tapraid5/cae-cae/cae-cae/cae00313/cae0200d13z | xppws | S=1 | 7/31/13 | 21:42 | Art: 2012-0312

APA NLM

Canadian Journal of Experimental Psychology / Revue canadienne de psychologie expérimentale



1196-1961/13/$12.00 DOI: 10.1037/a0032918

# Comparing Comprehension Measured by Multiple-Choice and Open-Ended Questions

AQ: au
AQ: 1

Yasuhiro Ozuru
University of Alaska Anchorage

Stephen Briner
University of Illinois at Chicago

Christopher A. Kurby
Grand Valley State University

Danielle S. McNamara
Arizona State University

This study compared the nature of text comprehension as measured by multiple-choice format and open-ended format questions. Participants read a short text while explaining preselected sentences. After reading the text, participants answered open-ended and multiple-choice versions of the same questions based on their memory of the text content. The results indicated that performance on open-ended questions was correlated with the quality of self-explanations, but performance on multiple-choice questions was correlated with the level of prior knowledge related to the text. These results suggest that open-ended and multiple-choice format questions measure different aspects of comprehension processes. The results are discussed in terms of dual process theories of text comprehension.

AQ: 2

*Keywords:* reading comprehension, measurement of comprehension, learning, memory

Open-ended and multiple-choice format questions are two of the most prevalent forms of questions used to measure text comprehension and learning in educational and research settings. Although some research suggests that these two types of questions assess comprehension differently, the nature of this difference is not well understood (see Campbell, 1999, for a review). Furthermore, some researchers argue that even if there are distinctions between the processes tapped by the two forms of questions, the distinctions between them may not be that clear cut due to the multitude of factors that influence readers' answers. Questions can elicit different performance depending on various question attributes such as the type of text, question content, the answer options, and whether or not the source text is available for reference during answering (Graesser, Ozuru, & Sullins, 2010; Magliano, Millis, Ozuru, & McNamara, 2007).

Empirical investigation of the potential differences in the type of comprehension processes tapped by multiple-choice and open-ended questions is critically important for both practical and theoretical reasons because this would affect interpretations of test scores measured by these two types of questions in educational and research settings. We specifically examined the nature of comprehension processes assessed by two versions (i.e., open-ended and multiple-choice) of the same questions regarding an expository text when they were administered as memory-based comprehension questions in which questions need to be answered without looking back at the source text. In particular, we examined these differences when readers actively processed key parts of the texts by generating self-explanations while reading (e.g., Chi, De Leeuw, Chiu, & Lavancher, 1994; McNamara, 2004).

We propose a process reinstatement hypothesis based on a transfer appropriate processing perspective (Morris, Bransford, & Franks, 1977). The reinstatement hypothesis postulates that performance on memory-based open-ended and multiple-choice versions of the same questions relies on different aspects of text comprehension because the cognitive processes involved in answering these questions reinstate different types of cognitive processes engaged at the time of reading the text (Singer & Remillard, 2004). In particular, we hypothesise that these two forms of questions differ in measuring the relative contribution of active information generation and passive automatic processing that occur at the time of reading a source text.

We first discuss the processes of memory retrieval involved in answering open-ended and multiple-choice comprehension questions, and then discuss the possibility that the memory retrieval involved in answering these two types of comprehension questions may reinstate the different cognitive processes (i.e., active generative processing vs. passive automatic activation of topic-specific knowledge) that occur during the time of reading a source text.

## Memory Processes Involved in Answering Open-Ended and Multiple-Choice Reading Comprehension Questions

Answering comprehension questions after reading an expository text without looking back at the text resembles a memory task.

Yasuhiro Ozuru, Department of Psychology, University of Alaska Anchorage; Stephen Briner, Learning Sciences Research Institute, University of Illinois at Chicago; Christopher A. Kurby, Department of Psychology, Grand Valley State University; Danielle S. McNamara, Department of Psychology, Arizona State University.

AQ: 8

Correspondence concerning this article should be addressed to Yasuhiro Ozuru, Department of Psychology, University of Alaska Anchorage, 3211 Providence Drive, Anchorage, AK 99508-4614. E-mail: yozuru@uaa.alaska.edu

AQ: 9

Test takers study the text content and try to remember the relevant information based on their episodic memory of the source text and their prior knowledge of the topic. Hence, in order to perform well on this type of comprehension question, one needs to remember the relevant information using the question as a retrieval cue. However, the amount and quality of the retrieval cues available to the test takers differ greatly depending on whether the format of the question is *open-ended* or *multiple-choice*. Typically, open-ended questions resemble cued-recall tasks which provide limited information as retrieval cues (Graesser et al., 2010; Magliano et al., 2007). The retrieval of source text information from memory in a cued-recall task primarily relies on recollection or controlled AQ: 3 retrieval processes (Jacoby, 1996), involving multiple steps of goal-oriented active searching, in which more effective and a greater number of retrieval cues are generated successively by step-by-step processing based on the cue.

In contrast with answering open-ended questions, the processes involved in answering multiple-choice questions are supported in part by automatic retrieval or familiarity. This is because the target information or its variant is present as a part of the answer options (see Yonelinas, 2002 for discussions on recollection and familiarity). According to this view, performance on multiple-choice questions may be aided by perceived familiarity of the target answer because the rich cues (i.e., target answers) present in multiple-choice questions often result in immediate retrieval of the target information. This retrieval may lead to successful identification of the correct answer among the answer options without an active memory search.

Certainly, multiple-choice questions can be constructed such that readers cannot rely on familiarity alone. Specifically, one can manipulate multiple-choice questions by (a) minimising the match of surface features (i.e., orthographic features) between the target and source information in the text; and (b) maximizing the conceptual and surface feature overlap between the target and distractor options (Graesser et al., 2010; Magliano et al., 2007). Logically speaking, as the distracter options become conceptually/perceptually similar to the target through the manipulation described above, the discrimination of a target based on familiarity becomes more difficult. Under such circumstances, test takers may rely more heavily on controlled processing to retrieve more source information for further discrimination among the answer options.

Therefore, multiple-choice questions may be different from yes–no recognition such that multiple-choice questions cannot be answered solely based on familiarity. However, according to research on memory, it is still unclear whether yes–no and multiple-choice questions differ in terms of how recollection and familiarity contribute to answering processes of these two types of questions (see Bastin & Van der Linden, 2003). Whereas some researchers argue that answering forced-choice questions (e.g., multiple-choice) relies on familiarity more so than does answering yes–no questions (Aggleton & Shaw, 1996), others argue for similar contributions of familiarity across forced-choice and yes–no recognitions (Khoe, Kroll, Yonellinas, Dobbins, & Knight, 2000), at least in the context of recognition memory research.

The assumption that answering open-ended questions requires greater active generation of source-text information from memory than does answering multiple-choice questions leads to a specific hypothesis that the two types of questions are differentially sensitive in measuring different processes in which readers engage during the reading of the source text. This hypothesis follows from the assumption that answering these two types of questions reinstates different types of cognitive processes that occur during the reading of the source text.

## Automatic and Controlled Processing in Reading Comprehension and Their Reinstatement

Various theories of text comprehension agree that a substantial part of text comprehension occurs in the form of relatively automatic activation of preexisting topic-specific knowledge related to the text content, triggered by text-based input (e.g., Adams & Collins, 1979; Anderson, 1984; Kintsch, 1988, 1998; Rumelhart, 1985). Such automatic activation of topic-specific knowledge results in a coherent mental representation of the text when the text provides sufficient (but often not complete) information about given events and/or situations relative to a reader's knowledge about the text topic (Kintsch, 1988, 1998; McNamara, Kintsch, Songer, & Kintsch, 1996). Automatic processing in comprehension is assumed to take the form of the relatively passive activation of relevant information in the reader's knowledge structure based on the working strength of associative links between textual information and preexisting topic-specific knowledge in a moment-by-moment fashion (Cook, Halleran, & O'Brien, 1998).

However, this type of passive activation may not be sufficient in itself to produce a desired level of comprehension. In particular, passive activation may not be sufficient for readers to form a coherent mental representation when a text does not provide sufficient information relative to readers' knowledge, such as when reading difficult expository materials to learn new information in an educational setting. In such cases, the reader must engage in more resource-consuming follow-up processes. These follow-up processes may involve activating difficult-to-access knowledge or memory of the relevant previous sections of the text using effortful, step-by-step reasoning activities (Kintsch, 1993; Long & Lea, 2005) such as self-explaining (McNamara, 2004).

In sum, reading comprehension involves the contribution of automatic passive activation of topic-specific knowledge and more controlled, goal-oriented active generation of information (e.g., Long & Lea, 2005; van den Broek, Rapp, & Kendeou, 2005). In other words, for a given (section of) text, the relative contributions of active generative processing and passive automatic activation of topic-specific knowledge is likely to vary across readers, with some readers relying more heavily on passive automatic activation of topic-specific knowledge and other readers engaging in more controlled and generative processing.

## Relating Performance on Open-Ended and Multiple-Choice Comprehension Questions to Individual Differences Based on the Process Reinstatement Hypothesis

The contributions of automatic and controlled processing to text comprehension can be viewed as the influence of how information is encoded during reading, which can then be used to answer subsequent comprehension questions. As discussed earlier, answering open-ended questions is expected to require active generation or recollection of the information from memory of the source text, whereas answering their multiple-choice counterparts is expected to be successful based in part on familiarity. Further, note that the transfer appropriate processing perspective (Morris et al., 1977) states that memory retrieval is best when the nature of the encoding process corresponds to the nature of the retrieval process.

According to this line of reasoning, the types of comprehension processes assessed by open-ended and multiple-choice versions of comprehension questions may differ. Further, the reinstatement hypothesis predicts that performance on multiple-choice and open-ended questions will be related to specific predictor variables such as the amount of active processing (e.g., as measured by the quality of self-explanations) and the level of topic-specific knowledge that can be easily activated upon processing of the text content.

Performance on open-ended questions, which requires active generation of information, should be positively correlated with the amount of active generative processing a reader engaged at the time of reading. This is because engaging in active generative processing during text comprehension should increase the probability that retrieval processes required to successfully answer open-ended questions overlap with processes engaged during comprehension, potentially reinstating those processes. Therefore, performance on open-ended questions is expected to correlate positively with the extent of active processing a reader engages during text comprehension (i.e., self-explanations). Further, this correlation should be larger than the correlation between the level of topic-specific knowledge and performance on open-ended questions. This is because the successful generation of correct answers to open-ended questions largely depends on how actively test-takers process the text by actually generating relevant information during comprehension irrespective of the amount of preexisting topic-specific knowledge.

Now, turning to performance on multiple-choice questions, the process reinstatement hypothesis predicts that performance on multiple-choice questions should not be strongly related to the extent of active processing that occurs during text comprehension. This is based on the following assumptions: (a) multiple-choice question answering processes are aided by familiarity in addition to recollection; and (b) familiarity improves performance beyond the level supported by active recollection alone. Because a reader can be familiar with information that cannot be spontaneously generated, the contribution of active processing in multiple-choice questions is expected to be overshadowed by the contribution of familiarity.

What about the influence of preexisting topic-specific knowledge on comprehension? First, as described earlier, the literature on reading comprehension indicates that text comprehension is influenced by prior knowledge in general, irrespective of the type of assessment (e.g., McNamara et al., 1996; McNamara & Kintsch, 1996). Thus, the prior literature supports the assumption that prior knowledge should show consistent correlations with performance on both open-ended and multiple-choice questions.

However, based on the process reinstatement hypothesis, we expect that the contribution of prior knowledge will be larger for multiple-choice questions than for open-ended versions of the same questions for the following reasons. First, variability in performance on open-ended questions largely depends on the type of processing engaged during reading, irrespective of prior knowledge. As discussed earlier, this is particularly true when the material is an expository text containing novel information to be learned because test-takers might be less likely to be able to generate correct answers if they did not generate the relevant information during text processing. Second, multiple-choice questions can be answered in part based on familiarity in addition to active recollection. Because the amount of information one can recognise based on familiarity (i.e., correctly identifying a target in multiple-choice) is greater than the amount of information that one can spontaneously generate in response to a limited cue, performance variability on multiple-choice questions is expected to strongly correlate with variability in participants' topic-specific knowledge.

From the former we can expect the contribution of topic-specific knowledge to performance on open-ended questions to decrease. From the latter, we can expect the contribution of topic-specific knowledge to performance on multiple-choice questions to increase. Therefore, the amount of topic-specific knowledge should correlate more strongly with performance on multiple-choice questions than with performance on open-ended questions.

We would like to emphasise that the larger contribution of prior knowledge to performance on multiple-choice relative to open-ended questions is particularly likely when prior knowledge is measured in terms of knowledge specifically related to the text topics (topic-specific knowledge). This is because passive/automatic activation of this type of relevant knowledge underlies automatic, as opposed to more controlled generative, processing in text comprehension which, we assume, is tapped by multiple-choice questions.

Finally, provided that these two questions measure different aspects of reading comprehension processes, we expect low correlations between performance on multiple-choice and open-ended versions of the same questions. Interestingly, however, this prediction is at odds with Ozuru, Best, Bell, Witherspoon, and McNamara (2007), who reported moderate to high (0.5 to 0.8) correlations between performance on open-ended and multiple-choice versions of the same questions when they were administered as memory-based comprehension questions. In addition, Ozuru and colleagues found that both multiple-choice and open-ended questions were moderately correlated with prior knowledge. Notably, there were some limitations to the Ozuru et al. (2007) study. Most importantly, only prior knowledge was examined as a potential factor contributing to question-answering performance. There were no measures collected that might have indicated the extent to which the participants engaged in active processing during reading. This is important because evidence is mixed regarding whether readers spontaneously engage in active processing while reading text. Some evidence suggests that readers engage in extensive semantic processing by default (e.g., Cook & Myers, 2004), and others suggest that readers do not tend to engage in active processing when reading difficult texts (Noordman, Vonk, & Kempff, 1992). This raises the possibility that readers in Ozuru and colleagues' previous study did not engage in extensive active processing.

To combat the possibility that participants would resist engaging in active processing while reading, we asked participants to generate self-explanations in the current study. The self-explanations also afforded a measure of differences in active processing. Specifically, participants were asked to generate self-explanations at preselected sentences that were judged to be relevant for answering the to-be-presented comprehension questions. After reading the text, participants answered open-ended, and then, multiple-choice versions of the same nine questions based on their memory for the text. Our predictions were: (a) the degree of active processing, as measured by the quality of the self-explanations, will be more strongly correlated with performance on open-ended questions than with performance on multiple-choice question; (b) participants' level of topic-specific knowledge will be more strongly correlated with multiple-choice questions than with performance on open-ended questions; and (c) the correlation between performance on the two versions of the questions will be low.

# Method

## Participants

Forty-one undergraduate students from the University of Memphis participated in the experiment in exchange for credit in an introductory psychology course. Age and gender of the participants were not recorded. The participants were all native speakers of English.

## Design

This study was designed as a correlational study in which two different predictor variables (i.e., level of prior topic-specific knowledge and quality of self-explanation generated during the reading phase) were correlated with text comprehension as measured by two different formats (i.e., open-ended and multiple-choice) of the same questions. Test format was manipulated as a within-subjects factor with a fixed order such that each participant first answered open-ended and then multiple-choice versions of the comprehension questions. We used a within-subjects manipulation of test format because a within-subjects manipulation affords better control of the targeted individual difference factors. That is, it is challenging to match two groups of participants who have the same profiles of preexisting topic-specific knowledge as well as skill in generating self-explanations. Additionally, variability in other individual differences across the two groups renders the interpretation of findings difficult. We used fixed order administration of the open-ended and then multiple-choice format questions because providing multiple-choice questions first would make performance on open-ended questions heavily confounded with the exposure to the answer options, one of which is correct answer. As such, we believe that the design used in this study is an optimal compromise given the constraints in achieving the objective of detecting potential differences in the contribution of two different factors to performance on two different types of comprehension questions.

## Materials

**Text.** The text was a 437-word expository passage entitled *Why is There Sex?* (see Appendix A). The passage discusses the evolutionary benefits and costs of sexual reproduction compared with asexual reproduction. The text was obtained from a chapter on evolutionary psychology in the psychology textbook *Learning and Cognition* (Leahey & Harris, 1997). Local coherence of the text was reduced by removing some connectives (e.g., *for example*, *however*, *therefore*) and argument overlap between adjacent sentences to induce active processing on the readers' part when explaining the sentences (e.g., McNamara, 2004).

**Text-presentation program.** The text was presented using a computer program that displayed sentences one at a time. Participants progressed through the text by clicking the NEXT button displayed on the screen. These sentences were presented cumulatively, such that previous sentences remained on the screen as new sentences appeared. Seven target sentences, which participants were required to explain, were presented in a red coloured font. The computer program also presented an explanation box, positioned at the bottom of the screen, to type in the explanations for the "red" sentences. Participants were instructed to type their explanation for each target sentence into the explanation box when it was presented on the screen. After completing each explanation, participants were instructed to click a "continue" button and were not permitted to view or revise completed explanations.

**Testing booklet.** The booklet contained comprehension questions about the science text and also included questions about participants' preexisting topic-specific knowledge. The booklet was divided into four sections. The first section contained nine open-ended comprehension questions. The second section contained multiple-choice versions of the same nine comprehension questions. The third section contained eight open-ended prior knowledge questions. Finally, the fourth section contained eight multiple-choice prior knowledge questions. The four sections were separated by a sheet of paper that instructed participants that they could not move back to a previous section once they moved on to a new section. The characteristics of comprehension and prior knowledge questions are described in the following sections.

**Comprehension questions.** A systematic method was used to construct the questions. We constructed open-ended questions designed to tap three different levels of comprehension: three text-based, three local bridging/inference, and three global bridging/inference questions. Based on guidelines for comprehension question construction in text processing and/or education literature (e.g., Anderson, 1972; Goldman & Duran, 1988), we developed three types of comprehension questions distinguished by the cognitive demands and processes associated with information searching and integration needed to answer the questions. A question was classified as text-based if its answer could be provided using information explicitly stated within a given sentence.

A question was classified as local bridging inference if its answer required the integration of information located within adjacent sentence pairs. Finally, global bridging inference questions are similar to local bridging inference questions but involve the integration of information located across larger distances (more than two sentences apart). With respect to scoring of the open-ended questions, we used answer keys based on the correct answer options in the multiple-choice version of the same questions. The only difference in scoring between the two types of questions is that we awarded partial credit (0.5) to seven of the nine open ended questions when the answer provided by the participants sufficiently fulfilled the criteria without fully matching the answer key. This way, we equated the difficulty level of the questions at least in relation to the nature of information requested as an answer by open-ended and multiple-choice version of the questions. However, a concern remains such that different scoring (i.e., awarding partial credit for open-ended question without doing the same for multiple-choice questions) may influence the analysis. In order to

APA NLM
tapraid5/cae-cae/cae-cae/cae00313/cae0200d13z | xppws | S=1 | 7/31/13 | 21:42 | Art: 2012-0312



address this concern, we repeated all the analyses after awarding full credit (1) to all the open-ended questions that received partial credit (0.5), and confirmed that the scoring methods of open-ended questions did not influence the findings.

We constructed nine multiple-choice questions from the open-ended questions by adding four answer options. The four answer options for each question consisted of one target and three types of distractors: near-miss, thematic, and unrelated. Near-miss distractors have a large conceptual overlap with the target answer and the idea is located in the passage but in an inappropriate context. This distractor is likely to seduce readers who focus primarily on the surface features of a text such as individual words or phrases, without processing the text at a deeper level.

Thematic distractors are answers that are plausible but contain erroneous information based on common misconceptions (not located within the passage). Thus, this distractor is likely to seduce readers who read the text by imposing their own erroneous understanding of the topic.

Unrelated distractors are highly improbable, or inconsistent with the theme of the passage. Thus, the inclusion of the two key distractors (i.e., near-miss and thematic) is based on the notion that successful text comprehension requires using appropriate knowledge to understand the text content by maintaining textual coherence. In this way, we attempted to control the quality of multiple-choice versions of comprehension question. Example comprehension questions are presented in Appendix B.

Cronbach's alpha of the open-ended version of the comprehension questions was .67, and multiple-choice version of the questions was .61.

**Prior knowledge questions.** There were 16 prior knowledge questions, which examined participants' knowledge relevant to evolutionary psychology, the nature of sexual reproduction, and the immune system. Half of the prior knowledge questions were in open-ended format and the other half in multiple-choice format. Both formats were used in order to minimise the effect of question format in the measurement of the prior knowledge, to control for the retrievability (familiarity and recollection) of prior knowledge (see beginning of article), and to assess various levels of prior knowledge. Cronbach's alpha of these 16 questions was .56. The text content targeted by the questions differed across and within multiple-choice and open-ended questions. Example prior knowledge questions are presented in Appendix C.

## Procedure

Participants were tested in small groups. Participants read the text presented on a computer screen and self-explained seven preselected target sentences at their own pace. Participants were asked to read each sentence carefully. For the production of explanations, participants were instructed to explain the target sentences displayed in red font. They were asked to explain what the sentences meant to them in order to better understand the meaning of the overall text. The instructions also contained an example of a good explanation, which demonstrated how a reader integrated ideas contained in separate sentences and used prior knowledge to interpret a situation described in a text. The instructions to this point did not mention that comprehension questions would follow the reading and self-explanation activity.

Immediately after reading and self-explaining the text, participants were provided with the booklet which contained comprehension questions and prior knowledge questions. We presented the prior knowledge questions at the end of the experiment to avoid exposing participants to topic-related questions that could influence participants' reading behaviour (see McNamara & Kintsch, 1996). We acknowledge that this may have affected the results. Care was taken to ensure that the texts did not contain answers to the prior knowledge questions. Participants' responses to the open-ended questions were scored using the answer key. The responses to the open-ended questions were scored independently and then compared by two raters. Interrater reliability was 95%.

## Coding of Self-Explanation Protocols

The quality of explanations was analysed in terms of various dimensions, such as whether explanations contained paraphrases, text-based, and/or knowledge based inferences. Two coders coded the explanation protocols. First, the coders identified whether explanations contained key information relating to the target sentences (*paraphrase*) in the source text. Paraphrases were further coded as either *Accurate* or *Inaccurate*. Accuracy concerned whether the paraphrase preserved the original meaning of the target sentences.

When an explanation contained ideas that could not be traced to the target sentences, they were coded as *inferences*. Each meaningful chunk of information that could not be traced to the target sentences was coded as an individual inference. When an explanation contained an inference, the source of the inference was identified in terms of the following types of information: (a) *near-bridging*, in which an inference was based on the preceding sentence; (b) *far-bridging*, in which the inference was based on a previous section of the text which was more distant than the previous sentence; (c) *general knowledge*, in which the inference was based on common knowledge, personal experiences, or semantic association; and (d) *domain knowledge*, in which the inference was based on knowledge specific to evolution, evolutionary psychology, and other biological knowledge related to evolution, the immune system, and reproduction. Hence, an explanation could contain any combination of paraphrase, near-bridging, far-bridging, general knowledge, and/or domain knowledge based inferences. When an explanation contained any type of inference, each inference was further classified in terms of whether it was relevant; that is, whether the inference contributed to the deeper and more coherent comprehension of the target sentence. Finally, when an inference was judged to be relevant, it was classified in terms of accuracy according to whether the inference was scientifically accurate.

The two coders initially coded 10 self-explanations together for the purpose of training specific to the text, to discuss and resolve discrepancies in coding. Informal counting of the agreement at the training stage confirmed the reliability of the coding. After the training, the coders independently coded approximately 5% of the self-explanations, which were randomly selected. We selected 5% of the items for the reliability analysis because the coders were experienced users of the coding scheme, which had been used to analyse self-explanations pertaining to many different texts. Levels of agreement reported below resembled reliability scores obtained in previous self-explanation analyses in which the same coding

APA NLM
tapraid5/cae-cae/cae-cae/cae00313/cae0200d13z | xppws | S=1 | 7/31/13 | 21:42 | Art: 2012-0312



scheme was used for different texts (McNamara, O'Reilly, Best, & Ozuru, 2006). Coders' agreement with respect to whether self-explanations contained paraphrases, inferences (all types), and the accuracy of paraphrases reached 85% or above. However, agreement concerning the relevance and accuracy of inferences, in particular general knowledge based inference, remained somewhat lower, around 70%. Disagreements were resolved via discussion. After the reliability analysis, the self-explanation data were divided in half. The two coders each independently coded half the self-explanations. Coders consulted with each other whenever ambiguities or difficulties were encountered in the coding process. The analyses reported in the results are based on the final coding.

## Results

The goals of the analyses were to examine: (a) the contributions of the quality of explanations and level of prior knowledge to performance on multiple-choice and open-ended questions, and (b) relations between participants' performance on multiple-choice and open-ended questions. First, however, we present overall performance on comprehension questions separately for three types of questions (i.e., text-based, local, global inference). Table 1 presents performance on the comprehension questions as a function of format (open-ended vs. multiple-choice) and type of question (text-based, local, global). T1

As shown in Table 1, overall performance on open-ended questions ($M$ = .47, $SD$ = .21) was lower than performance on multiple-choice questions ($M$ = .69, $SD$ = .22), $t(38) = 5.1$, $p <$ .001, $d$ = .83. Also, performance on global questions ($M$ = .45, $SD$ = .15) was lower than performance on local ($M$ = .63, $SD$ = .23), $t(38) = 5.3$, $p <$ .001 $d$ = .85, and text-based ($M$ = .66, $SD$ = .22) questions, $t(38) = 5.98$, $p <$ .001, $d$ = .95. The difference between text-based questions and local questions was not statistically significant, $t(38) < 1.0$. In sum, the questions generally behaved in the predicted way, with multiple-choice questions being easier than open-ended questions and questions involving more information integration being more difficult than questions that do not require information integration. These two findings make sense because: (a) baseline performance on multiple-choice questions (25%) is higher than that of open-ended questions (0%); and (b) answering global or local questions is generally more difficult than answering text-based questions because these questions demand inferential processing that links multiple ideas in the text.

Next we present data on participants' self-explanations and performance on prior knowledge questions (see Table 2), which were then correlated with performance on comprehension questions for the main analysis. First, in order to represent the overall T2

Table 1
*Proportion Correct on Comprehension Questions as a Function of Question Difficulty and Question Format*

| | Open-ended Mean (*SD*) | Multiple-choice Mean (*SD*) |
|---|---|---|
| Text-based | .58 (.29) | .74 (.32) |
| Local | .61 (.28) | .72 (.29) |
| Global | .21 (.21) | .68 (.26) |

Table 2
*Performance on Self-Explanations and Prior Knowledge Questions*

| | Mean | *SD* | Min | Max |
|---|---|---|---|---|
| Average overall quality of self-explanation (the number of different types of explanations included) | 1.08 | 0.51 | 0.29 | 2.14 |
| Average number of inaccurate self-explanations | 0.41 | 0.26 | 0 | 1.00 |
| Proportion correct on prior knowledge questions | 0.40 | 0.16 | 0.12 | 0.84 |

self-explanation quality of each participant, we counted, for each of the seven self-explanations, how many of the five different dimensions (i.e., accurate paraphrase, accurate general knowledge based inference, accurate domain knowledge based inference, and accurate near- and far-bridging inference) were present, and then we averaged the count across the seven target sentences for which self-explanations were generated (see Appendix D for an example coding). This score represented self-explanation quality. Thus, overall self-explanation quality represents how many of the five dimensions of self-explanation were present in the person's self-explanations for a given target sentence on average. The value of overall self-explanation quality can be viewed as representing the extensiveness of relevant and accurate self-explanation generated by participants. In addition, we also computed the average number of inaccurate self-explanations included. Similar to the overall self-explanation quality measure, participants can have up to five inaccurate self-explanations (i.e., in case a participant included all five dimensions of self-explanation but all of the self-explanations were inaccurate). Table 2 indicates that on average, participants included 1.0 accurate and 0.4 inaccurate (out of 5) self-explanation dimensions per utterance (i.e., for a given target sentence), with the occurrence of inaccurate self-explanations varying from 0 (minimum) to 1 (maximum). With regard to prior knowledge, Table 2 indicates that performance on prior knowledge questions varied from .16 (minimum) to .84 (maximum). Thus, the data indicate fairly large variability in levels of prior knowledge as well as the quality of explanations that participants produced. A Pearson correlation indicated that participants' level of prior knowledge was not significantly correlated with overall self-explanation quality (−.08) or the occurrence of inaccurate self-explanations (.17), suggesting that self-explanation quality did not depend on participants' level of prior knowledge in this study.

Next, we examined whether performance on open-ended and multiple-choice questions was correlated with the quality of self-explanations or prior knowledge. Table 3 presents the correlations T3 between scores on open-ended and multiple-choice questions and overall self-explanation quality and prior knowledge separately for text-based, local, and global bridging questions.

Table 3 shows that performance on open-ended questions was significantly positively correlated with the quality of self-explanations but was uncorrelated with the level of prior knowledge across all the three difficulty levels of questions. In contrast, performance on multiple-choice questions was significantly positively correlated with level of prior knowledge, but was uncorrelated with quality of self-explanation. A Fisher's $z$ test comparison

Table 3
*Correlations Between Comprehension Question Performance, Prior Knowledge, and Self-Explanation Quality for Open-Ended and Multiple-Choice Questions*

| Question type | Open-ended questions | | Multiple-choice questions | |
|---|---|---|---|---|
| | Self-explanation quality | Prior knowledge | Self-explanation quality | Prior knowledge |
| Text-based | .64** | .04 | .12 | .35* |
| Local | .48** | .09 | .05 | .41** |
| Global | .38* | −.06 | −.11 | .56** |
| Average | .64** | .04 | .10 | .58** |

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

of these two average correlations indicates that self-explanation quality is more strongly correlated with performance on open-ended questions ($r = .64$) than the performance on multiple-choice questions ($r = .10$), $p < .01$. In contrast, the level of prior knowledge is more strongly correlated with performance on multiple-choice questions ($r = .58$) than open-ended questions ($r = .04$), $p < .01$.

As a follow up analysis, we performed a set of hierarchical linear regressions to identify how well overall performance on open-ended and multiple-choice questions is predicted from the five different dimensions of self-explanation and performance on prior knowledge questions. Thus, in this analysis, we used these six predictor variables: (a) frequency of accurate paraphrases, (b) frequency of accurate near-bridging inferences, (c) frequency of accurate far-bridging inferences, (d) frequency of accurate general knowledge based inferences, (e) frequency of accurate domain knowledge based inferences, and (f) overall performance on prior knowledge questions.

With regard to open-ended question performance, we examined the extent to which self-explanation quality predicted performance above and beyond performance on prior knowledge questions. We entered prior knowledge scores in Step 1 and the five measures of self-explanation quality in Step 2. The analysis indicated that prior knowledge explained less than 1% of the variance. Adding the five measures of self-explanation quality increased $R^2$ to .60 (adjusted $R^2$ to .51), with $F$ associated with the $R^2$ change (5, 27) = 7.98, $p < .01$. The overall model was significant, $F(6.\ 27) = 6.68$, $MSE = .022$, $p < .01$, with the frequency of accurate close bridging inferences, ($b = .41$, $p < .05$), distant bridging inferences ($b = .32$, $p < .05$) and domain knowledge based inferences ($b = .31$, $p < .05$) being significant predictors.

With regard to performance on multiple-choice questions, the order of the entry was reversed to examine the extent to which performance on prior knowledge questions predicts the performance, above and beyond self-explanation quality. The results indicated that the five measures of self-explanation quality in Step 1 explained 9% of the variance, $F(5, 27) < 1.0$, but none of the five self-explanation beta weights were statistically significant. Adding prior knowledge scores in Step 2 increased $R^2$ to .33 (adjusted $R^2$ to .18), with $F$ associated with the $R^2$ change (1, 27) = 9.56, $p < .01$. The overall model was marginally significant, $F(6, 27) = 2.02$, $MSE = .03$, $p = .07$. In this model, prior knowledge alone was a significant predictor ($b = .56$, $p < .01$). The results based on these regression analyses were consistent with the correlational analysis.

One issue that is not clear from the above analysis is how the accuracy of self-explanation affected the performance on comprehension questions. In order to address this issue, we examined the correlation between the average frequency of inaccurate explanations (see Table 2) and performance on open-ended and multiple-choice questions. Table 4 presents the results of this analysis. The analysis indicated that performance on open-ended questions was significantly negatively correlated with the frequency of inaccurate explanations across the three types of questions (i.e., −.63, −.44, −.42), but there were no systematic relations between the frequency of inaccurate explanations and performance on multiple-choice questions (−.15, .03, .11). Thus, this analysis indicates that self-explanation accuracy only affects performance on open-ended questions. T4

Finally, we correlated performance on open-ended and multiple-choice questions with each other, separately for text-based, local, and global questions. This provides an assessment of agreement between the question formats. This analysis is important for a methodological reason. As mentioned in the Method section, we administered open-ended questions before the multiple-choice questions without counterbalancing the presentation order. One concern for this method is that asking participants to answer open-ended question first may have created a phenomenon similar to a testing effect (Roediger & Karpicke, 2006), in which active information retrieval in response to open-ended questions affects subsequent performance on multiple-choice questions. If this is the case, performance on multiple-choice questions is likely to be positively correlated with performance on open-ended questions. The correlational analysis between performance on open-ended and multiple-choice questions across participants indicates no significant correlations: The overall correlation between performance on open-ended and multiple-choice questions was .17, $p = .30$,

Table 4
*Correlations Between Performance on Comprehension Questions and Frequency of Inaccurate Explanations as Function of Question Format*

| Question type | Open-ended questions | Multiple-choice questions |
|---|---|---|
| Text-based | −.63** | −.15 |
| Local | −.43** | .03 |
| Global | −.42** | .11 |
| Average | −.63** | −.05 |

** $p < .01$.

and separate correlations for text-based, local, and global questions were .07, $p = .68$; .16, $p = .32$; and $-.17$, $p = .29$, respectively.

In addition to this correlational analysis across participants, we also examined, for each individual, whether correctly answering an open-ended question influenced subsequent performance on the multiple-choice version of the same question. In order to perform this analysis, we first identified whether a given participant correctly (or partially correctly) answered an open-ended question and then checked whether that person correctly answered the multiple-choice version of the same question. This operation allowed us to analyse whether the proportion of correctly answered multiple-choice questions differed as a function of correctly answering the open-ended version of the same questions. A paired samples *t* test indicated no significant difference in performance on multiple-choice questions, $t(38) < 1.0$, as a function of providing the correct answer ($M = .71$, $SD = .26$) or not ($M = .70$, $SD = .29$) to the open-ended version of the same question previously. Thus, these analyses indicate that it is highly unlikely that performance on multiple-choice questions was influenced by answering open-ended questions.

## Discussion and Conclusion

This study examined whether multiple-choice and open-ended versions of the same comprehension questions assessed different aspects of comprehension, namely automatic passive and more controlled active processing, respectively. Specifically we predicted: (a) the amount of active processing as measured by the quality self-explanation should be more strongly positively correlated with performance on open-ended questions compared to multiple-choice questions; (b) the level of topic-specific prior knowledge should be more strongly positively correlated with performance on multiple-questions compared to open-ended questions; and (c) there should be very little, if any, relationship between performance on multiple-choice and open-ended questions. All of these predictions have been confirmed as discussed below.

### Differential Contribution of Active Processing to Performance on Open-Ended and Multiple-Choice Questions

First, the findings indicated that quality of self-explanation was more strongly correlated with performance on open-ended questions than with performance on multiple-choice questions, lending support to the first prediction. In addition, a follow-up analysis correlating the performance on open-ended questions with occurrence of inaccurate self-explanations indicated significant negative correlations between performance on open-ended questions and occurrence of inaccurate self-explanations. This latter finding suggests what is important for performance on open-ended questions is not just the generation of any idea, but the generation of relevant and accurate information which actually helps readers to answer the comprehension questions. Thus, the relation between the extensiveness of accurate self-explanations and performance on open-ended questions is not due to just verbosity or verbal fluency of the participants. Rather, readers needed to generate meaningful and accurate explanations during reading to perform well on the open-ended questions. These findings collectively suggest that open-ended questions are more sensitive than multiple-choice questions in measuring the quality of active generative processing of relevant and accurate ideas that occur at the time of encoding, irrespective of the amount of prior knowledge.

There is at least one alternative interpretation for the lack of correlation between performance on multiple-choice questions and the nature of active generative processing during text comprehension (i.e., extensiveness and accuracy) in this study. Specifically, this may be due to a testing effect-like phenomenon of answering open-ended versions of the same questions before answering multiple-multiple-choice questions in this study.

Although we admit that testing effects play important roles in general, we argue that our results are not due to testing effects for several reasons. First, if answering open-ended questions increased participants' performance on multiple-choice questions in general by a testing effect-like phenomenon, then the variability in performance on multiple-choice questions should have been smaller than that of open-ended questions, which we did not observe. Multiple-choice performance was well below ceiling. Also, if answering open-ended questions influenced the performance on multiple-choice questions, then we should observe a significant positive correlation between performance on the two types of questions, which we did not observe. In addition, our more detailed item-by-item comparison within individual participants indicated that performance on multiple-choice questions did not differ as a function of whether the participants correctly answered the open-ended version of the same questions. Therefore, we are inclined to conclude that the lack of correlation between the quality of self-explanation and performance on multiple-choice questions supports the reinstatement hypothesis, which specifies that answering multiple-choice questions does not reinstate the active generation of information during comprehension processes.

One question about these results, which show a lack of correlation between multiple-choice questions and active processing, is that the finding is somewhat at odds with the results reported by Magliano and colleagues (e.g., Magliano & Millis, 2003; Magliano et al., 2007). For example, Magliano and Millis (2003) showed that active processing such as paraphrases and bridging inferences, based on think-aloud protocol, has a significant positive correlation with both open-ended and true-false question performance. However, there are many differences between the Magliano and Millis (2003) study and the current study. Although we asked participants to self-explain the text, Magliano and Millis (2003) asked participants to think-aloud during the reading phase. Although both think-aloud and self-explanation produce verbal protocols, we assume that they are different, with self-explanation tending to result in more active processing (McNamara, 2004). McNamara and Magliano (2009) suggest that verbal protocols reflect relatively unaltered processing, whereas self-explanation changes comprehension during reading, which suggests that think-alouds and self-explanations induce and tap into different processes. In addition, the type of text used in this study is different from the Magliano and Millis study, which used relatively cohesive narrative and history texts. As such we believe these studies are not directly comparable. Nonetheless, it would be informative for future studies to directly test the impact of self explanation versus think-aloud protocols on reading comprehension, and whether these two types of verbal protocol techniques tap into different processes as a function of text type.

### Differential Contributions of Topic-Specific Knowledge to Performance on Multiple-Choice and Open-Ended Questions

Participants' levels of topic-specific knowledge were found to be more strongly correlated with performance on multiple-choice questions than with performance on open-ended questions. It is also worth mentioning that performance on multiple-choice questions was not correlated with the occurrence of inaccurate self-explanations. These findings indicate that multiple-choice question performance is mainly driven by topic-specific knowledge that can be retrieved passively via familiarity whereas performance on open-ended questions is not, lending support to the process reinstatement hypothesis.

Although the overall findings are in line with our prediction, the lack of correlation between performance on prior knowledge questions and performance on open-ended questions may require an explanation because this finding is at odds with theories of reading comprehension as well as some empirical findings. For example, the Construction-Integration model of reading comprehension postulates that relevant prior knowledge is a critical element in reading comprehension because reading comprehension is, in part, the integration of text content with prior knowledge (Kintsch, 1988, 1998). Evidence generally indicates that prior knowledge is a critical driving factor in reading comprehension across many different situations with different test formats including open-ended questions (e.g., McNamara et al., 1996; McNamara & Kintsch, 1996). In this sense, we acknowledge that this finding is somewhat surprising.

However, prior research has indicated that self-explanation training helps low knowledge readers to overcome knowledge disadvantages. Students with less domain knowledge particularly benefit from the comprehension strategy training because they learn effective strategies that compensate for their knowledge deficits (McNamara, 2004; O'Reilly, Best, & McNamara, 2004). When participants with low levels of prior knowledge self-explain text, while using effective reading strategies, they reword the text, make connections between ideas, and make use of whatever logic and common knowledge they have at their disposal. These active strategies help low knowledge readers catch up to high knowledge readers, which lowers the correlation between knowledge and comprehension. These effects are most pronounced on text-based questions, where all of the information to answer the question is available in the text (and doesn't call upon information available solely from prior knowledge). Likewise, in the current study, the correlation between self-explanation quality and performance on open-ended questions was highest for text-based questions (.64), suggesting that the effect of self-explanation was most beneficial for questions that could be answered based on information available explicitly in the text. These results confirm prior findings that high quality self-explanation helps low knowledge readers overcome knowledge deficits, and these benefits will be most apparent on questions that tap the text-base level of comprehension.

Finally, we note that the present findings contradict one finding by Ozuru et al. (2007) in which relatively high correlations ($.5 < r > .8$) were observed between performance on multiple-choice and open-ended versions of the same questions across the two experiments. The difference in these findings is likely related to the fact the participants in Ozuru et al. (2007) were not asked to engage in a concurrent active processing task during reading. It is known that readers tend to process a text relatively passively without actively generating inferences when they read an unfamiliar text without any specific reading purpose (Noordman et al., 1992). We suspect that Ozuru et al. (2007) participants read the text relatively passively without reliably engaging active processes. Under such circumstances, performance on multiple-choice and open-ended questions are likely to be more highly correlated due to lack of variability in controlled processing during reading; passively activated pre-existing knowledge may equally contribute to answering both types of questions.

Another noteworthy finding is that there were no differences in performance on multiple-choice versions of the questions irrespective of whether readers had correctly answered the open-ended version of the same questions previously. This is a surprising finding because intuitively, if a reader can correctly answer open-ended questions, the reader should be able to correctly identify the target among the answer options when answering a multiple-choice version of the same question. Unfortunately, we do not have data to address how this occurred. Thus, our account here is limited to speculation.

One possibility is that even when participants provide an appropriate answer to an open-ended question, they may not know whether the answer is correct. There is evidence indicating that readers' metacognitive accuracy of their confidence ratings when answering multiple-choice questions is far from perfect, with gamma correlations around .3 to .5 (Ozuru, Kurby, & McNamara, in press). Thus, at least according to data based on multiple-choice questions, readers' ability to monitor whether their answer is correct is quite limited. If the ability to monitor correctness in answering questions is limited, subsequent exposure to multiple answer options with several seductive distracters may lead readers to choose an option that is different from the one they generated at the time of answering an open-ended version of the same question because they may not be confident about what they provided earlier. This account is somewhat consistent with findings that when students are given an initial test and later take a second test, performance on the second test is greater if the students receive feedback on the initial test than if they receive no feedback (Kang, McDermott, & Roediger, 2007). It should further be noted that, as the examples of multiple-choice questions indicate (see Appendix B), the answer options of multiple-choice questions were constructed to be as seductive as possible. Thus, the presence of multiple seductive distracters may have presented challenges or difficulties that are different from the challenges present in answering open-ended versions of the same question. Nonetheless, this account is presently speculative at best. Future studies need to follow up this notion of the different type of challenges present in answering open-ended and multiple-choice versions of the same question.

Finally, we would like to note some limitations of the study here. First, given that participants self-explained the text while reading the text, our findings may not generalise to regular testing situations in which students read texts without such a concurrent task. An examination of how individual difference factors might contribute to different types of reading comprehension questions without self-explanations has been conducted in the past (see Ozuru et al., 2007). Therefore, our focus here is on how individual differences such as prior knowledge and active processing contribute to comprehension test performance when they are encouraged to actively process the text.

A second limitation of the current study is that we focused on covariation between performance on two types of questions and two types of predictor variables representing active processing during reading (overall self explanation quality) and (possibly passive) use of topic-specific knowledge across participants. As such, the grain size of the analysis is relatively coarse. In the future, studies should be conducted to examine the relations between specific types of processing and resulting performance on specific questions in more detail.

That being said, however, this study importantly provides a first step providing support to the proposal that open-ended and multiple-choice questions measure different aspects of comprehension processes, even when the multiple-choice questions are well-constructed. The present study makes a significant contribution to the field of text comprehension research and education by empirically demonstrating that multiple-choice questions and open-ended questions have, at least, a potential to measure different types of comprehension when administered in conjunction with a task that facilitates active processing (i.e., self-explanation) during reading.

## Résumé

Cette étude compare la nature de la compréhension de textes, mesurée par des questions à choix multiple et des questions à réponse libre. Les participants devaient lire un court texte et expliquer des phrases sélectionnées au préalable. Après la lecture, les participants ont répondu à deux versions des mêmes questions, à choix multiple et à réponse libre d'après ce qu'ils avaient retenu de la teneur du texte. Les résultats ont révélé que le rendement pour les questions à réponse libre était en corrélation avec la qualité des explications des participants, tandis que le rendement pour les questions à choix multiple était en corrélation avec le niveau de connaissances préalables se rapportant au texte. Ces résultats suggèrent que les questions à réponse libre et les questions à choix multiple mesurent différents aspects des processus de compréhension. La discussion des résultats est axée sur les théories de processus doubles pour la compréhension de textes.

*Mots-clés* : compréhension de la lecture, mesure de la compréhension, apprentissage, mémoire

## References

Adams, J., & Collins, A. (1979). A schema-theoretic view of reading. In R. O. Freedle (Eds.), *New directions in discourse processing, Vol. 2: Advances in discourse processes.* Norwood, NJ, USA: Ablex.

Aggleton, J. P., & Shaw, C. (1996). Amnesia and recognition memory: A re-analysis of psychometric data: *Neuropsychologia, 34,* 51–62. doi:10.1016/0028-3932(95)00150-6

Anderson, R. C. (1972). How to construct achievement tests to assess comprehension. *Review of Educational Research, 42,* 145–170.

Anderson, R. C. (1984). Role of the reader's schema in comprehension, learning, and memory. In R. C. Anderson & R. J. Tierney (Eds.), *Learning to read in American schools: Basal readers and content texts* (pp. 372–384). Hillsdale, NJ, USA: Erlbaum. AQ: 4

Bastin, C., & Van der Linden, M. (2003). The contribution of recollection and familiarity to recognition memory. *Neuropsychology, 17,* 14–24. doi:10.1037/0894-4105.17.1.14

Campbell, J. R. (1999). *Cognitive processes elicited by multiple-choice and constructed-response questions on an assessment of reading comprehension*. Unpublished doctoral dissertation, Temple University, Philadelphia, PA, USA.

Chi, M. T. H., De Leeuw, N., Chiu, M., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18,* 439–477.

Cook, A. E., Halleran, J. G., & O'Brien, E. J. (1998). What is readily available? A memory-based view of text processing. *Discourse Processes, 26,* 109–129. doi:10.1080/01638539809545041

Cook, A. E., & Myers, J. L. (2004). Processing discourse roles in scripted narratives: The influence of context and world knowledge. *Journal of Memory and Language, 50,* 268–288. doi:10.1016/j.jml.2003.11.003

Goldman, S. R., & Durán, R. P. (1988). Answering questions from oceanography texts: Learner, task and text characteristics. *Discourse Processes, 11,* 373–412. doi:10.1080/01638538809544710

Graesser, A. C., Ozuru, Y., & Sullins, J. (2010). What is a good question? In M. McKeown & L. Kucan (Eds.), *Bringing reading research to life* (pp. 112–141). New York, NY: Guilford Press.

Jacoby, L. L. (1996). Dissociating automatic and consciously-controlled effects of study/test compatibility. *Journal of Memory and Language, 35,* 32–52.

Kang, S. H. K., McDermott, K. B., & Roediger, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology, 19,* 528–558. doi:10.1080/09541440601056620

Khoe, W., Kroll, N. E. A., Yonelinas, A. P., Dobbins, I. G., & Knight, R. T. (2000). The contribution of recollection and familiarity to yes-no and forced-choice tests in healthy subjects and amnesiacs. *Neuropsychologia, 38,* 1333–1341. doi:10.1016/S0028-3932(00)00055-5

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95,* 163–182. doi:10.1037/0033-295X.95.2.163

Kintsch, W. (1993). Information accretion and reduction in text processing: Inferences. *Discourse Processes, 16,* 193–202. doi:10.1080/01638539309544837

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York, NY, USA: Erlbaum.

Leahey, T. H., & Harris, R. J. (1997). *Learning and cognition*. Upper Saddle River, NJ, USA: Prentice Hall.

Long, D. L., & Lea, B. R. (2005). Have we been searching for meaning in all the wrong places? Defining the "search after meaning" principle in comprehension. *Discourse Processes, 39,* 279–298.

Magliano, J. P., & Millis, K. (2003). Assessing reading skill with a think-aloud procedure and latent semantic analysis. *Cognition and Instruction, 21,* 251–283. doi:10.1207/S1532690XCI2103_02

Magliano, J. P., Millis, K., Ozuru, Y., & McNamara, D. S. (2007). A multidimensional framework to evaluate reading assessment tools. In D. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (pp. 107–136). Mahwah, NJ, USA: Erlbaum. AQ: 5

McNamara, D. S. (2004). SERT: Self explanation reading training. *Discourse Processes, 38,* 1–30. doi:10.1207/s15326950dp3801_1

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction, 14,* 1–43. doi:10.1207/s1532690xci1401_1

APA NLM

tapraid5/cae-cae/cae-cae/cae00313/cae0200d13z | xppws | S=1 | 7/31/13 | 21:42 | Art: 2012-0312



McNamara, D. S., & Kintsch, W. (1996). Learning from text: Effects of prior knowledge and text coherence. *Discourse Processes, 22,* 247–288. doi:10.1080/01638539609544975

McNamara, D. S., & Magliano, J. P. (2009). 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). New York, NY, USA: Routledge.

McNamara, D. S., O'Reilly, T. P., Best, R. M., & Ozuru, Y. (2006). Improving adolescent students' reading comprehension with iSTART. *Journal of Educational Computing Research, 34,* 147–171. doi:10.2190/1RU5-HDTJ-A5C8-JVWE

Morris, D. M., Bransford, J. S., & Franks, J. J. (1977). Levels of processing vs. transfer appropriate processing. *Journal of Verbal Learning & Verbal Behavior, 16,* 519–533. doi:10.1016/S0022-5371(77)80016-9

Noordman, L. G., Vonk, W., & Kempff, H. J. (1992). Readers' knowledge and the control of inferences in reading. *Language and Cognitive Processes, 7,* 373–391. doi:10.1080/01690969208409392

O'Reilly, T., Best, R., & McNamara, D. S. (2004). Self-explanation reading training: Effects for low-knowledge readers. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th Annual Conference of the Cognitive Science Society* (pp. 1053–1058). Mahwah, NJ, USA: Erlbaum.

Ozuru, Y., Best, R., Bell, C., Witherspoon, A., & McNamara, D. S. (2007). Influence of question format and text availability on the assessment of expository text comprehension. *Cognition and Instruction, 25,* 399–438. doi:10.1080/07370000701632371

Ozuru, Y., Kurby, C. A., & McNamara, D. S. (in press). The effect of metacomprehension judgment task on comprehension monitoring and metacognitive accuracy. *Metacognition and Learning*.

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17,* 249–255. doi:10.1111/j.1467-9280.2006.01693.x

Rumelhart, D. (1985). Toward an interactive model of reading. In H. Singer & R. B. Ruddell (Eds.), *Theoretical models and processes of reading* (3rd ed., pp. 722–750). Newark, DE, USA: International Reading Association.

Singer, M., & Remillard, G. (2004). Controlled and automatic influences in text retrieval. *Memory & Cognition, 32,* 1223–1237. doi:10.3758/BF03206314 AQ: 6

van den Broek, P., Rapp, D. N., & Kendeou, P. (2005). Integrating memory-based and constructionist processes in accounts of reading comprehension. *Discourse Processes, 39,* 299–316.

Yonelinas, A. P. (2002). The nature of recollection and familiarity. *Journal of Memory and Language, 46,* 441–517. doi:10.1006/jmla.2002.2864

## Appendix A

AQ: 7

## Text Used for Self-Explanation and Reading Comprehension Tasks

### Why Is There Sex?

We tend to take the existence of sexual reproduction for granted. From an evolutionary standpoint this is a serious puzzle. The simplest and most direct way of reproducing one's genes is fission, creating a copy of yourself. Sexual reproduction incurs many costs. One has to find a mate, and has to risk mingling one's genes with another individual's, whose fitness is unknown. Animals have to go to great lengths to find them, choose the best ones, and try to guarantee that the offspring are their own. Sexual reproduction is a costly and risky business. **Why does it not only exist, but exist in so many species**?

The question has been addressed by many theories over the years. The central feature of reproduction that they focus on is the fact that it introduces variation into nature. Sex reshuffles genes into new combinations. **This makes offspring quite different from either parent.** The currently favoured account of the evolutionary value of reshuffling is the "Red Queen" theory, named after the character in Alice in Wonderland. **The reshuffling of genes keeps us one step ahead of our most dangerous enemies: viruses, bacteria, and parasites**.

**Imagine owning a home threatened by clever burglars.** You put locks on doors, but they figure out how the locks work and get in. Our immune systems are the locks that our bodies use to repel micro-organisms and parasites. They are the burglars who figure out how to pick the locks. Now, if we reproduced by fission, our offspring would be exact copies of ourselves. **They possess the same locks of the immune system.** Soon, viruses, bacteria, and parasites would evolve into perfect adaptation with our immune system. We would be defenseless.

**Shuffling our genes results in new immune systems.** This protects ourselves against the organisms who had adapted themselves to our last generation. They will catch up to us, and we have to keep running away to stay just ahead of them. Hence the name for the theory.

The Red Queen turns up in many areas of evolution. Any time organisms interact competitively, it may be found. **Each competitor evolves to adapt to the strategies of the other, who in turn evolves new strategies of defense**.

XXX: target sentences that participants self-explained.

(*Appendices continue*)

## Appendix B

## Examples of Comprehension Questions

### Open-Ended Questions

When finding a mate, why does mixing one's own genes with another person's genes pose a risk?

What is the role of the immune system in a living organism?

Explain the gist of the Red Queen theory in evolution.

### Multiple-Choice Questions

When finding a mate, why does mixing a one's own genes with another person's genes pose a risk?

a. One cannot be certain about another individual's fitness.

b. Another individual's genes may carry parasites. (near-miss)

c. Another individual's genes may be stronger and dominate the offspring (unrelated).

d. One cannot be certain whether two sets of genes fit together. (thematic)

Which of the following most accurately describes the role of the immune system in a living organism?

a. Prevents micro-organisms from entering the digestive system. (thematic)

b. Helps the organism remove poisonous materials from the body. (unrelated)

c. Prevents micro-organisms from invading the system.

d. Prevents the organism from having genetic abnormalities. (near-miss)

What is the gist of "Red Queen" theory in evolution?

a. Weaker organisms need to keep running away from stronger organisms. (unrelated)

b. Sexual reproduction introduces variation in nature. (near-miss)

c. Organisms need to continuously evolve to outrun their enemies which try to catch up.

d. Only individuals that successfully adapt to the environment survive to reproduce their genes. (thematic)

## Appendix C

## Examples of Domain Knowledge Questions

What is a chromosome? What is a pathogen?

An individual's sex chromosomes determine whether they are male or female. Human females have two X chromosomes (XX), and human males have ________.

From an evolutionary perspective, an organism's individual fitness is measured in terms of ____ .

a. Length of life

b. the number of reproductive partners

c. the number of offspring

d. Height

Which of the following is a non-specific defense mechanism of the human immune system?

a. Skin

b. Nerve Cell

c. Muscle

d. Bone

(*Appendices continue*)

APA NLM

tapraid5/cae-cae/cae-cae/cae00313/cae0200d13z | xppws | S=1 | 7/31/13 | 21:42 | Art: 2012-0312



## Appendix D

## Example Coding of Self-Explanation

| Target sentence | They (off spring) possess the same locks of the immune system. | |
|---|---|---|
| Self-explanation | If offspring were reproduced by fission, then the locks would always be the same. Therefore, the bacteria would not have to do anything different to infect the offspring since the offspring have the same defenses. | |
| **Coding Type** | **Coding** | **Remark (relevant part of the SE for the coding)** |
| Paraphrase | Accurate paraphrase | "then the locks would always be the same. since the offspring have the same defenses" |
| Near bridging | Accurate near bridging | "If offspring were reproduced by fission" (see Appendix A for preceding sentence) |
| Far bridging | Not present | |
| General knowledge | Relevant and accurate GK inference | "Therefore, the bacteria would not have to do anything different to infect the offspring" |
| Domain knowledge | Not present | |

Received June 25, 2012
Accepted April 4, 2013 ■