Effects of Font Type of Embedded Key Terms on Judgments of Learning and Memory from Textbooks

by

Andrew Scott Roberts

A thesis submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Masters of Science

Auburn, Alabama
May 6th, 2013

Keywords: judgments of learning, reading ability, working memory, perceptual salience

Approved by
Aimee Callender, Assistant Professor of Psychology
Alejandro Lazarte, Associate Professor of Psychology
Ana Franco-Watkins, Associate Professor of Psychology

Abstract

Previous research by Rhodes and Castel (2008) provided evidence that metacognitive predictions of future memory performance (i.e. judgments of learning (JOLs)), are significantly affected by perceptual information (i.e. font size) present during learning with no actual benefits to memory. Based on these findings, the current study investigated whether font changes used to emphasize key terms within texts (i.e. bold font) affect metacognition and reading comprehension. Participants read two texts that contained five embedded key terms (per passage) that were presented in either bold or non-bold fonts and were counterbalance for topic and font. For each passage, participants provided two global-level JOLs (i.e. predictions based on their memory for the entire passage prior to and immediately following testing), several term-specific JOLs (i.e. memory predictions for each key term), and completed a multiple choice test that consisted of an equal number of term-specific and inference-based questions (five questions per type). In addition to JOLs and performance, participants' reading ability and working memory were evaluated to determine if individual differences predicted differences in the effects of key term font on JOLs and memory. Results from the current study did not provide evidence that bold key terms influences either global or term-specific JOLs; however, reading ability did significantly predict differences in performance and in some cases, differences in participants bias when making JOLs. Evidence was also present to suggest that under certain circumstances, embedded bold key terms can significantly improve performance based on text difficulty.

Table of Contents

List of Tables

List of Figures

Effects of Font Type of Embedded Key Terms within Texts on Students' Judgments of Learning and Performance

Due to the demands placed on students (amount of material to-be-learned, time constraints, etc.), Rawson and Dunlosky (2007) argue that the task of successful students is not to achieve high levels of learning but to efficiently regulate their study behaviors. Although successful monitoring can improve students' learning, failures in self-monitoring accuracy could lead to detrimental study choices (Maki, Willmon, & Pietan, 2009). If students are overconfident in their preparation, they may prematurely suspend their studies; however, if they are underconfident, they could choose to continue to study despite diminishing returns for their extra effort. In light of the potential consequences associated with these choices, ongoing metacognitive research—the study of individuals' knowledge of their own knowledge (Metcalfe, 2009)— is necessary to further our understanding of the factors that influence and potentially compromise students' study choices.

Within metacognition, one method researchers employ to evaluate individuals' understanding of their own knowledge when studying is through the use of judgments of learning. Judgments of learning (JOLs) can be considered measures of metacognition because they provide an index of individuals' confidence about what they *believe* they know while studying (Metcalfe & Finn, 2008). Metcalfe and Finn (2008) demonstrated that when students are asked to provide JOLs for recently studied materials, their future study choices are affected more by these JOLs than prior test performance. In other words, when students believe they

knew the material (based on their JOLs), they were less likely to choose to study this material again even when their prior test performance was poor (Metcalfe, 2009).

Similar to metacognition, failures in absolute metacomprehension—the difference in accuracy between readers' predictions and their actual test performance when reading texts—are well documented (Dunlosky & Lipko, 2007; Maki, Shields, Wheeler, & Zacchilli, 2005; Thiede, Griffin, Wiley, & Anderson, 2010; Wiley, Griffin, & Thiede, 2005). Students' accuracy is generally quite poor when they are asked to make global predictions—confidence judgments of their knowledge for *all* recently studied materials (Rawson & Dunlosky, 2007)— and correlations between these JOLs and test performance rarely exceed +.40 (Dunlosky, Rawson, & Middleton, 2005). Even when students are asked to make term-specific judgments—predictions based on the definition of key terms—their accuracy does not improve significantly; however, term-specific predictions' accuracy improves when students are forced to recall definitions prior to making these JOLs.

Metacomprehension accuracy is influenced by readers' characteristics as well as factors specific to the text. Individual differences in both reading comprehension and working memory may affect the ability to monitor comprehension which can affect monitoring accuracy. Likewise, perceptual saliency of targeted information in texts could affect metacomprehension accuracy. Each of these factors and how they may affect metacomprehension are discussed below.

Reading Comprehension

Gernsbacher and Varner (1990) suggest that the goal of reading comprehension is to build a cohesive, mental structure to represent knowledge. According to their *structure building framework* theory, mental structures begin with a basic foundation and when new information

coheres or relates to prior knowledge, it can be added onto these pre-existing structures. But if the new information is less coherent, a new substructure is developed, which is less likely to activate previous knowledge resulting in reduced accessibility to previously built structures

Kintsch (1988; 2005) proposed a *construction-integration model* of text comprehension. In his theory, text comprehension emerges from an interaction of bottom-up and top-down processes that occur on different levels of text representation. Initial processing is based on surface level characteristics of the text (individual words and their linguistic relationships to each other) (Kintsch, 1994). When semantic information is extracted from the surface characteristics, the text base emerges, in which, the content of what was directly written is represented in memory as a network of propositions (McNamara, Kintsch, Songer, & Kintsch, 1996); however, understanding the text base, while sufficient for reproduction, does not guarantee comprehension.

To account for the deeper levels of understanding necessary for comprehension, Kintsch (1998) advocated the development of a situation model which shares similarities with the mental structures proposed by Gernsbacher, Varner, and Faust (1990). According to this theory, the acquisition of a situation model requires that readers have adequate prior knowledge *and* use this repository to make active inferences while reading (McNamara et al., 1996). By accessing prior knowledge (long-term memory) to elaborate on the text base, readers can (hopefully) generate predictive inferences while reading which will permit them to integrate the text base into a situation model (Kintsch, 1994; McNamara et al., 1996); however, Kintsch (1994) argues that if these processes are to take place, specific assumptions about the types of inferences readers' generate must be made in order to determine ways in which texts can facilitate readers' propensity to make inferences.

Perceptual Salience

In the current study, a common assumption about how to improve learning from textbooks is through using font changes (bold and/or italicized) to highlight key concepts. Despite being a common publishing practice, it remains to be seen if these visual enhancement promote improvements in learning or actually result in some readers misperceiving perceptual clarity as a predictor of comprehension (i.e. making failures in metacomprehension). Although students have demonstrated the ability to accurately judge the learnability (i.e. readability) of the text based on how much they perceive they will learn (Britton, Van Dusen, Gülgöz, Glynn, and Sharp, 1991), a question that must be asked is whether the accuracy of JOLs for key concepts (term-specific JOLs) are influence by the font type(s) used to present the words within texts and whether students global assessment of learning are influenced by these font changes?

Although a review of the literature failed to find any previous studies that investigated the effects of font type of embedded key terms with texts on JOLs, perhaps the most convincing argument about the potential flaws of this practice can be found in Rhodes and Castel (2008), who demonstrated that perceptually salient information influences JOLs. By manipulating the font size (18 pt. font vs. 48 pt. font) of to-be-remembered words within word lists, they demonstrated that participants provided significantly larger JOLs for words presented in the much larger font and that font size was not predictive of recall. In other words, even though participants gave larger JOLs for larger words, they were not more likely to recall these words later compared to the smaller font size. Furthermore, this perceptual illusion persisted even when participants were explicitly informed that memory was not better for words presented in the larger font and instructed to ignore differences in font size when making JOLs.

Recently, McDonough and Gallo (2011) replicated the findings from Rhodes and Castel

(2008) by using a recognition test. Difference in font size did not predict difference in participants' recognition despite their significantly larger JOLs for word presented in the larger font. Furthermore, they demonstrated that when participants were instructed to selectively search through their memory for words presented in either large or small font, the occurrence of false recognitions were significantly reduce for words present in the larger font.

Even though the results obtain in Rhodes and Castel (2008) and McDonough and Gallo (2011) represent very extreme examples of the effects of font-type on JOLs, the ramifications of these works should be noted. Participants in these studies relied on salient perceptual information during encoding to make predictions about their memory for the to-be-learned words. In spite of their confidence (JOLs), words that were more perceptual salient (i.e. larger in font size) were not more likely to be recalled or recognized later.

One explanation for why JOLs could be sensitive to perceptual information is Koriat's (1997) *cue-utilization approach*. In this theory, he suggests that when JOLs are made, individuals rely on currently available cues to make predictions about their performance in the future. These cues are problematic because they inflate JOLs during learning (Koriat & Bjork, 2005) and their absence during testing results in poorer performance. Based on the results obtained by Rhodes and Castel (2008) and McDonough and Gallo (2011), participants may have been relying on intrinsic cues –specific characteristics of the study items that indicate the ease/difficultly of learning these items (Koriat, 1997)—to make their predictions because the font size of words inflated participants' JOLs.

The implications of this theory suggest that the salience of readily available "cues" in short-term memory can affect JOLs (Koriat, 1997). Although the font changes within textbooks may not be quite as extreme or as perceptually salient as those used in Rhodes and Castel (2008)

and McDonough and Gallo (2010), it seems reasonable to suggest based on their findings and Koriat's (1997) *cue-utilization approach* that JOLs for embedded key concepts could vary as a function of the font these words are presented in.

Working Memory

Working memory is generally conceptualized as the cognitive resources one has available to dedicate to a task (Linderholm & Broek, 2002) and is believed to play a critical role in integrating information when reading (Daneman & Carpenter, 1983; Lee, Tedder, & Xie, 2006) through two primary functions, maintenance and retrieval (Unsworth & Engle, 2007). Maintenance is a highly involved process of trying to maintain novel information in memory in spite of both capacity limitations—the amount of information one can successfully maintain at a given time— and the presence of potential distracters that could disrupt this process by using some of these limited attentional resources. Because working memory has limitations, retrieval of information from maintenance may occur at times in the presence of irrelevant information. Therefore, the success of retrieval is dependent upon individuals' ability to discriminate between relevant and irrelevant information in working memory using contextual cues, but when more attentional resources are required during reading, comprehension can be reduced (Gasser, Boeke, Haffernan, & Tan, 2005).

One of the potential outcomes of differences in working memory is that individuals may vary in their reading efficiency (i.e. how successfully than can integrate information while reading to create a situation model). For instance, less efficient readers may devote too much capacity to processing incoming words, leaving fewer resources available for comprehension, which Daneman and Carpenter (1983) suggest would be the functional equivalent of having a smaller working memory capacity. Along these lines, slower readers have smaller perceptual

spans (regions within text where they obtain useful information) and are believed to use more resources to encode and understanding words they fixate on during reading (Rayner, Slattery, & Bélanger, 2010). Essentially, one could argue that because these readers require more resources to represent and understand the text base, they have fewer resources available to constructing a situation model. Therefore, it should not be surprising that low working memory readers tend to make fewer predictive inferences, repeat texts more, and generally recall less content than high working memory readers (Linderholm & Broek, 2002).

In a recent study, Linderholm, Cong, and Zhao (2008) used working-memory capacity to evaluate individual differences between readers predicted performance (JOLs) and actual performance on reading comprehension tests. By using difference scores to indicate participants' absolute monitoring accuracy, they were able to investigate readers' bias as a function of reading purpose. The results of their study demonstrated that low working-memory capacity individuals, despite spending more time reading during a study condition, performed poorer when tested and were more overconfident (i.e. biased) making their JOLs in comparison to high working-memory capacity individuals. As a consequence of spending more time reading, low working-memory capacity readers, who are typically considered to be less efficient readers, overestimated their performance because of the amount of effort (time) required for them to complete the reading. Nevertheless, the pattern of inflated JOLs for participants with lower working-memory capacity is consistent with the general finding that lower verbal ability students show overconfidence in making their JOLs compared to higher verbal ability students (Maki, Shields, Wheeler, & Zachhilli, 2005).

Although Linderholm, Cong, and Zhao (2008) did not examine the effects of font type in their study, several studies have demonstrated that font changes can affect reading processes.

Phrases within text that are presented in different fonts can be distracting and disrupt reading processes (Carlson, Hasher, Connelly, & Zacks, 1995). For example, readers of all ages are more likely to fixate on distractors (random words) embedded in sentence that differ in font (italics) compared to differences in font color (red), suggesting that changes in font type (spacing and orientation of letters) are more difficult to process when reading (Kemper & McDowd, 2006). Furthermore, reading rates are faster for passages presented in Times New Roman in comparison to Harrington and Script MT bold fonts (Slattery & Rayner, 2009).

Rationale for Current Study

The results from Kemper and McDowd (2006) and Slattery and Rayner (2009) would seem to suggest bold and italicized font could be detrimental when reading, but due to methodological differences between these studies and the current study, we cannot make definitive conclusions. For starters, entire passages were presented in one font type in Slattery and Rayner (2009) and the font changes in Kemper and McDowd (2006) were used only for distracter items. In the current study, only select key terms—not distracters—were presented in bold font which is more consistent with font changes used by textbook authors.

The changes in text font in Kemper and McDowd (2006) were designed to investigate participants' ability to inhibit distracting text from disrupting their reading processes. In contrast, the current study attempted to investigate the effects of embedded key terms' font on two concurrent processes involving attention and reading comprehension: 1) the acquisition of relevant information related to key concepts; and 2) the ability to inhibit the effects of perceptually salient key terms during the process of making JOLs. Because of novel nature of the current study, individual differences in reading comprehension and working memory were investigated using the Multi-media Comprehension Battery (MMCB) and the Operation Span

Task (OSPAN) to determine if differences in these abilities would predict differences in JOLs for key terms embedded in text and performance.

Hypothesis

Hypothesis 1: Reading ability will significantly correlate with working memory (OSPAN) and these measures combined will significantly predict changes in JOLs and criteria performance better than either measure by itself

Hypothesis 2: Judgments of learning (JOLs) for bold-faced key terms will be significantly larger overall than judgments of learning (JOLs) for non-bold key terms.

Hypothesis 3: Judgments of learning (JOLs) for lower reading ability participants will be significantly larger for bold-faced key terms compared to lower reading ability participants; judgments of learning for non-bold key terms and high reading ability participants' judgments of learning for both bold-faced and non-bold key terms because they rely on perceptually salient information more to make predictions about their future performance.

Method

Participants & Design

Eighty-nine undergraduates enrolled in introductory to psychology courses at Auburn University were given extra credit for participating in the current study. Basic demographic information (age, gender, year in school) was obtained from all participants prior to participating. Participants were randomly assigned to one of four text conditions that were counterbalanced according to reading topic and the font type of five key terms embedded in each reading.

The current study was a 2 x 2 x 2x 2 between-subjects design with two levels of key term fonts (bold vs. non-bold), two reading ability groups (high vs. low), two reading topics (anxiety disorders vs. mood disorder), and order (first text vs. second text). The dependent variables consisted of participants' performance on two multiple choice tests (one per reading topic), 10 term-specific judgments of learning (five per text), and pre- and post-test global judgments (per text). In addition, working memory and reading ability were assessed and treated as discrete variables during analysis.

Materials

Multi-Media Comprehension Battery. To assess participants' reading ability, they completed the Multi-Media Comprehension Battery (MMCB) developed by Gernsbacher & Varner (1988). The MMCB consists of six stories presented in three modalities (two stories per modality): auditory sentences, written sentences, and visual pictures. After finishing a story, participants are asked to answer comprehension questions based on what they have read, heard, or viewed. The MMCB has been shown to have predictive validity and measures general

comprehension skill (Maki & Maki, 2002). Verbal performance on the SAT and comprehension in the written sentences conditions on the MMCB are highly correlated (r = .64) (Gernsbacher, Varner, & Faust, 1990). Because this measure was being used to assess participants' reading ability, the version used in the current study consisted of four verbal stories (the two auditory stories were converted into written sentences). Performance on the MMCB was used as a continuous variable during analysis to predict differences in participants' JOLs and test performance.

Automated Operational Span Task. An automated version of the operation span task (OSPAN) developed by Unsworth, Heitz, Schrock, & Engle (2005) was used to assess participants working memory (WM). It is an easy to administer self-guided assessment of working memory that correlates with other working memory measures (r = 0.45, p < .01) and has good internal reliability (alpha = .78). During this task, participants must remember the correct order of a series of letter as they answer simple math equations. The length of the series incremental increases over time and performance on this measure is based on the correct letter recall while maintaining at least 85% accuracy on the math portion. Performance on this measure was used as a continuous variable during analysis to predict differences in participants' JOLs and test performance.

Reading Samples. In order to determine the effects of font type of embedded key terms on participants' judgments of learning and performance, two reading samples, anxiety disorders (~1,250 words, Flesch-Kincaid = 11.04) and mood disorders (~1,150 words, Flesch-Kincaid = 12.10) were created by dividing up a longer reading sample (~2,400 words, Flesch-Kincaid = 11.53) obtained from a "Psychological Disorders" chapter published in Myers' *Psychology* (*9th*

ed) that was piloted in a previous study[1]. Both samples were presented in their entirety (participants had to scroll through the text) and contained five embedded key terms that could appear in either bold or non-bold font depending on the reading condition (see Appendices A and B).

Multiple Choice Tests. To test participants' knowledge of each reading sample, they were given two multiple choice tests that contained five term-specific questions and five inference-based questions (per test) with four answer choices per question (see Appendices C and D). Test questions were randomized and presented individually with participants being required to select an answer to each question before the subsequent question was presented.

Demographic. A brief demographic survey develop by the researcher was administered on a paper to each participant and consisted of four questions: age, gender, year in school, and major.

Procedure

The current experiment occurred during two separate sessions with a 1-hour time limit per session. During the first session, participants filled out a brief demographic and then completed the MMCB followed by the OSPAN to obtain measures of reading ability and working memory, respectively. During the second session, participants were randomly assigned to a reading condition. Each condition consisted of three reading passages, a short practice text and two reading samples (anxiety disorders and mood disorders) that were counterbalanced for topic and the font of key terms between conditions.

Prior to beginning the reading condition, participants were given a general set of verbal instructions about the experiment. The experimenter briefly explained the tasks and participants were instructed to read each sample like they normally would if they were studying for an

[1] To match samples for length, a small section based on somatoform disorder was added to the piloted text.

upcoming exam because they would be tested on the reading materials. Participants were also made aware that they would be asked to judge their confidence on a 0 to 100 scale with 100 indicating absolute confidence that they could correctly answer questions about the material they have read. When making these judgments, the experimenter emphasized to participants that they could use any value within this range, but they should try to be as accurate and honest as possible because they would be asked to answer questions about the material.

Following these verbal instructions, participants began their assign reading condition on a computer. This portion of the experiment was an entirely self-contained, self-paced program with on-screen instructions presented throughout the program, which permitted participants to control their rate of progression throughout the session without any additional instructions from the experimenter. Each reading condition began with the short practice text and after reading this text, participants were reminded that they would be asked to make predictions based on their knowledge of the material they read and that they should be as accurate and honest as possible in making their predictions because they would be tested on the material.

Immediately following this reminder, participants received a prompt to provide a pre-testing global JOL (prediction) that asked them to predict how well they would do on a test (0-100) if they were tested on all the material they just read. To make this prediction, participants had to type in their response. Following the global JOL, participants then were prompted to make three term-specific JOLs. During this portion of the experiment, each key term from the practice reading was individually presented on the screen (non-bold) and participants had to provide a response on a 0 -100 scale to indicate how confident they were that they could answer a question about each key term. Once participants completed these term-specific JOLs, they were given a brief five-item multiple choice test (three questions were term-specific). After completing the

practice test, participants were then prompted to provide a post-testing global JOL (postdiction) that asked them to predict their performance (0-100) based on the test they just completed. When participants submitted their post-diction for the practice test, they were informed that they were ready to begin the experiment.

The main experiment began with participants reading the first of two reading samples (anxiety disorders or mood disorders) with five key terms embedded in the text appearing in either bold or non-bold font. When participants finished reading the first sample, they provided a pre-testing global JOL followed by a series of five term-specific JOLs according the procedure described above. Following the last term-specific JOL, participants completed a 5 minute distractor task (Tetris) and were given a multiple choice test based on the first reading sample. Immediately after testing, participants provide a post-testing global JOL based on their perceived performance for the first multiple choice test.

For the second reading sample, which was always the topic not previously read and contained the opposite font type for key terms that was not presented in the first reading sample, participants complete all steps according to the procedure described above and were then dismissed from lab.

Out of the 89 participants whom completed the current experiment, several participants' data was excluded from further analysis. Three participants had too many math errors on the Operational Span Test and one participant performed below chance level on the Multimedia Comprehension Battery. Eight participants failed to provide an appropriate response to at least one of the judgment of learning (JOLs) prompts and one participant was determined to be outliers based on their first text performance. These exclusions resulted in an $N = 76$ that was used during analysis for the current study.

Results

Individual Differences Measures

Reading ability. A median-split was conducted on participants' performance on the Multi-media Comprehension Battery (MMCB) to create two reading ability groups to be used in analysis (*Mdn* = 33.5, range from 17 to 44 out of a possible 48). Scores of 33 or less were classified as low reading ability (*n* = 38) and scores of 34 or more were classified as high reading ability (*n* = 38).

Working memory. A median-split was also conducted on participants' performance on the Operational Span Task (OSPAN) to create two working memory ability groups to be used in analysis (*Mdn* = 48, range from 3 to 75 out of a possible 75). Scores of 48 or less were classified as low working memory (*n* = 40) and scores of 49 or more were classified as high working memory (*n* = 36).

Test Performance

Order. A 2 x 2 x 2 repeated measures analysis of variance with order (first and second text performance) as a within-subjects factor and reading ability (high vs. low) and working memory (high vs. low) as between-subject factors was conducted on the multiple choice performance to determine if overall performance (ignoring topic and key term font) was affected by text order. Results revealed significant main effects of order, $F(1, 72) = 5.97$, $p = .02$, $\eta_p^2 = .08$ and reading ability, $F(1, 72) = 9.19$, $p = .003$, $\eta_p^2 = .11$. Working memory was not significant, $F(1, 72) = 3.24$, $p = .08$, $\eta_p^2 = .04$, and there were no significant interactions between order,

reading ability, or working memory.

Pairwise comparison of the means with Bonferroni corrections determined that first text performance (M = 77.53, SE = 2.12) was significantly better than second text performance (M = 70.38, SE = 2.13) and that the average of high reading ability participants' performance between the two texts (M = 78.61, SE = 2.15) was significantly better than low reading ability participants' performance (M = 68.30, SE = 2.19).

The implications of this initial analysis are that text order, independent of topic and key term font, affected individual participants' test performance based on when each text was read, with the second text resulting in poorer performance. In other words, performance within identical conditions (same topic and key term font) could differ based entirely on which text was read first. Therefore, order cannot be collapsed across in subsequent analyses.

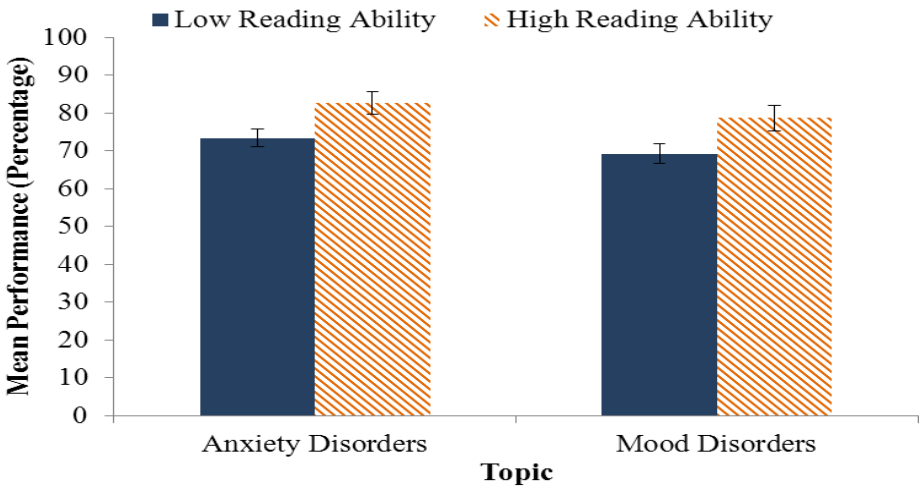
Regarding measures of individual differences, there was only a significant effect of reading ability. Working memory was not significant and did not interact with reading ability or order. Contrary to what was hypothesized, the combination of reading ability and working memory does not explain performance better than either measure by itself and continuing forward, only reading ability will be used during analysis to predict changes in performance.

Text Topic and Font

Because the prior analysis examined performance regardless of key term font and determined that there were effects of order, two separate 2 x 2 x 2 ANOVAs were conducted using performance based on a specific topic (anxiety disorders and mood disorders) as the dependent variable and key term font (bold vs. non-bold), order (first vs. second text), and reading ability (high vs. low) as between-subjects factors to determine if there were any effects of topic and/or font on performance.

For both topics, there was a significant main effect of reading ability (see Figure 1), $F(1, 68) = 6.38$, $p = .01$, $\eta_p^2 = .09$; $F(1, 68) = 3.93$, $p = .05$, $\eta_p^2 = .06$. However, there was not a significant effect of font on performance for either anxiety disorders (bold: $M = 80.56$, $SE = 2.34$ vs. non-bold: $M = 82.53$, $SE = 2.45$) or mood disorders (bold: $M = 71.76$, $SE = 3.32$ vs. non-bold: $M = 64.26$, $SE = 3.17$), $F(1, 68) = 0.34$, $p = .56$, $\eta_p^2 = .01$; $F(1, 68) = 2.67$, $p = .11$, $\eta_p^2 = .04$, or significant effects of order, $F(1, 68) = 3.73$, $p = .06$, $\eta_p^2 = .05$; $F(1, 68) = 3.29$, $p = .07$, $\eta_p^2 = .05$, and there were no significant interactions.

Figure 1. Comparisons of performance by reading ability and topic.

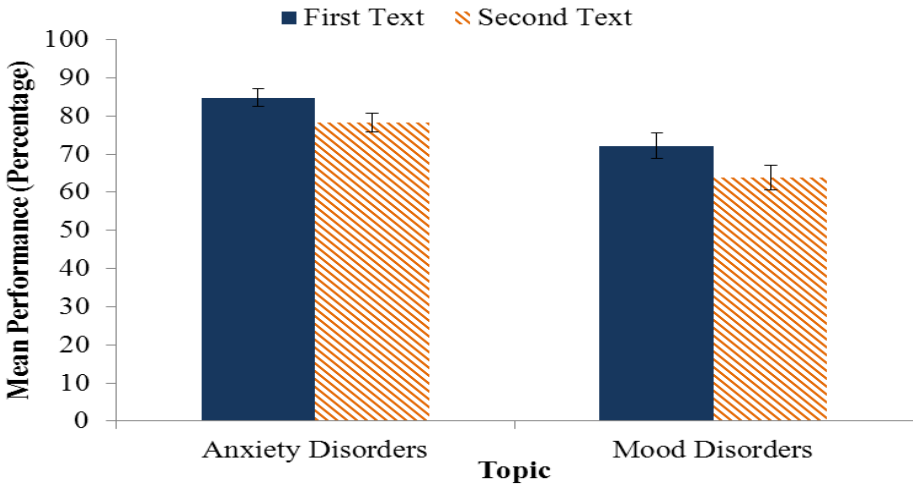


Pairwise comparisons with Bonferroni corrections revealed that for both topics, high reading ability participants' performance was significantly better than low reading ability participants' performance (anxiety disorders: $M = 86.25$, $SE = 2.41$ vs. $M = 76.85$, $SE = 2.39$; mood disorders: $M = 74.18$, $SE = 3.26$ vs. $M = 61.85$, $SE = 3.24$).

Although there were no significant effects of order on performance within each topic (see Figure 2), the patterns were trending toward significance (especially for anxiety disorders, $F(1, 68) = 3.73$, $p = .058$, $\eta_p^2 = .05$), which suggests that performance is still affected by the order of

the text.

Figure 2. Comparisons of performance by topic and order.



In light of the general pattern in performance based on the order of the texts, it is important to investigate the effects of font and topic on performance within each separate text position (first or second text) in order to better detect differences in performance and metacognition regarding the font of key terms. As it currently stands, there are no effects or interactions of font, but failure to detect difference may be due the effects of font on performance being nullified by the order of testing.

Performance by Text Position

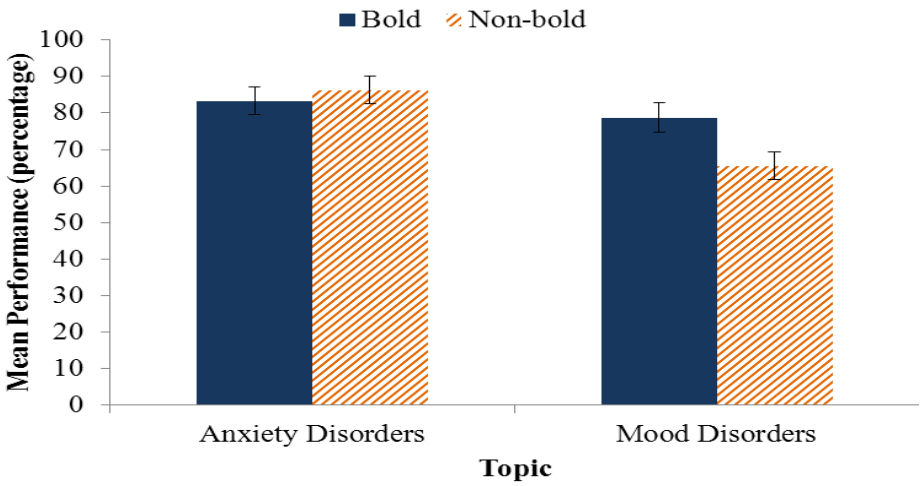
First text. For the first text position, a 2 x 2 x 2 ANOVA was conducted using reading ability (high vs. low), topic (anxiety disorders vs. mood disorders) and font (bold vs. non-bold) as between-subjects factors. Results revealed a significant main effect of reading ability, $F(1, 68) = 7.98$, $p = .01$, $\eta_p^2 = .11$, topic, $F(1, 68) = 10.70$, $p < .01$, $\eta_p^2 = .14$, and a significant interaction between topic and font, $F(1, 68) = 4.35$, $p = .04$, $\eta_p^2 = .06$.

Pairwise comparisons with Bonferroni corrections determined that high reading ability participants' performance ($M = 83.96$, $SE = 2.74$) was significantly better than low reading

ability participants' (M = 73.04, SE = 2.73) and performance on anxiety disorders (M = 84.82, SE = 2.67) was significantly better than performance on the mood disorders (M = 72.17, SE = 2.80).

The main effect of topic was qualified by a two-interaction between topic and font (see Figure 3). Upon examination of the means, the font did not affect performance on the anxiety disorders text whereas for the mood disorders text the bold font resulted in better performance than the non-bold font. Pairwise comparisons with Bonferroni corrections determined that participants' performance was significantly lower for the mood disorders text without bold key terms (M = 65.60, SE = 3.85) compared to both the mood disorder text with bold key terms (M = 78.75, SE = 4.05), $F(1, 68) = 5.53$, $p = .02$, $\eta_p^2 = .08$, and the anxiety disorders text without bold key terms (M = 86.31, SE = 3.85), $F(1, 68) = 14.44$, $p < .001$, $\eta_p^2 = .08$

Figure 3. First text performance by topic and key term font



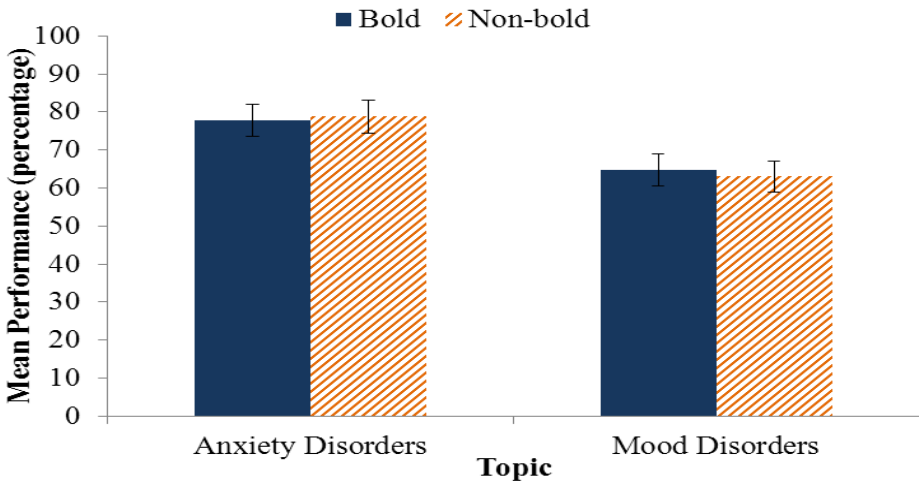
Second text. For the second text position, a 2 x 2 x 2 ANOVA was also conducted using reading ability (high vs. low), topic (anxiety disorders vs. mood disorders) and font (bold vs. non-bold) as between-subjects factors. Results revealed only significant main effects of reading ability, $F(1, 68) = 6.62$, $p = .01$, $\eta_p^2 = .09$, and topic, $F(1, 68) = 11.82$, $p < .001$, $\eta_p^2 = .15$.Font

was not significant, $F(1, 68) = 0.11$, $p = .92$, $\eta_p^2 < .01$ and there were no significant interactions.

Like the first text position, pairwise comparisons with Bonferroni corrections determined that high reading ability participants' performance ($M = 76.46$, $SE = 2.98$) was significantly better than low reading ability participants' ($M = 65.66$, $SE = 2.96$) and performance was better for the anxiety disorders test ($M = 78.27$, $SE = 3.04$) compared to the mood disorders test ($M = 63.84$, $SE = 2.90$)

These results provide further confirmation that individuals with better reading skills continue to perform better than participants with lesser reading skills (regardless of order, topic, or font) and that overall performance on the anxiety disorders test is superior to performance on the mood disorders test. Compared to the first text, there was not a significant interaction between topic and key term font for the second text (see Figure 4).

Figure 4. Second text performance by topic and key term font.



Question Type Performance. In light of the fact that performance was significantly different between topics and bold key terms significantly improved performance on the mood disorders test, a series of four 2 x 2 x 2 univariate tests were conducted using reading ability (high vs. low), key term font (bold vs. non-bold), and order (first text vs. second text) as

between-subject factors to determine if performance for specific types questions (term-specific vs. inference-based) were different based on reading topic (see Table 1).

Results of these analyses revealed only significant differences between reading ability groups based on term-specific performance on the anxiety disorders test, $F(1, 68) = 15.79$, $p < .001$, $\eta_p^2 = .18$, and inference-based performance on the mood disorders test, $F(1, 68) = 6.04$, $p = .02$, $\eta_p^2 = .08$. In both cases, lower reading ability participants' performance on the respective topic-question type (anxiety term-specific: $M = 3.41$, $SE = .16$; mood inference-based: $M = 2.40$, $SE = .19$) was significantly lower than higher reading ability participants' (anxiety term-specific: $M = 4.32$, $SE = .16$; mood inference-based: $M = 3.05$, $SE = .19$). The results of these analyses provide further evidence that performance was affected by difference in difficulty between topics and the implications of these findings will be discussed more in depth below.

Table 1. Mean Performance by Question Type and Reading Ability

	Anxiety Disorders		Mood Disorders	
	Term-specific	Inference-based	Term-specific	Inference-based
Reading Ability	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
Low	3.39 (1.03)	4.29 (.77)	3.68 (1.07)	2.39 (1.10)
High	4.29 (.87)	4.53 (.73)	4.08 (1.05)	2.92 (1.15)

Global Judgments of Learning: Prediction and Postdiction Bias

Prediction and postdiction bias was calculated for each text by subtracting participants' performance from each of their corresponding global JOLs. As a reminder, predictions occurred immediately following each reading, whereas, postdictions occurred immediately following testing. For both biases, positive values indicate participants' were overconfident in their abilities (meaning their JOLs were larger than their actual performance) and negative values indicate they were underconfident (meaning their JOLs were smaller than their actual performance).

First text biases. Prediction and postdiction bias were analyzed using two separate

ANOVAs with reading ability (high vs. low), topic (anxiety disorders vs. mood disorders) and font (bold vs. non-bold) as between-subjects factors (see Figure 6 for a comparison of prediction and postdiction bias means by topic and reading ability).

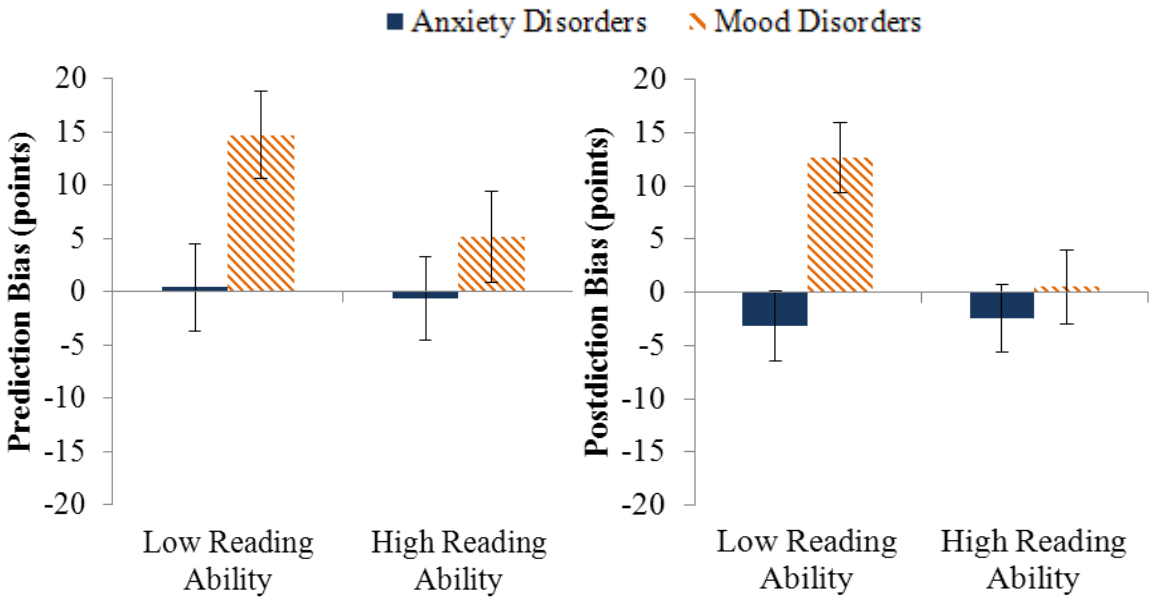
For prediction bias, there was a significant main effect of topic, $F(1, 68) = 5.96$, $p = .08$, $\eta_p^2 = .02$. There was no significant effects of reading ability, $F(1, 68) = 1.68$, $p = .20$, $\eta_p^2 = .02$, or font, $F(1, 68) = .88$, $p = .35$, $\eta_p^2 = .01$, and no interactions were significant. A pairwise comparison with a Bonferroni correction revealed prediction bias for anxiety disorders ($M = -.09$, $SE = 2.83$) was significantly different from prediction bias for mood disorders ($M = 9.90$, $SE = 2.96$). Based on these results, participants who read anxiety disorders first were accurate predicting their performance, whereas, the participants who read mood disorders were very overconfident predicting their performance (see Figure 5).

Analysis of first text postdiction bias also revealed a significant main effect of topic, $F(1, 68) = 8.91$, $p = .004$, $\eta_p^2 = .12$. There was not a significant effect of reading ability, $F(1, 68) = 3.48$, $p = .07$, $\eta_p^2 = .05$, or font, $F(1, 68) = .07$, $p = .79$, $\eta_p^2 < .01$, on postdiction bias; however, there was a significant interaction between reading ability and topic, $F(1, 68) = 4.37$, $p = .04$, $\eta_p^2 = .06$. Pairwise comparisons with Bonferroni corrections for postdiction bias revealed a significant difference between participants' bias for anxiety disorders ($M = -2.79$, $SE = 2.29$) compared to mood disorders ($M = 7.09$, $SE = 2.39$).

Regarding the two-way interaction observed between reading ability and topic, analyses determined that low reading ability participants' bias for mood disorders ($M = 13.63$, $SE = 3.30$) was significantly different compared to high reading ability participants' bias for the same text ($M = .54$, $SE = 3.47$), $F(1, 68) = 7.49$, $p = .01$, $\eta_p^2 = .10$, and low reading ability participants' bias

for anxiety disorders (M = -3.16, SE = 3.30), $F(1, 68) = 12.97$, $p = .001$, $\eta_p^2 = .16$.

Fi*gure* 5. Comparison of first text prediction and postdiction bias by reading ability and topic



The implications of this analysis are that key term font does not affect participants' global JOLS, but topic difficulty does. For low reading ability participants, they were significantly overconfident predicting and postdicting their performance for the harder topic (mood disorders). Even though they adjusted their overconfidence following testing, they did not adjust their bias enough compared to higher reading participants (hence, the interaction observed for postdicted performance).

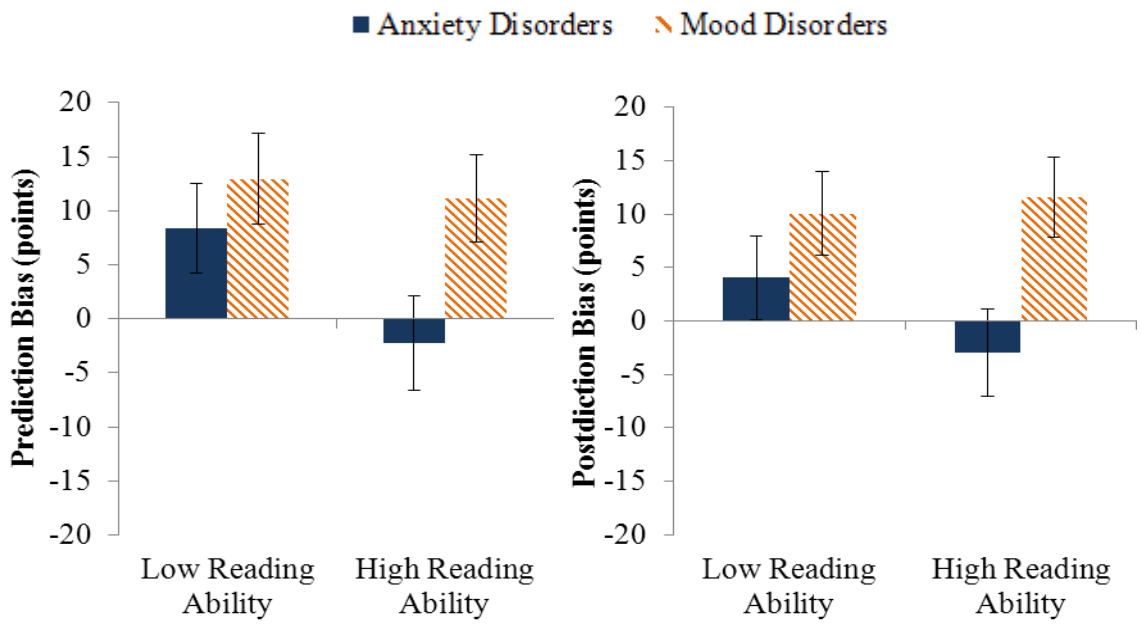
Second text biases. Based on the second text position, prediction and postdiction bias were also analyzed using two separate one-way ANOVAs with reading ability (high vs. low), topic (anxiety disorders vs. mood disorders) and font (bold vs. non-bold key terms) as between-subjects factors (see Figure 6).

Analysis of prediction bias revealed a significant main effect of topic, $F(1, 68) = 4.63$, $p = .04$, $\eta_p^2 = .03$. There were no significant main effects of font, $F(1, 68) = .11$, $p = .74$, $\eta_p^2 < .01$,

or reading ability, $F(1, 68) = 2.23$, $p = .14$, $\eta_p^2 = .03$, and there were no significant interactions. Pairwise comparison with a Bonferroni correction determine that prediction bias for anxiety disorders ($M = 3.07$, $SE = 3.02$) was significantly different than prediction bias for mood disorders ($M = 12.06$, $SE = 2.89$)

For postdiction bias, there was significant main effect of topic, $F(1, 68) = 6.90$, $p = .01$, $\eta_p^2 = .09$. There were no significant main effects of font, $F(1, 68) = .11$, $p = .73$, $\eta_p^2 < .01$, or reading ability, $F(1, 68) = .50$, $p = .48$, $\eta_p^2 < .01$, or any significant interactions. A pairwise comparison with a Bonferroni correction determine that prediction bias for anxiety disorders ($M = .53$, $SE = 2.84$) was significantly different than prediction bias for mood disorders ($M = 10.84$, $SE = 2.71$)

Figure 6. Comparison of second text prediction and postdiction bias by reading ability and topic.



The outcome of the analysis of second text bias is that with the exception of high reading ability participants' biases when reading the easier text (anxiety disorders), participants were consistently overconfident predicting their performance, which suggests that participants did not

account for or anticipate the effects of fatigue on their performance. However, the argument can be made that high reading ability participants' overconfidence for mood disorders may not be an effect entirely of fatigue, but also an indirect consequence of the amount of effort they exerted when reading the mood disorders text.

Term-specific JOLs

Average term-specific JOLs by font. Because two of the hypotheses in the current study were concerning the magnitude of term-specific JOLs in relation to font, two separate analyses were conducted.

To compare the overall magnitude of the average bold term-specific JOLs to the average non-bold term-specific JOLs for each participant, a paired sample T-test was conducted. Contrary to the hypothesis that JOLs for bold terms would be significantly larger than JOLs for non-bold terms, results indicated that there were no significant differences between bold (M = 81.81, SD = 13.49) and non-bold JOLs (M = 82.47, SD = 13.00), t (75) = -.31, p = .76.
For the second analysis, a 2 x 2 repeated measures analysis was conducted with the average term-specific JOL based on font type as a within-subjected variable and reading ability (high vs. low) as a between subjects variable. Results revealed no significant main effects of font-type on the average of term-specific JOLs, F(1, 74) = .09, p = .76, η_p^2 < .01, and no significant interaction between font and reading ability, F(1, 74) = .04, p = .85, η_p^2 < .01 (see Table 2).

Table 2: Comparison of Term-Specific Judgments of Learning (JOLs) by Reading Ability and Key Term Font

	Bold	Non-bold
Reading Ability	*M* (*SD*)	*M* (*SD*)
Low (*n* = 38)	81.81 (13.49)	82.47 (13.01)
High (*n* = 38)	81.99 (8.47)	82.14 (12.89)

Although it was hypothesized that low reading ability participants term-specific JOLs for

bold key terms would be significantly larger than all other groups' term-specific JOLs because they are more likely to be influenced by perceptual cues when predicting their performance, the current study failed to find evidence to support this hypothesis. Planned comparisons with Bonferroni corrections determine that there were no significant differences between reading ability groups for bold term-specific JOLs, $F(1, 74) = 2.18$, $p = .14$, $\eta_p^2 = .03$. In fact, there were no significant differences between low reading ability participants' average bold and non-bold term-specific JOLs, $F(1, 74) = .12$, $p = .73$, $\eta_p^2 < .01$.

Gamma. An attempt to calculate gamma correlations by font type for every participant was made by comparing each individual term-specific JOL with the responses to the corresponding term-specific questions. For texts with bold key terms, gammas could not be calculated for 31 participants (26 out of 31 cases were due to participants answering all term-specific questions correctly); for texts with non-bold key terms, gammas could not be calculated for 25 participants (18 out of 25 cases were due to participants answering all term-specific questions correctly).

Out of 76 participants, only 33 participants had gammas calculated for both fonts (23 low reading ability and 10 high reading ability). For a breakdown of gammas by font type and reading ability, please refer to Table 3.

Table 3: Gamma Correlations Based on Key Term Font and Reading Ability

	Gammas	
Reading Ability	Non-bold	Bold
Low	*M* = .18 *SD* = .69 *n* = 30	*M* = .21 *SD* = .84 *n* = 29
High	*M* = .48 *SD* = .66 *n* = 21	*M* = .11 *SD* = .92 *n* = 16
Total	*M* = .30 *SD* = .69 *n* = 51	*M* = .17 *SD* = .86 *n* = 45

Despite the small proportion of participants with gammas calculated for both fonts, a 2 x 2 repeated measures analysis with gamma type (bold vs. non-bold) as a within-subjects factor and reading ability (high vs. low) as between subjects variable was conducted to determine whether gammas varied due to key term font and/or reading ability.

Results of this analysis determine that there were no significant effects of gamma type, $F(1, 31) = .20$, $p = .66$, $\eta_p^2 < .01$, or any significant interactions between reading ability and gamma type, $F(1, 31) = .92$, $p = .35$, $\eta_p^2 = .03$. Although there was not a significant interaction, pairwise comparisons with Bonferrroni corrections were computed to determine if gamma type varied significantly within a reading ability group (meaning participants' gammas were biased more based on a specific key term font). This analysis revealed that within the reading ability groups, neither low or high reading ability participants' gammas significantly varied between font types, $F(1, 31) = .22$, $p = .64$, $\eta_p^2 = .01$ and $F(1, 31) = .70$, $p = .41$, $\eta_p^2 = .02$, respectively.

Because there was not a significant difference between gamma within reading ability groups, two separate 2 x 2 x 2 one-way ANOVAs were conducted comparing the each gamma

type (bold gamma vs. non-bold gamma) by reading ability (high vs. low), topic (anxiety disorders vs. mood disorders), and order (first text vs. second text). The benefit to conducting the current analysis in this manner is that within a specific gamma type, the data for all participants who had a gamma calculated for that specific gamma type (i.e. 51 non-bold gammas and 45 bold gammas) could be compared based on reading ability, topic, and order.

Results of these analysis determined that for non-bold gammas, there was a significant main effect of topic, $F(1, 42)= 4.76$, $p = .04$, $\eta_p^2 = .10$. Reading ability, $F(1, 42)= .41$, $p = .53$, $\eta_p^2 = .01$, and order, $F(1, 42) = .88$, $p = .36$, $\eta_p^2 = .02$, were not significant and there were no significant interactions. For bold gammas, there were no significant main effects of topic, $F(1, 36) = .32$, $p = .58$, $\eta_p^2 = .01$, reading ability, $F(1, 36) = .12$, $p = .74$, $\eta_p^2 < .01$, order, $F(1, 36) = .04$, $p = .84$, $\eta_p^2 < .001$, or any significant interactions observed.

Based on the main effect of topic on non-bold gammas, a pairwise comparison with a Bonferroni correction revealed that gammas for participants who read mood disorders without bold key terms ($n = 27$, $M = .58$, $SE = .16$) were significantly higher than participants who read anxiety disorders without bold key terms ($n = 23$, $M = .08$, $SE = .16$). Based on these findings, it would appear that the term-specific JOLs made when mood disorders had no bold key terms were more accurate overall, meaning that terms given higher JOLs had a higher probability of being answered correctly.

Reliability: Anxiety Disorders and Mood Disorders Tests

Cronbach's alpha was calculated for both multiple choice tests to assess the reliability of these measures. Results were $\alpha = .455$ for anxiety disorders and $\alpha = .528$ for mood disorders. Although neither of these tests would be considered very reliable, the discrepancy in mean performance for anxiety disorder test ($M = 82.50$, $SD = 1.50$) suggest that combination of the

anxiety disorders text and multiple choice test was less difficult than the combination of the mood disorders text and multiple choice test (M = 65.40, SD = 1.90).

Discussion

Results from the current study failed to provide evidence that the font of key terms embedded in texts affects students' confidence of their knowledge for these concepts or their overall test performance. Contrary to what was hypothesized, the magnitude of JOLs for bold key terms was not significantly larger than JOLs for non-bold key terms and the magnitude of lower reading ability participants' term-specific JOLs for bold key terms was not significantly larger than all other reading ability-key term groups. Failure to find evidence to support these hypotheses may have been the result of several confounds present in the current study.

Regardless of key term font and topic, there were significant effects of order throughout the study. Participants' performance was consistently poorer for the second text they read, which could be due to fatigue because they were required to read a practice text and then each of the reading samples. It is worth noting that these order effects are concerning given the length of each reading sample (~ two pages of content). Realistically most students read more than two pages of material at a time when studying and if the onset of fatigue occurs early during this process, the efficacy of their studies could be greatly compromised which may result in poorer retention and performance (especially for less efficient readers).

An additional problem that affected the current study was the effect of topic on performance and bias among groups. The two texts used in the current study were created by dividing a longer text that had been previously piloted into two shorter texts. Although this previous text was determined to be reliable, an unexpected consequence of dividing up this text

was that performance could vary according to topic. Even though every attempt was made to match these samples for length and difficulty, based on the difference in reliability and performance on the multiple choice tests, there is reason to believe that anxiety disorders topic was easier for participants compared to mood disorders.

One plausible explanation for these effects of topic would be that participants may have been more familiar with the content of the anxiety disorders. In the current study, participants were allowed to participant only if they had not covered the content of either topic in their introductory courses; however, prior knowledge was not tested or controlled for. Although prior knowledge is a concern of any reading comprehension study, given the key term concepts in the anxiety disorders text (post-traumatic stress disorder, obsessive-compulsive disorder, etc) and the fact that half of the multiple choice questions related directly to these concepts, participants with more prior knowledge would have significant advantage during testing based on the current methodology.

An analysis of performance based on question type, topic and reading ability determined that low reading ability participants correctly answered fewer term-specific questions for the anxiety disorders test compared to high reading ability participants, but there were no differences for term-specific questions between groups for mood disorders. If difference in responding between topics for term-specific questions are due to difference in prior knowledge, these findings imply that having more prior knowledge and better reading skills (i.e. high reading ability) is an advantage when reading easier texts because these readers are more likely to acquire a comprehensive situation model.

Another important outcome of the analysis of question type is that there was also a significant difference in correct responses for inference-based questions on the mood disorders

test, with higher reading ability participants correctly responding to more inference questions compared to lower reading ability participants. The conclusion that could be made is that when readers have less prior knowledge, being a more efficient reader (even for more difficult texts) can compensate for deficits in prior knowledge by allowing readers to generate predictive inferences within the text. Unfortunately, because prior knowledge was not evaluated, the extent to which this conclusion may be true cannot be determined. However, a follow-up of this portion of the current study may be worth pursing in the future in light of the theoretic implications these findings have regarding the construction-integration model theory.

The other theoretical point of interest in the current study was whether the presence of font type of key terms can affect participants' performance and subsequent biases. With the one exception observed in an interaction between the topic and key term font on performance for the first text participants read (depicted in Figure 4), font type of key terms did not affect performance or bias. Although previous research has demonstrated that font size affects participants JOLs (McDonough & Gallo, 2010; Rhodes & Castel, 2008), failure to detect any potential differences in JOLs due to font may also be due to several methodological differences between this study and previous studies which make it difficult to compare the results to previous research.

Both Rhodes and Castel (2008) and McDonough and Gallo (2010) used word and manipulated the font size of to-be-remembered words by creating a very large contrasts in fonts (12-pt. vs. 48-pt. font) within the lists. In the current study, we elected to examine the effects of font type (bold vs. non-bold) on key terms within texts. Compared to the degree of the perceptual changes used in these previous studies, it is possible that the manipulation in the current study, though perceptually salient, were simply less extreme and thus, less effective. Furthermore,

because bold key terms in texts are quite common compared to font changes within word lists, participants may have pre-existing strategies for reading texts when bold key terms are present. Together, these factors could limit the strength of perceptual salience of bold key terms embedded in text s to evoke changes in students' confidence toward their future performance.

Assuming that the argument above is indeed true, regarding the cue-utilization approach (Koriat, 1997), presenting key terms in bold may not be strong enough of a cue to influence students' JOL when reading. Interestingly, topic difficulty does appear to be an important cue, based on the interaction observed between topic and font on performance for the first reading sample. When participants were not fatigued, the presence of bold key-terms significantly improved performance for difficult texts compared to difficult texts without bold key-terms; thus, bold key terms may make difficult texts easier to read. Nevertheless, the assumption that bold key terms make texts easier to read will need to be further investigated because despite improvements in overall performance, low reading ability participants were still significantly more overconfident predicting and postdicting their first text performance for mood disorder regardless of font; high reading ability participants were slightly overconfident predicting their performance, but corrected their bias following testing.

Another possibility that may need to be considered is that bold font in texts may actually influence students' JOLs more after a delay. In the current study, students' term-specific JOLs were only examined immediately after reading. Although it could be hypothesized that effects of bold font on students' JOL immediately following reading would be more prevalent given the cue-utilization approach, if one takes into account previous findings, it is possible that after a longer delay, students' memory of the presentation style of the concepts could be superior to their actual memory of what the concepts are. Unfortunately, definitive support of this argument

is lacking given prior and current methodologies. Future studies will need to further investigate the effects of font type on JOLs with fewer confounds present in order to provide more definitive conclusions.

In conclusion, in spite of prior findings and the theoretical implications of the cue-utilization approach, the current study did not find evidence to suggest that the presence of bold key terms in texts affects participants' metacognition. However, due to novel nature of the current methodology, additional studies are warranted because of the theoretical implications associated with reading texts with bold key terms and students' metacognition for content within these texts. It is encouraging that bold key terms appear to improve overall performance in some conditions, but if some individuals' persist to be metacognitivly naïve, the effectiveness of these font changes to improve performance may be negated (especially if students prematurely suspend their studies). Therefore, it is important that researchers determine ways to improve both students' performance and metacognition because obtaining proficiency in both domains could permit students to become more efficient and productive in their academics.

References

Britton, B. K., Van Dusen, L., Gülgöz, S., Glynn, S. M., & Sharp, L. (1991). Accuracy of learnability judgments for instructional texts. *Journal of Educational Psychology* 83 (1): 43–47.

Carlson, M. C., Hasher, L., Connelly, S. L., & Zacks, R. T. (1995). Aging, distraction, and the benefits of predictable location. *Psychology and Aging, 10,* 427-436.

Daneman, M., & Carpenter, P. A. (1983). Individual differences in integrating information between and within sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*(4), 561-584.

Dunlosky, J., & Lipko, A. R. (2007). Metacomprehension: A brief history and how to improve its accuracy. *Current Directions in Psychological Science, 16*(4), 228-232.

Dunlosky, J., Rawson, K. A., & Middleton, E. L. (2005). What constrains the accuracy of metacomprehension judgments? Testing the transfer-appropriate-monitoring and accessibility hypotheses. *Journal of Memory and Language, 52,* 551-565.

Gasser, M., Boeke, J., Haffernan, M. & Tan, R. (2005). The influence of font type on information recall. *North American Journal of Psychology, 7*(2), 181-188.

Gernsbacher, M. A. & Varner, K. R. (1988). The multi-media comprehension battery. (Tech. Rep. No. 88-07). Eugene, OR: University of Oregon, Institute of Cognitive and Decision Sciences.

Gernsbacher, M. A., & Varner, K. R., & Faust, M. E. (1990). Investigating differences in general

comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(3), 430-445.

Kemper, S., & McDowd, J. (2006). Eye movements of young and older adults while reading with distraction. *Psychology and Aging, 21*(1), 32-39.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95*(2), 163-182.

Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist, 49*, 294-303.

Kintsch, W. (2005). An overview of top-down and bottom–up effects in comprehension: The CI perspective. *Discourse Processes, 39*(2&3), 125-128.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*(4), 349-370

Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one's knowledge during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(2), 187-194.

Lee, M. J., Tedder, M. C., & Xie, G. (2006). Effective computer text design to enhance readers' recall: Text formats, individual working memory capacity and content type. *Journal of Technical Writing and Communication, 36*, 57-73.

Linderholm, T., Cong, X., & Zhao, Q. (2008). Differences in low and high working-memory capacity readers' cognitive and metacognitive processing patterns as a function of reading for different purposes. *Reading Psychology, 29*, 61-85.

Linderholm, T., & Broek, P. V. D. (2002). The effects of reading purpose and working memory capacity on the processing of expository text. *Journal of Educational Psychology, 94*,

778-784.

Maki, W. S., & Maki, R. H. (2002). Multimedia comprehension skill predicts differential outcomes of web-based and lecture courses. *Journal of Experimental Psychology: Applied*, *8*(2), 85-98.

Maki, R. H., Shields, M., Wheeler, A. E., & Zacchilli, T. L. (2005). Individual differences in absolute and relative metacomprehension accuracy. *Journal of Education Psychology*, *97*, 723-731.

Maki, R. H., Willmon, C., & Pietan, A. (2009). Basis of metamemory judgments for text with multiple-choice, essay, and recall tests. *Applied Cognitive Psychology*, *23*, 204-222.

McDonough, I. M., & Gallo, D. A. (2011). Illusory expectation can affect retrieval-monitoring accuracy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 1-14.

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), 1-43.

Metcalfe, J. (2009). Metacognitive judgments and control of study. *Current Directions in Psychological Science*, *18*, 159-163.

Metcalfe, J., and Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review*, *15*, 174-179.

Rawson, K. A., & Dunlosky, J. (2007). Improving students' self-evaluation of learning for key concepts in textbook materials. *European Journal of Cognitive Psychology*, *19*, 559-579.

Rayner, K., Slattery, T. J., & Bélanger, N. N. (2010). Eye movements, the perceptual span, and reading speed. *Psychonomic Bulletin & Review*, *17*(6), 834-839

Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual

information: Evidence for metacognitive illusions. *Journal of Experimental Psychology: General, 137,* 615-625.

Thiede, K. W., Griffin, T. D., Wiley, J., & Anderson, M. C. M. (2010). Poor metacomprehension accuracy as a result of inappropriate cue use. *Discourse Processes, 47,* 331-362.

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review, 114,* 104-132.

Wiley, J., Griffin, T. D., & Thiede, K. W. (2005). Putting the comprehension in metacomprehension. *The Journal of General Psychology*, *132*(4), 408-428.

Appendix A

Anxiety Disorders Reading Sample

Generalized Anxiety Disorder

For the past two years, Tom, a 27-year-old electrician, has been bothered by dizziness, sweating palms, heart palpitations, and ringing in his ears. He feels edgy and sometimes finds himself shaking. With reasonable success, he hides his symptoms from his family and co-workers. But he allows himself few other social contacts, and occasionally he has to leave work. His family doctor and a neurologist can find no physical problem.

Tom's unfocused, out-of-control, negative feelings suggest **generalized anxiety disorder**. The symptoms of this disorder are commonplace; their persistence is not. People with this condition (two-thirds are women) worry continually, and they are often jittery, agitated, and sleep-deprived. Concentration is difficult, as attention switches from worry to worry, and their tension and apprehension may leak out through furrowed brows, twitching eyelids, trembling, perspiration, or fidgeting.

One of the worst characteristics of this disorder is that the person cannot identify, and therefore cannot deal with or avoid, its cause. To use Sigmund Freud's term, the anxiety is *free-floating.* Generalized anxiety disorder is often accompanied by depressed mood, but even without depression it tends to be disabling (Hunt et al., 2004 Moffitt et al., 2007b). Moreover, it may lead to physical problems, such as high blood pressure.

Many people with generalized anxiety disorder were maltreated and inhibited as children (Moffitt et al., 2007a). As time passes, however, emotions tend to mellow and by age 50, generalized anxiety disorder becomes rare (Rubio & López-Ibor, 2007).

Panic Disorder

Panic disorder is an anxiety tornado. It strikes suddenly, wreaks havoc, and disappears. For the 1 person in 75 with this disorder, anxiety suddenly escalates into a terrifying *panic attack*—a minutes-long episode of intense fear that something horrible is about to happen. Heart palpitations, shortness of breath, choking sensations, trembling, or dizziness typically accompany the panic, which may be misperceived as a heart attack or other serious physical ailment. Smokers have at least a doubled risk of panic disorder (Zvolensky & Bernstein, 2005). Because nicotine is a stimulant, lighting up doesn't lighten up.

One woman recalled suddenly feeling "hot and as though I couldn't breathe. My heart was racing and I started to sweat and tremble and I was sure I was going to faint. Then my fingers started to feel numb and tingly and things seemed unreal. It was so bad I wondered if I was dying and asked my husband to take me to the emergency room. By the time we got there (about 10 minutes) the worst of the attack was over and I just felt washed out" (Greist et al., 1986).

Phobias

Phobias are anxiety disorders in which an irrational fear causes the person to avoid some object, activity, or situation. Many people accept their phobias and live with them, but others are incapacitated by their efforts to avoid the feared situation. Marilyn, an otherwise healthy and happy 28-year-old, so fears thunderstorms that she feels anxious as soon as a weather forecaster

mentions possible storms later in the week. If her husband is away and a storm is forecast, she may stay with a close relative. During a storm, she hides from windows and buries her head to avoid seeing the lightning.

Other *specific phobias* may focus on animals, insects, heights, blood, or close spaces. People avoid the stimulus that arouses the fear, hiding during thunderstorms or avoiding high places.

Not all phobias have such specific triggers. *Social phobia* is shyness taken to an extreme. Those with a social phobia, an intense fear of being scrutinized by others, avoid potentially embarrassing social situations, such as speaking up, eating out, or going to parties—or will sweat, tremble, or have diarrhea when doing so.

People who have experienced several panic attacks may come to fear the fear itself and avoid situations where the panic has struck before. If the fear is intense enough, it may become *agoraphobia,* fear or avoidance of situations in which escape might be difficult or help unavailable when panic strikes. Given such fear, people may avoid being outside the home, in a crowd, on a bus, or on an elevator.

After spending five years sailing the world, Charles Darwin began suffering panic disorder at age 28. Because of the attacks, he moved to the country, avoided social gatherings, and traveled only in his wife's company. But the relative seclusion did free him to focus on developing his evolutionary theory. "Even ill health," he reflected, "has saved me from the distraction of society and its amusements" (quoted in Ma, 1997).

Obsessive-Compulsive Disorder

As with generalized anxiety and phobias, we can see aspects of our own behavior in **obsessive-compulsive disorder (OCD)**. We may at times be obsessed with senseless or offensive thoughts that will not go away. Or we may engage in compulsive behaviors, rigidly checking, ordering, and cleaning before guests arrive, or lining up books and pencils "just so" before studying.

Obsessive thoughts and compulsive behaviors cross the fine line between normality and disorder when they persistently interfere with everyday living and cause the person distress. Checking to see you locked the door is normal; checking 10 times is not. Washing your hands is normal; washing so often that your skin becomes raw is not. At some time during their lives, often during their late teens or twenties, 2 to 3 percent of people cross that line from normal preoccupations and fussiness to debilitating disorder (Karno et al., 1988). The obsessive thoughts become so haunting, the compulsive rituals so senselessly time-consuming, that effective functioning becomes impossible.

OCD is more common among teens and young adults than among older people (Samuels & Nestadt, 1997). A 40-year follow-up study of 144 Swedish people diagnosed with the disorder found that, for most, the obsessions and compulsions had gradually lessened, though only I in 5 had completely recovered (Skoog & Skoog, 1999).

Post-Traumatic Stress Disorder

As an Army infantry scout during the Iraq war, Jesse "saw the murder of children, women. It was just horrible for anyone to experience." After calling in a helicopter strike on one house where he had seen ammunition crates carried in, he heard the screams of children from within. "I didn't know there were kids there," he recalls. Back home in Texas, he suffered "real bad flashbacks" (Welch, 2005).

Our memories exist in part to protect us in the future. So there is biological wisdom in not being able to forget our most emotional or traumatic experiences—our greatest embarrassments, our worst accidents, our most horrid experiences. But sometimes, for some of us, the unforgettable takes over our lives. The complaints of battled-scarred veterans such as Jesse—recurring haunting memories and nightmares, a numbed social withdrawal, jumpy anxiety, and insomnia—are typical of what once was called "shellshock" or "battle fatigue" and now is called **post-traumatic stress disorder (PTSD)** (Hoge et al., 2004; Kessler, 2000).

PTSD symptoms have also been reported by survivors of accidents, disasters, and violent and sexual assaults (including an estimated two-thirds of prostitutes) (Brewin et al., 1999; Farley et al., 1998; Taylor et al., 1998). A month after the 9/11 terrorist attacks, a survey of Manhattan residents indicated that 8.5 percent were suffering PTSD, most as a result of the attack (Galea et al., 2002). Among those living near the World Trade Center, 20 percent reported such telltale signs as nightmares, severe anxiety, and fear of public places (Susser et al., 2002).

Appendix B

Mood Disorders Reading Sample

Somatoform Disorders

Among the most common problems bringing people into doctors' offices are "medically unexplained illnesses" (Johnson, 2008). Ellen becomes dizzy and nauseated in the late afternoon—shortly before she expects her husband home. Neither her primary care physician nor the neurologist he sent her to could identify a physical cause. They suspect her symptoms have an unconscious psychological origin, possibly triggered by her mixed feelings about her husband. In **somatoform disorders**, such as Ellen's, the distressing symptoms take a somatic (bodily) form without apparent physical causes. One person may have a variety of complaints—vomiting, dizziness, blurred vision, difficulty in swallowing. Another may experience severe and prolonged pain.

Mood Disorders

The emotional extremes of **mood disorders** come in two principal forms: (1) *major depressive disorder*, with its prolonged hopelessness and lethargy, and (2) *bipolar disorder* (formerly called manic-depressive disorder), in which a person alternates between depression and mania, an overexcited, hyperactive state.

Major Depressive Disorder

If you are like most college students, at some time during this year—more likely the dark months of winter than the bright days of summer—you will probably experience a few of depression's symptoms. You may feel deeply discouraged about the future, dissatisfied with your life, or socially isolated. You may lack the energy to get things done or even to force yourself out of bed; be unable to concentrate, eat, or sleep normally; or even wonder if you would be better off dead. Perhaps academic success came easily to you in high school, and now you find that disappointing grades jeopardize your goals. Maybe social stresses, such as feeling you don't belong or experiencing the end of a romance, have plunged you into despair. And maybe brooding has at times only worsened your self-torment. You are not alone. In one survey of 90,000 American college and university students, 44 percent reported that on one or more occasions within the last school year, they had felt "so depressed it was difficult to function" (ACHA, 2006).

Depression has been called the "common cold" of psychological disorders—an expression that effectively describes its pervasiveness but not its seriousness. Although phobias are more common, depression is the number-one reason people seek mental health services. At some point during their lifetime, depressive disorders have plagued 12 percent of Canadian adults and 13 percent of U.S. adults (Chasm et al., 2005; Patten et al., 2006). Moreover, it is the leading cause of disability worldwide (WHO, 2002). In any given year, a depressive episode plagues 5.8 percent of men and 9.5 percent of women, reports the World Health Organization.

As anxiety is a response to the threat of future loss, depressed mood is often a response to past and current loss. About one in four people diagnosed with depression is simply struggling with the normal emotional impact of a significant loss, such as a loved one's death, a ruptured marriage, a lost job (Wakefield et al., 2007). To feel bad in reaction to profoundly sad events is to be in touch with reality. In such times, depression is like a car's low—oil-pressure light—a

signal that warns us to stop and take protective measures. Recall that, biologically speaking, life's purpose is not happiness but survival and reproduction. Coughing, vomiting, and various forms of pain protect the body from dangerous toxins. Similarly, depression is a sort of psychic hibernation: It slows us down, defuses aggression, and restrains risk taking (Allen & Badcock, 2003). To grind temporarily to a halt and ruminate, as depressed people do, is to reassess one's life when feeling threatened, and to redirect energy in more promising ways (Watkins, 2008). There is sense to suffering.

But when does this response become seriously maladaptive? Joy, contentment, sadness, and despair are different points on a continuum, points at which any of us may be found at any given moment. The difference between a blue mood after bad news and a mood disorder is like the difference between gasping for breath after a hard run and being chronically short of breath.

Major depressive disorder occurs when at least five signs of depression (including lethargy, feelings of worthlessness, or loss of interest in family, friends, and activities) last two or more weeks and are not caused by drugs or a medical condition. To sense what major depression feels like, suggest some clinicians, imagine combining the anguish of grief with the sluggishness of jet lag.

Bipolar Disorder

With or without therapy, episodes of major depression usually end, and people temporarily or permanently return to their previous behavior patterns. However, some people rebound to, or sometimes start with, the opposite emotional extreme— the euphoric, hyperactive, wildly optimistic state of **mania**. If depression is living in slow motion, mania is fast forward. Alternating between depression and mania signals **bipolar disorder**.

Adolescent mood swings, from rage to bubbly, can, when prolonged, produce a bipolar diagnosis. Between 1994 and 2003, U.S. National Center for Health Statistics annual physician surveys revealed an astonishing 40-fold increase in diagnoses of bipolar disorder in those 19 and under—from an estimated 20,000 to 800,000 (Carey, 2007; Moreno et al., 2007). The new popularity of the diagnosis, given in two-thirds of the cases to boys, has been a boon to companies whose drugs are prescribed to lessen mood swings.

During the manic phase of bipolar disorder, the person is typically overtalkative, overactive, and elated (though easily irritated if crossed); has little need for sleep; and shows fewer sexual inhibitions. Speech is loud, flighty, and hard to interrupt. The person finds advice irritating. Yet they need protection from their own poor judgment, which may lead to reckless spending or unsafe sex.

To simulate mania's racing thoughts, such as you may have experienced when excited about a new idea, Emily Pronin and Daniel Wegner (2006) invited students to read a series of statements at either double or half the normal reading speed. Those who had just raced through the material reported feeling happier, more powerful, more energetic, and more creative. A racing mind arouses an upbeat mood.

In milder forms, mania's energy and free-flowing thinking does fuel creativity. George Frideric Handel (1685-1759), who may have suffered from a mild form of bipolar disorder, composed his nearly four-hour-long Messiah during three weeks of intense, creative energy (Keynes, 1980). Robert Schumann composed 51 musical works during two years of mania (1840 and 1849) and none during 1844, when he was severely depressed (Slater & Meyer, 1959). Those who rely on precision and logic, such as architects, designers, and journalists, suffer bipolar disorder less often than do those who rely on emotional expression and vivid imagery,

reports Arnold Ludwig (1995). Composers, artists, poets, novelists, and entertainers seem especially prone (Jamison, 1993, 1995; Kaufman & Baer, 2002; Ludwig 1995).

It is true of emotions as of everything else: What goes up comes down. Before long, the elated mood either returns to normal or plunges into a depression. Though bipolar disorder is much less common than major depressive disorder, it is often more dysfunctional, claiming twice as many lost workdays yearly (Kessler et al., 2006). In adults, it afflicts men and women about equally.

Appendix C

Anxiety Disorders Criterion Test

Term-specific Questions

1) This disorder is characterized by a persistent nervousness and uneasiness that occurs for at least 6 months:
   a) Panic Disorder
   b) Generalized Anxiety Disorder
   c) Major Depression
   d) Post-Traumatic Stress Disorder

2) Irrational fear that causes anxiety is a characteristic of _____?
   a) Obsessions
   b) Phobias
   c) Anxiety Disorders
   d) Compulsions

3) Which of the following is not a characteristic of panic disorder?
   a) Sudden Onset
   b) Occurs Without Warning
   c) Intense Feelings of Terror
   d) Has a Specific Cause

4) OCD stands for _____?
   a) Obsession-Compulsion Disorder
   b) Obsessive-Compulsive Disorder
   c) Obsessive-Counting Disorder
   d) Obsession-Compulsive Disorder

5) Re-accruing nightmares, social withdrawal, and anxiety are symptoms of _____?
   a) Phobias
   b) PTSD
   c) Bipolar
   d) Depression

Inference-based Questions

6) Maggie worries continuously about a variety of things and has difficulty sleeping at night. Maggie most likely suffers from:
   a) Insomnia
   b) Specific Phobias
   c) Generalized Anxiety
   d) Panic

7) Which of the following statements is true of people who suffer from social phobias?
   a) These individuals are terrified of being alone and prefer to be in large crowds of people
   b) These individuals tend to display inappropriate behaviors in public
   c) These individuals fear being scrutinized by others and try to avoid social situations
   d) These individuals' fears cause other people to be afraid of them

8) OCD, PTSD, and Phobias are examples of_____?

a) Fear Disorders
c) Mood Disorders
b) Social Disorders
d) Anxiety Disorders

9) In obsessive-compulsive disorder, compulsions are:

a) Thoughts
c) Behaviors
b) Fears
d) Illogical Dreams

10) Jill is afraid of heights and refuses to go on a rollercoaster with her friends. Jill's behavior is an example of a _____ ?

a) Specific phobia
c) Specific anxiety
b) Specific compulsion
d) Specific PTSD

Correct Answers:1) b 2) b 3) d 4) b 5) b 6) c 7) c 8) d 9) c 10) a

Appendix D

Mood Disorder Criterion Test

Term-specific Questions

1) What are the major types of mood disorders?
    a) Major Depressive Disorder and Bipolar Disorder
    b) Generalized Anxiety Disorder and PTSD
    c) Mania Disorder and Depressive Disorder
    d) Panic Disorder, Phobias, and Bipolar

2) Lethargy, lack of interests, and feelings of worthlessness that last two weeks or more are symptoms of _____?
    a) Antisocial Disorder
    b) Major Depressive Disorder
    c) Depression Disorder
    d) Bipolar Disorder

3) States of hyperactivity, described as upward mood changes, are examples of _____?
    a) Nerves
    b) Mania
    c) Anxiety
    d) Jitters

4) The two main symptoms of bipolar disorder are_____ and _____?
    a) Depression and Anxiety
    b) Mania and Depression
    c) Fear and Mania
    d) Anxiety and Fear

5) Somatoform disorders are characterized by _____?
    a) Intense feelings of depression
    b) Loss of sleep and appetite
    c) Rapid changes in mood
    d) Symptoms without apparent causes

Inference-based Questions

6) What disorder is the most common reason people seek mental health services?
    a) Bipolar
    b) Manic
    c) Somatoform Disorder
    d) Depression

7) Which statement about depression is true?
    a) Depression is a response to threats of future losses
    b) Depression is a biologically adaptive response

c) Depression occurs after individuals experience anxiety
d) Depression occurs more often in males than females

8) Somatic symptoms can be characterized by _____?
   a) Depression
   b) Re-acquiring fears
   c) Inability to sleep
   d) Physical ailments

9) Creativity is often a byproduct of _____?
   a) Bipolar Disorder
   b) Somatoform Disorder
   c) Depression
   d) Major Depressive Disorder

10) Which statement about bipolar disorder is not true _____?
   a) It is more common than major depressive disorder
   b) Men and women are equally affected by it
   c) Individuals cycle from “highs” and “lows”
   d) Drug companies have benefit from increases diagnosis

Correct Answers:1) a 2) b 3) b 4) b 5) d 6) d 7) b 8) d 9) a 10) a