.e., S- and D- divided by the total number of strategies) and subject-matter knowledge was significant \( (r = .39, t = 5.11, df = 144, p < .001) \). This strong relation lends credence to the
argument that, for this sample, deep processing may only have been effective if one possessed the requisite knowledge to deploy the strategic process in an appropriate way. This conclusion is consistent with the relations between strategic processing and knowledge forwarded by Alexander and Judy (1988).

Differences Between the Passage Recall Task and Open-Ended Task

The third question investigated differences in the paths between three of the variables of interest (i.e., depth of processing, the interaction of depth of processing and type of text, and the interaction of depth of processing and subject-matter knowledge) and the two outcomes (i.e., passage recall task and the open-ended task). An initial examination of these paths revealed two significant paths (from the interaction of type of text and depth of processing to the open-ended task and the path from the interaction of type of text and depth of processing to the open-ended task). The other paths (the interaction of depth of processing and subject-matter knowledge to both outcomes and depth of processing to both outcomes) were nonsignificant and should be investigated further (given the nature of the effects). Since the scale of outcome variables (the open-ended and passage recall tasks) could not be taken as equivalent, due to the problems with comparisons of unstandardized paths (Kwan & Chan, 2010), Hotchkiss (1976) suggested standardizing each of the outcome variables for comparison. After both outcome variables were standardized, a comparison of the path between the interaction of type of text and depth of processing to passage recall and the interaction of type of text and depth of processing to the open-ended task were significant using this method ($t = 2.15$, $df = 141$, $p = .033$).

Furthermore, in this sample the standardized effects were larger in general for the open-ended response than for the passage recall task (except for the subject-matter knowledge and the three way interaction in the model). Comparing these values should be interpreted with caution as these paths did not attain significance at the .05 level. The differences in these effects, particularly the significant depth of processing and type of text interaction between these outcomes, may be due to the demands of these two tasks and the participants’ familiarity with these types of tasks. First, the passage recall task and open-ended response place different cognitive demands on participants. While the passage recall task only requires surface-level memory (e.g., Kintsch, 2004) to reproduce a correct answer, the open-ended tasks requires justification of a position, a more higher-order task (e.g., Alexander, Dinsmore, Parkinson, & Winters, 2011; Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956). Unlike mere reproduction of information relying on surface-level memory, the justification task would likely require readers to construct their own situation model during and after reading. In other words, the open-ended task is a production task rather than a reproduction task.

Second, with regard to the participants’ familiarity with the tasks, the passage recall task is a more typical school-type task. Passage recall is not only commonplace in the classroom, but is a typical task for the standardized testing to which most of these individuals have been exposed. However, given that justification of knowledge is not as commonplace in schools, it could be that this task was something unfamiliar or novel to many. These two issues (the task demands and familiarity of the participants with them) could have contributed to the significant interaction of depth of processing and type of text as well the significant difference in the two outcomes.
Conclusions and Implications

This study was designed to test the interactions of type of text and subject-matter knowledge on the relation between depth of processing and two outcomes, a passage recall task and an open-ended task. Overall, the findings of this study support the contention that a consideration of the relation between levels of processing and performance without a consideration of moderating factors may be contributing to the mixed findings in the literature. There were significant differences in the interaction of depth of processing and type of text on the open-ended tasks, as well as differences in this interaction between the open-ended task and the passage recall task. Although there was no conclusive evidence of a significant interaction between depth of processing and subject-matter knowledge on the open-ended response, this should be further investigated.

Overall, the results of this investigation provide some evidence that the relation between depth of processing and performance is conditional. The evidence in regard to the situation (i.e., type of text) and the target of learning (i.e., the passage recall and open-ended tasks) having a conditional effect was statistically significant. However, the evidence in regard to individual differences (i.e., subject-matter knowledge) should be further investigated before drawing more definitive conclusions. Despite the nonsignificant finding in regard to subject-matter knowledge, it is clear that considering depth of processing and performance in isolation may hamper attempts to fully understand this complex relation.

Thus, these data contradict the prevailing assumption in the literature that deep processing leads to better outcomes. Depending on the nature of the learner, nature of the text, or how the outcome is measured, surface processing, or a mix of deep and surface-level strategies, may be more advantageous for some readers. This may be dependent on characteristics of the learner, such as their level of knowledge related to reading (e.g., Perfetti, Beck, Bell, & Hughes, 1987) or the topic of the passage (e.g., Afflerbach, 1990). Furthermore, differences in processing and outcomes will be dependent on the nature of the text presented (e.g., Kendeou & van den Broek, 2007), even when the difficulty of the passages is similar. The task in which learners are asked to demonstrate their comprehension influences the relation between processing and performance.

Given these findings, this study adds to the extant literature in two substantive ways. First, this study investigated not only the effects that the learner, text, and task characteristics have on reading outcomes, but it also investigated how these characteristics influenced the nature of the relation between levels of processing and outcomes. Second, because of the evidence of interactions among these characteristics, levels of processing, and outcomes, this study highlights the need to examine these constructs in a multidimensional way. A unidimensional examination of these constructs without investigating relevant interactions may lead to false assumptions. For example, higher topic knowledge may only lead to better reading outcomes if the assessment requires the learning to use their knowledge in employing deeper-levels strategies. If the “search and destroy tactic” is sufficient to answer reading outcomes questions, deeper-level strategies are not necessary.

Limitations of the Present Study

Although the proposed study forwarded the literature on deep and surface processing, there were a few limitations present. First, as with any measure, the concurrent think-aloud protocol was likely to elicit only a certain range of processes (Pressley & Afflerbach, 1995). It is probable
that processes that are overly complex or overly simple (skillful reading) were not reported by participants. Therefore, it is likely that the think-aloud protocol elicited a range of strategic processes but certainly did not capture all strategic processes or skillful reading. Specifically, the number of utterances may not necessarily correspond to the number of strategies a participant uses. They may be more likely to report deep strategies (because they have better access).

Second, processing in this investigation was measured for one instance with one passage. Certainly there would be an expectation that one’s depth of processing may differ from situation to situation depending on the properties of task at hand. Last, the learning outcome, as with any experimental task was necessarily constrained beyond that of what individuals might do in their everyday lives. For example, individuals did not have access to materials that they might normally use to complete a task (e.g., other resource materials such as textbooks). Here, they were purposefully being limited to their prior knowledge and the text they read in their given text condition.

Future Directions for Research

The conclusions of the present study warrant further investigation of deep and surface processing in two areas: measurement and the modeling of the relations between processing and performance.

Measuring deep and surface processing. Four recommendations for measuring deep and surface processing will be discussed including: measurement of depth of processing with multiple indicators; the need for multitrait multimethod studies; and, the avoidance of dichotomization of deep and surface processing. First, in regard to the measurement of depth of processing, it may be the case that a single indicator (e.g., the proportion of deep- vs. surface-processing strategies) may not be adequate to represent the construct of depth of processing. It is possible that strategies aimed at the text base (surface-processing strategies) and those aimed at the situation model (deep-processing strategies) may not be adequate indicators of depth of processing. Other indicators, such as monitoring and control processing, may also play an integral part in the relation between levels of processing outcomes when included in a latent construct regarding processing.

In addition to more fully representing the construct of depth of processing, including multiple indicators of depth of processing may also provide some convergent validity evidence (i.e., whether theoretically related indicators are actually related to each other in practice; Campbell & Fiske, 1959) for depth of processing. One preliminary study to provide convergent validity evidence would be to design and carry out a multitrait multimethod (Campbell & Fiske, 1959). Multitrait multimethod designs work by measuring several traits with several different measures (or methods). By examining the resulting correlation matrix, convergent validity can be determined by examining the validity diagonal in each method block (i.e., the measures should correlate highly with each other for the same trait) and divergent validity can be determined by examining the heterotrait-monomethod triangles (i.e., the measures should not correlate highly with each other for different traits) and the heterotrait-heteromethod triangles (i.e., different measures with different traits should also not correlate highly with one another). In addition to measured variable multitrait multimethod designs first described by Campbell and Fiske (1959), latent variable multitrait multimethod designs are also possible (e.g., Marsh, 1989).

Third, dichotomization of depth of processing (i.e., designating an individual as a deep processor or a surface processor) should likely be avoided. To legitimately dichotomize depth of processing, one would have to conceptually claim that individuals use primarily deep-processing
strategies or primarily surface-processing strategies. This runs counter to the models presented here (the Construction-Integration Model and the Model of Domain Learning). In addition, evidence from this study (see Figure 1 in particular) suggests that there were many participants whose proportion of deep-level strategies was neither near 0% or 100%. Therefore, the evidence presented here suggests that the dichotomization of depth of processing is both theoretically and empirically unjustified.

Fourth, the data regarding the appropriateness of strategies in this investigation was exploratory. Further investigation into the appropriateness of strategy use should be undertaken with an emphasis on multiple viewpoints on whether or not a strategy is appropriate or useful. Specifically, viewpoints should include the participants themselves to help coders make determinations of appropriateness or inappropriateness of strategy use. Preliminary findings have indicated that participants themselves do not differentiate between the usefulness of deep- and surface-level strategies (but indicate monitoring and evaluative strategies as much less useful; Dinsmore et al., in press).

The issues of measuring, identifying, and categorizing deep- and surface-level strategies are difficult ones. In future research, careful consideration needs to be given to how these categories are developed and used. The categorization scheme used in this study is different from other conceptual schemes that have examined deep and surface processing. This study used the Construction-Integration Model to identify deep versus surface processes, whereas studies such as Vermetten, Vermunt, and Lodewijks (2002) study used approaches to learning (Vermunt, 1998) to identify deep versus surface approaches to learning. Thus, as the studies of strategic processing continue to mature, it will be increasingly important for the underlying conceptualization and ensuing categorization of strategies to be taken into account when trying to compare the level or depth of students’ strategic engagement across these studies.

In addition, for studies using think-aloud methods, the issue of grain size also needs to be further addressed. The present study examined deep- and surface-level processing at a larger grain size by categorizing processes as either deep or surface level. Using a smaller grain size (e.g., looking at the relation of individual strategies such as inferring) may elucidate more intricate relations between depth of processing and outcomes than that afforded by larger grain size analyses. This issue is currently being debated in the self-regulated learning literature, where coding schemes at the micro and macro levels are being compared (e.g., Greene & Azevedo, 2009). Consequently, for future research on deep- and surface-level processing, it may be beneficial to examine relations at different grain sizes to see if similar patterns emerge.

**Modeling Deep and Surface Processing**

Beyond further refinements to the measures and measurements of deep and surface processing, statistically modeling the relation between depth of processing and performance also needs to be a continuing consideration. These modeling considerations comprise: including other pertinent individual and situational variables; the latent modeling of these variables; and inclusion of the usefulness of strategies used.

First, while the present study provided evidence that an additive model (i.e., one without interactions) is likely not sufficient to model the relation between deep and surface processing, much remains to be seen as to what other individual or situational variables may also play a key role in the relation between depth of processing and performance. In related areas of investigation
(e.g., metacognition and self regulation) other variables that have played key roles in theoretical models include interest and epistemic beliefs. Interest (i.e., the feeling- and value-related attributes toward an object; Schiefele, 1999), which is also one of the forces considered in the Model of Domain Learning, may also potentially moderate the relation between depth of processing and performance. Epistemic beliefs, while not a force modeled in the Model of Domain Learning, have been theorized to play a key role in self-regulated learning and other theoretical frameworks (e.g., Muis & Franco, 2009). In addition, situational variables discussed previously, such as text coherence and text difficulty, should also be systematically investigated for interaction effects in future models.

Second, dealing with interactions in the relation between depth of processing and performance (e.g., interaction effects and multiple outcomes) brings with it many analytic challenges. Casual modeling (i.e., structural equation modeling) can be a helpful approach in this endeavor for a variety of reasons. These types of models can address “theory-driven causal research questions” at the latent variable level (Hancock & Mueller, 2006, p. 6). In addition, the latent variable approach is designed to estimate the effects of latent constructs (e.g., depth of processing), rather than a single indicator (e.g., proportion of deep-processing strategies; Cohen et al., 2003; Pedhazur, 1997) and measurement errors in the model can be somewhat accounted for (Byrne, 1994). Using a multivariate framework, these models are flexible in that they can consider multiple outcomes (Pedhazur, 1997) and are capable of estimating the interactions of latent variables (e.g., Marsh et al., 2007). Last, these models move away from simple rule based procedures to model building that relies on a scientific epistemology (Rogers, 2010). For example, in the present study the decision to retain the null hypothesis in terms of the subject-matter and depth of processing interaction is rather difficult to make, because the p value is so close to the cutoff value (i.e., $\alpha = .05$). Although these models typically require large sample sizes, the benefits of such models make the effort worthwhile.

One particularly helpful outcome of the present study is that parameter estimates from the current model can assist with future modeling. Specifically, the current model may serve another indicator of parameter estimates for a priori power analyses, whether for testing particular parameters as in this study or for model fit as a whole. However, as was the issue in this case, focal and peripheral parameters values were chosen from studies using different conceptualizations and operationalizations of deep and surface processing, so if different measures are used or different variables are modeled, these estimates will need to be continually revised.

The parameter estimates included in this study, however, may be more difficult to replicate—particularly those related to the open-ended task. Specifically, the relation between the interaction of subject-matter knowledge and depth of processing and learning outcomes may be less stable when tasks (and measures such as subject-matter knowledge) could have multiple possible responses versus tasks with only one correct response. Each layer of construct complexity is going to introduce more error of measurement. Thus, it is important that future studies replicate these findings with a variety of tasks and measures (such as the multitrait multimethod approach suggested previously).

Third, one of the more difficult considerations in modeling the relation between deep depth of processing and performance is how the usefulness of the strategies used (whether deep or surface) plays into the overall relation between depth of processing and performance. For example, if individuals use deep-processing strategies, but these strategies are ineffective (e.g., an unhelpful elaboration about an event unrelated to the text), should these strategies really be counted as
deep-processing strategies? Although this approach sounds reasonable, preliminary attempts to ascertain whether strategies were appropriate from participants’ think-aloud utterances were difficult to make. One possible avenue to investigate this issue may rely on a multipronged approach using both think-aloud protocols and semi-structured interviews. For example, it may be helpful to use the think-aloud transcript as a way to query participants’ reasons for engaging in and their judgments of the success of a particular strategy.

It is clear that the recommendations for measurement and model building will not be addressed by a single study. These issues require a systematic program of research to investigate valid measures of depth of processing, appropriate interaction terms for the model, and performance outcomes that assess learning differently. We hope that programs of research from a variety of different theoretical positions can shed new light on how depth of processing relates to task performance.

AUTHOR NOTES

Daniel L. Dinsmore received his doctorate in Human Development with a specialization in Educational Psychology from the University of Maryland in 2011. His research interests include strategic processing and learning from text. Patricia A. Alexander is a Distinguished Professor and the Jean Mullan Professor of Literacy at the University of Maryland, College Park. Her research interests include strategic processing, learning from text, and academic development.

ACKNOWLEDGEMENTS

The authors thank Meghan Parkinson and Emily Fox for their help with coding. The authors also thank Peter Afflerbach, Ann Battle, Gregory Hancock, Kathryn Wentzel, Allan Wigfield, and the members of the Disciplined Reading and Learning Research Laboratory for their helpful comments and feedback on previous drafts of this article.

REFERENCES


