12-9-2016

The Impact of Metacomprehension Accuracy on Control Processes during Comprehension

Elaine Wei-Ling Tan

The impact of metacomprehension accuracy on control processes during comprehension

By

Elaine Wei-Ling Tan

A Dissertation
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
in Cognitive Science
in the Department of Psychology

Mississippi State, Mississippi

December 2016

The impact of metacomprehension accuracy on control processes during comprehension

By

Elaine Wei-Ling Tan

Approved:

Deborah K. Eakin
(Major Professor)

Jarrod Moss
(Committee Member)

Gary L. Bradshaw
(Committee Member)

Wendy Herd
(Committee Member)

Kevin J. Armstrong
(Graduate Coordinator)

Rick Travis
Interim Dean
College of Arts & Sciences

Name: Elaine Wei-Ling Tan

Date of Degree: December 9, 2016

Institution: Mississippi State University

Major Field: Cognitive Science

Committee Chair: Deborah K. Eakin

Title of Study: The impact of metacomprehension accuracy on control processes during comprehension

Pages in Study 294

Candidate for Degree of Doctor of Philosophy

The aim of this dissertation was to investigate whether improving metacomprehension accuracy via the monitoring process impacted learning strategy selection implemented by the control process so that comprehension was also improved. A new paradigm—the multi trial metacomprehension paradigm—was introduced to investigate this aim. Participants studied a text using an effective or ineffective learning strategy, made metacomprehension predictions about their future comprehension, and took a comprehension test; there were three trials of this procedure. The goal was to determine whether metacomprehension accuracy improved—leading to improved comprehension accuracy—for the third trial.

Experiment 1 tested whether metacomprehension accuracy improved across multiple trials when compared against single trials. Although no difference in metacomprehension accuracy between multiple and single trial conditions was found, comprehension accuracy did improve with multiple trials. However, for a subset of participants whose metacomprehension accuracy across trials did improve, their comprehension accuracy also improved. Although there was no effect of learning strategy on either metacomprehension accuracy or comprehension accuracy overall, the effective learning strategy produced the highest metacomprehension accuracy on the first trial, leaving no room for improvement at later trials. Metacomprehension accuracy only improved when using the ineffective learning strategy if it was used on multiple trials, but never to the same degree as when using an effective learning strategy.

Experiment 2 tested whether improved metacomprehension accuracy affected the control process of learning strategy selection by allowing participants to select which learning strategy to use during the third trial. Participants overwhelmingly selected the ineffective learning strategy, even in case in which metacomprehension accuracy improved across trials. This finding calls into question the theory that improved monitoring accuracy informing the meta level leads to better implementation of control process on the object level. However, while metacomprehension accuracy might be necessary to improve comprehension accuracy—and to result in selection of effective learning strategies toward that end—it might not be sufficient. Students should not just be told to use an effective learning strategy; they should also be taught how to use cues during the monitoring process that are diagnostic of future comprehension.

Keywords: metacomprehension, comprehension, reading comprehension, metacognitive monitoring, self-regulated study.

DEDICATION

For my father, whose wisdom and resilience taught me to have the grit I needed for this journey. You will always be close to my heart.

For my mother, whose love and unwavering support kept me from giving up countless times. You are my supermum.

ACKNOWLEDGEMENTS

I would like to first express my gratitude to my advisor, Deborah Eakin, who believed in me, provided honest feedback, and invested six long years mentoring me to be the better scientist that I am today. Thank you, Doc!

I would like to thank my committee members, Jarrod Moss, Gary Bradshaw, and Wendy Herd, for their advice and encouragement throughout my dissertation journey. I would also like to thank Carrick Williams for his support from the moment I landed in Mississippi, thank you for always asking the tough questions.

I would like to thank my fellow scientists in the Eakin Memory and Metamemory Lab, both past and present, for being part of my graduate school support system. To the undergraduate research assistants who diligently helped with materials generation, data collection and processing, I appreciate your hard work.

Last but not least, I would like to acknowledge the steadfast cheer and love I received from my Malaysian and American families. To mum, Alicia, Evelyn, Karla, Alexis, my darling Drew, and my guardian Syd, you are my heart. I cannot thank each one of you enough. Thank you for your faith in me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Assessing one's own comprehension is called metacomprehension and is a crucial component to the comprehension process. Metacomprehension accuracy is the degree to which judgments of comprehension match actual comprehension, and is represented using a correlation score ranging from -1 (perfectly inaccurate) to +1 (perfectly accurate). A shared conclusion is clear from a review of the metacomprehension literature; people are generally poor at assessing the degree to which they comprehend texts (Dunlosky & Lipko, 2007; Glenberg & Epstein, 1985; Glenberg, Sanocki, Epstein, & Morris, 1987; Maki, 1998b; Maki & Serra, 1992; Thiede & Anderson, 2003; Thiede, Anderson, & Therriault, 2003).

A meta-analysis (Fukaya, 2010) conducted using data from 63 different studies varying across manipulations showed that metacomprehension accuracy peaked at +.27. In addition, Maki (1998a) reported the average monitoring accuracy across 25 studies conducted within her laboratory to be at +.27. With such a low correlation between predicted and actual comprehension, a hypothetical student studying two texts to varying levels of comprehension would only be 14% better than chance, on average, at predicting their comprehension level accurately for each text (Fukaya, 2010). These are not new findings. Dating back to the initial metacomprehension paradigm introduced by Glenberg and Epstein (1985), people have been shown to be less than accurate at assessing their comprehension. These results have a certain consequence for text comprehension. If students are unable to accurately assess their comprehension, they will not be able to effectively make decisions about what learning strategies to use while reading. To improve comprehension, students need to periodically assess their current level of comprehension while reading and use that assessment to decide whether a sufficient level of comprehension has been achieved. If the desired level of comprehension has not been met, based on their assessment, they must decide whether adjustments need to be made to their current learning strategy. For example, if a student assesses that they have adequately learned the material from the text, they might make a decision to stop studying. However, if a student assesses that they do not fully understand the text, they might decide to increase study time and change their learning strategy from highlighting to rereading (Dunlosky & Lipko, 2007; Metcalfe, 2009). Clearly, understanding the processes that occur between metacomprehension and comprehension is crucial toward improving comprehension.

The experiments in this dissertation aimed to investigate whether students were able to improve their metacomprehension accuracy via repeated trials and experience with learning strategies of various levels of effectiveness. Additionally, the experiments aimed to investigate whether improved metacomprehension accuracy subsequently led to selection of a learning strategy that also improved comprehension. As will be discussed next, the relationship between metacomprehension and comprehension is a dynamic one, with both processes interacting to impact each other.

Metacomprehension

Following the theory proposed by Nelson and Narens (1990) to explain the relationship between metamemory and memory, the process of metacomprehension involves the interaction of two levels: a) the object level and b) the meta level. For text comprehension, the object level is the actual comprehension process at the cognitive level, including the actual strength of the memory traces for information from the text. For instance, a text on “Alien Planets” might include, “There are out there in the depth of space. There are giant ones, small ones, weird ones, and most likely ones we can’t even imagine.” Memory traces of information from the text, such as “alien planets exists” and “some planets are unimaginable”, now exist at the object level; some traces are stronger than others. The state of comprehension at the object level is then monitored by the monitoring process which informs the meta level.

Theoretically, direct access to the comprehension process at the object level is not possible (Koriat, 1998; Koriat, 2000). Instead, object-level processes are represented at the meta level; this representation is informed by the monitoring process. Some examples of the object-level processes that are monitored by the monitoring process include cues pertaining to the characteristics of the text, the level of encoding of the text, and assessments of whether a piece of information will be retained in memory. The monitoring process is one of two processes that operate between the object and meta levels. The monitoring process continuously assesses the state of the object level and makes inferences about the state of the object level based on that information. This information is received at the meta level, which uses this information to update its representation of the state of text comprehension. (The factors that are available to monitoring will be discussed in Chapter II.)

Based on the information received via monitoring of the state of comprehension of the text at the object level, the meta level representation of that state of comprehension is updated. For example, assessments of the state of comprehension such as, “This will take time to learn” or “This is difficult information”, will now inform the meta level. Based on these memory assessments during monitoring, decisions about how to change the state of the object level, such as which learning strategy to use, will be made and implemented by the control process. Therefore, if given the opportunity to study again, the control process might select a more effective learning strategy to restudy the text if it was assessed to be not well learned. Depending on the degree to which the state of comprehension as represented at the meta level matches with the reader’s goal state of comprehension, the meta level implements processes to alter the state of comprehension at the object level via the control process. The control process operates on the object level via processes such as the continuation or termination of study and the application of learning strategies that could modify the object level toward improved comprehension. The relationship between monitoring and control is a dynamic and ongoing process during the course of comprehension, and the interplay amongst these processes result in metacomprehension (Nelson & Narens, 1990). Figure 1 depicts the dynamic relationship between the object and meta level and the monitoring and control processes involve when studying a text called “Alien Planets”.

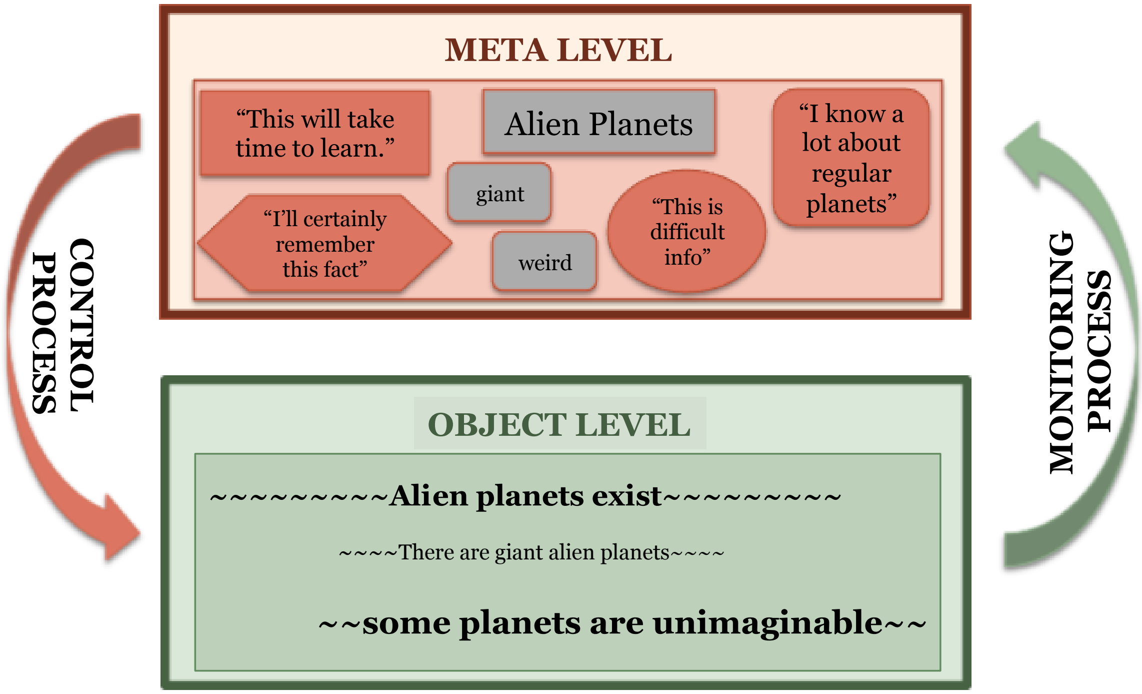

Figure 1. The relationship between object and meta level during the comprehension process

An example of what occurs at the object and meta level assisted by the monitoring and control process while reading the text "Alien Planets".

The Monitoring Process in Text Comprehension

Monitoring is measured during a special metacomprehension phase as part of the typical metacomprehension paradigm. This paradigm starts with reading a text during a comprehension phase. Then, during the phase measuring the monitoring process, metacomprehension assessments are elicited about the degree to which the content of the text has been understood and learned. Later a comprehension test is given that tests that understanding. Metacomprehension accuracy is determined by the correlation of the assessment of and the actual test measure of comprehension.

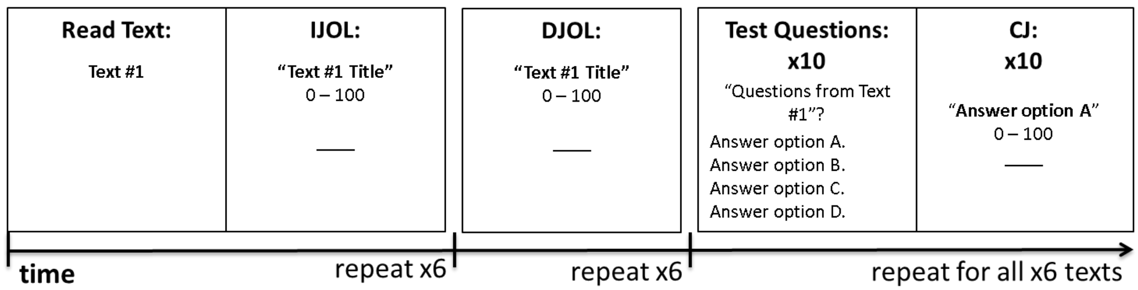

Figure 2. A basic metacomprehension paradigm used to monitor the comprehension process

An example of a typical metacomprehension paradigm with both IJOLs and DJOLs prediction phases.

There are a variety of types of metacomprehension assessments, depending on when they are elicited during the metacomprehension paradigm. *Judgments of Learning* (JOL) are assessments made about the degree to which the text has been learned at the time of assessment. The type of JOL varies, depending on both the timing and granularity of these judgments. In terms of timing, *immediate* JOLs (IJOLs) are made immediately after the reading of the text, whereas *delayed* JOLs (DJOLs) are made during the retention interval between study and test. In terms of granularity, *global* judgments are made about comprehension of the entire text, whereas *absolute* judgments are made about individual questions or idea units.

Maki (1998a) conducted an experiment with a series of timing manipulations between the reading comprehension and metacomprehension judgment phases of each text. In the experiment, Maki (1998a) had participants read a total of twelve texts. Half of the participants were assigned to the immediate-judgment condition; immediately after reading each text, they made global IJOLs about how well they would perform on a future test on the text using a likert-scale ranging from 1 (very well) to 6 (very poorly).

The other half was assigned to the delayed-judgment condition; they read all twelve texts before making global DJOLs for each text. The findings were that IJOLs were more accurate than DJOLs. This finding was contrary to the typical finding in metamemory research showing that DJOLs are more accurate than IJOLs (i.e., Nelson & Dunlosky, 1991). However, early metacomprehension research also reported this pattern of higher accuracy for IJOLs than for DJOLs (Glenberg et al., 1987). One reason for this contrary effect is because the definitions of DJOLs and IJOLs differ somewhat between the metamemory and metacomprehension literature. In metamemory research, IJOLs are predictions made immediately after studying each word pair whereas DJOLs are collected at a separate phase after all word pairs have already been studied. In metacomprehension research, IJOLs are collected immediately after each text is read whereas DJOLs are collected at a separate phase after all texts have been read. In this sense, the IJOL in the metacomprehension literature is more like the DJOL in the metamemory literature. For IJOLs in metacomprehension to be more analogous to IJOLs in metamemory, they would have to be collected after each main idea unit is read in the text. Further reasons for the lack of DJOL superiority over IJOLs in the metacomprehension literature will be discussed in Chapter II under the Metacomprehension Judgment Phase section.

Dunlosky, Rawson, and Middleton (2005) conducted a study comparing global and absolute judgments. Dunlosky et al. (2005) had participants read seven expository texts; each contained four key terms presented in capital letters within the text. After reading each text, participants were prompted to make a global judgment about how well they would be able to complete a test over the text when cued with the text title; the scale used ranged from 0 (definitely would not be able to) to 100 (definitely will be able to). After making the global judgment, participants made absolute judgments. Each term was presented and participants predicted how well they would be able to define each one using the same rating scale as was used for the global judgment. Participants repeated this procedure for all seven texts. They were then given a test on which they had to define each of the four key terms from each of the texts. No differences were found between the mean global judgments (*M* = 51.5), mean absolute judgments (*M* = 52.8), and test performance (*M* = 51.6). Additionally, correlations between mean test performances did not differ for both global judgments (*M* = .52) and absolute judgments (*M* = .57). Therefore, the findings from this experiment showed little difference between global and absolute judgments for texts. However, this experiment does not conclusively show that the two are interchangeable because—in practice—both judgments were the same assessment. Although the global assessment was not term-specific, all four key terms were presented in capital letters within the text. This factor implicitly highlighted the information that would be tested, even though it was not explicitly specified during the global assessment. In addition, testing on definitions may not be the most ideal type of test for text comprehension because definitions can be answered using surface details from the text that does not require consolidation of information from text with long-term memory (Kintsch & van Dijk, 1978).

In addition to assessing comprehension after reading the text, comprehension can be assessed after testing. *Confidence Judgment* (CJ) is a post-diction (as opposed to predictions like JOLs) metacomprehension judgment about confidence in the answer provided on a test question. Mengelkamp & Bannert (2010) investigated participants' ability to make accurate CJs about their comprehension of texts. Participants studied a chapter about principles and applications of operant conditioning via hypermedia with links, diagrams, and texts. After studying for 10 minutes, participants were given a 20-question intermediate test before studying the text for another 20 minutes before taking the final 20-question comprehension test. Participants made CJs about how confident they were about their answer using a scale of 0 (lowest confidence) to 100 (highest confidence). Their comprehension accuracy was relatively high ($M = .79$). In terms of CJ absolute accuracy, participants' CJs were well calibrated indicating that participants were able to, post hoc, consistently predict their overall performance quite accurately; correlation was +.62. This higher metacomprehension accuracy for post-diction CJs than for predictions such as the DJOL is a typical finding in the literature (Glenberg & Epstein, 1985; Maki & Serra, 1992; Pierce & Smith, 2001).

In the experiments conducted in this dissertation, participants made only absolute DJOL metacomprehension assessments. After reading a text, participants completed an interval task to allow for a delay between the study and metacomprehension assessment phase. Each comprehension question that would later be used on the comprehension test was presented and participants made DJOLs to assess how likely it were that they would remember the answer for each question on the future test. After making DJOLs, they provided an answer to each question on the comprehension test. After answering each comprehension question, participants also made a CJ about how confident they were about the answer they just provided. All metacomprehension judgments were made using a scale of 0 (certain not to remember) to 100 (certain to remember). Using the same scale across judgments facilitated comparison between the two judgment types and comprehension accuracy scores.

The Control Process in Text Comprehension

The control process can be measured by asking participants about their learning strategies, such as whether they plan to use rereading or highlighting during comprehension. Because the interplay between monitoring and control is a dynamic process that changes and updates during the comprehension process, the monitoring process continuously monitors the impact of the learning strategy on comprehension (Nelson & Narens, 1990). In this way, the meta level is continuously updated regarding the state of comprehension and continuously adjusts control processes based on that perceived state. Therefore, this dynamic relationship between the meta and object levels—via the monitoring and control processes—contributes toward the final outcome of comprehension. This dynamic relationship also demonstrates that understanding comprehension requires an understanding of the interaction between metacomprehension and comprehension. Similar to the monitoring process, the control process varies with and has been investigated at various stages of the comprehension process.

A first implementation by the control process in metacomprehension occurs prior to any interaction with the text. According to the discrepancy-model of self-regulated learning, students typically set a learning goal for the text material (e.g., Nelson & Narens, 1990; Son & Metcalfe, 2000; Thiede & Dunlosky, 1999; Thiede, Dunlosky, Griffin, & Wiley, 2005). This goal is the desired state of the meta level representation of the object level; this predetermined threshold is set and informs control processes such as study termination (Schunk & Rice, 1989; 1991). During the reading comprehension phase, the student periodically monitors their level of understanding of the text to determine the current state of comprehension, as represented by the meta level. The information garnered from monitoring is then used as a baseline to decide whether the learning goal has been achieved. If the current state of the meta-level representation of the object level meets or exceeds the threshold set by the learning goal, then the control process will be implemented to terminate study. On the contrary, if the current state of meta level has not reached the desired state, the control processes of allocating more study time (Son & Metcalfe, 2000) and/or adjusting learning strategies in order to achieve the desired state of the learning goal more efficiently will be implemented. Monitoring of the object level informs the meta-level representation continuously until there is no more discrepancy between the current state and the learning goal state (Nelson & Narens, 1990). When there is no discrepancy, the control process of terminating study is implemented.

After the initial control process of setting a learning goal state is achieved, the next control process is determining the type of learning strategy to implement while reading the text. Unfortunately, most students have a tendency to not select effective learning strategies that optimize comprehension (Karpicke, Butler, & Roediger, 2009). Karpicke et al. (2009) conducted a survey with 177 students who scored at least an average score on their SAT, thus making their sample a fairly representative group of college undergraduates. Students were asked to list all learning strategies they would use when studying for an exam, and to rank order the learning strategies by their frequency of use. A total of 11 different learning strategies were listed, with the average student listing a mean of 2.9 learning strategies. The most common learning strategy was rereading. A total of 83.6% of the students surveyed listed rereading as one of the learning strategies they normally used, and half of them ranked it as the learning strategy used most often. Only 11% of the students reported using retrieval practice as a learning strategy. The findings are troubling because rereading has been shown to be an ineffective learning strategy for average students and can promote the illusion of competence (Rawson & Dunlosky, 2011; Serra & Dunlosky, 2010; Koriat, 2012). On the contrary, practicing retrieval/self-testing after reading of the text has been shown to be effective (Butler & Roediger, 2007; Karpicke & Roediger, 2007; 2008; Roediger & Karpicke, 2006; Karpicke et al, 2009). In addition to listing learning strategies, Karpicke et al. (2009) gave students a hypothetical scenario in which they had to select what they would do after reading a textbook chapter for an impending exam. Slightly more than half of the students selected the option of rereading, while 20% of the students selected practicing retrieval of the materials of some sort; the remaining students reported using a different learning strategy.

Tan and Eakin (2012) applied the survey findings from Karpicke et al. (2009) in an experimental setting. They gave participants the option to select one of three learning strategies—rereading, highlighting, and retrieval practice—to use to study texts. Consistent with Karpicke et al. (2009), participants selected rereading ($n = 38$) or highlighting ($n = 39$) at a higher rate than they selected the more effective learning strategy of retrieval practice ($n = 24$). Participants then read a text using their self-selected learning strategy. Mean comprehension test scores were significantly higher for participants who selected retrieval practice ($M = .60$, $SE = .04$) than participants who selected rereading ($M = .47$, $SE = .03$) or highlighting; those participants had the lowest comprehension test scores (*M* = .32, *SE* = .02). In a subsequent trial, participants were given another opportunity to select a learning strategy to study a new text. Of the participants who initially selected an ineffective learning strategy (i.e., rereading or highlighting), 80% of the participants reselected ineffective learning strategies whereas the other 20% switched to the effective learning strategy for the new text. These findings showed experimentally that, even when the ineffectiveness of a learning strategy is demonstrated, participants do not always select the effective learning strategy when given a second chance.

Taken together, the findings from Karpicke et al. (2009), and Tan and Eakin (2012) provide evidence that students are more inclined to select the learning strategy that is presumably less effortful (i.e., rereading or highlighting) as opposed to those requiring more effort and time (i.e., retrieval practice). Despite the evidence that students do not typically select them, the literature is clear in demonstrating that learning strategies vary in terms of their effectiveness toward comprehension. Bretzing and Kulhavy (1979) assigned participants one of four learning strategies to use while reading texts. Two learning strategies were presumed to be effective—summarization and paraphrase—and two were presumed to be less effective: verbatim and letter-search. Results showed that participants assigned to summarization or paraphrase obtained significantly higher comprehension test scores than those assigned to verbatim, and letter-search. Similarly, Chi, de Leeuw, Chiu, and LaVancer, (1994) instructed participants to either use self-explain as they read a text or to reread the text twice. When tested on comprehension, participants who engaged in the self-explanation learning strategy showed a significantly higher gain in knowledge than participants who merely reread the text twice. A further review of these studies will be conducted in Chapter II, but the conclusion here is that there are learning strategies that the literature has identified to be effective toward optimal comprehension and learning strategies that have been identified as ineffective.

In the experiments conducted in this dissertation, a learning strategy that has been shown to be effective toward optimal comprehension—delayed explanation—and one that has been shown to be ineffective toward optimal comprehension—keyword-identification—were used (Bretzing & Kulhavy, 1979; Chi et al., 1994). Delayed explanation has the added benefit in that has shown to also improve metacomprehension accuracy (Thiede & Anderson, 2003). Participants were assigned to (Experiment 1) or self-selected (Experiment 2) the learning strategy to use while reading a text. It was anticipated that metacomprehension and comprehension accuracy would be higher when the delayed explanation learning strategy was used than when the keyword-identification learning strategy was used.

The Dynamic Interplay between Monitoring and Control

As theorized by Nelson and Narens (1990), the monitoring process does not have direct access to the object level. This theoretical supposition is supported by findings in the metamemory literature showing that memory and metamemory can be impacted differently by the same factor. For instance, age deficits have been obtained in memory, but metamemory appears to be mostly unaffected by aging (Eakin & Hertzog, 2006; Eakin & Hertzog, 2012; Eakin, Hertzog, & Harris, 2014). In addition, retroactive interference effects have been obtained in memory, but not metamemory; in fact, the two processes are dissociated under retroactive interference (Eakin, 2005). These findings suggest that what is being monitored are cues from object-level processing from which inferences are made about the state of comprehension at the object level. Theories about what these cues are will be discussed in Chapter II. Sometimes monitoring is based on the same factors that will later affect comprehension, but sometimes it is not. The degree to which metacomprehension is accurate is based on the degree to which what is being monitored is also diagnostic of future comprehension (Hertzog, Dunlosky, Sinclair, 2010). If monitoring during comprehension infers future comprehension based on information that is not diagnostic of comprehension, not only will metacomprehension be inaccurate, but control processes, such as selection of appropriate learning strategies, will also be affected. This dynamic interplay between monitoring and control is important when considering why students are so poor at metacomprehension. If students select ineffective learning strategies, the kind of cues that are produced for monitoring are things like surface features, such as keyword, that do not generate connections amongst concepts and do not focus on meaning. Effective learning strategies are also effective toward accurate metacomprehension because the kinds of cues generated while using strategies, such as delayed explanation, include meaning, inferences, connections amongst concepts, and even when retrieval of a concept fails. Monitoring based on this information, which affects future comprehension accuracy, also produces accurate metacomprehension. This potential explanation for metacomprehension accuracy has been explored, resulting in the bulk of the research on metacomprehension focusing on improving monitoring (Maki, 1998b; Thiede & Anderson, 2003; Thiede et al., 2003; Thiede et al., 2005).

Thiede and Anderson (2003) have shown that metacomprehension accuracy can be improved to a level of +.60 (as compared to the typical +.27) when effective learning strategies are used. In their study, participants read six texts using one of three learning strategies: a) writing a summary immediately after reading the text (immediate summarization), b) writing a summary about each text after reading all six texts (delayed-summarization), and c) not writing a summary at all (no summary). After reading and writing summaries for all the texts, participants made global JOLs about how well they understood the texts, prompted with the text title, using a scale of 1 (very poorly) to 7 (very well). Although comprehension test scores did not differ significantly amongst the three conditions ($M = .71$ for no-summary; $M = .78$ for immediate-summary; and $M = .74$ for delayed-summary)—a typical finding that will be discussed later—metacomprehension accuracy was highest for the delayed-summary condition (+.60) as compared to the immediate-summary (+.22) and no-summary (+.24) conditions. Although comprehension was not improved by using the "effective" learning strategy, this learning strategy produced cues that, when monitored, were more diagnostic of further comprehension than those produced by the "ineffective" learning strategies.

The conclusion from Thiede and Anderson (2003), and others showing improved metacomprehension accuracy with effective learning strategies, is that what these strategies are effective as is producing cues that are diagnostic of comprehension, even when comprehension itself did not improve (Thiede et al., 2003; Thiede et al., 2005). As comprehension theories suggest, the comprehension process involves the integration of incoming information with information from long-term memory (e.g., Kintsch & van Dijk, 1978; Gernsbacher, 1997; Zwaan, Langston, & Graesser, 1995), focusing on meaning and making connections amongst concepts in the text (Parr & McNaughton, 2014; Trabasso, Secco, & van den Broek, 1984; van den Broek, White, Kendeou, & Carlson, 2009). Therefore, when a learning strategy allows participants to generate cues at the metacomprehension judgment phase that are diagnostic of comprehension, metacomprehension accuracy improves over using learning strategies that do not produce diagnostic cues, even when comprehension itself does not change. In the case of Thiede and Anderson (2003), participants who only read the text without writing a summary were likely to focus on surface factors, such as how fluent the text was to read. Basing metacomprehension judgments on fluency, for example, produced low metacomprehension accuracy because fluency of the text is not diagnostic of the reader's comprehension of the text. Although summarization focuses more on connecting information to concepts in long-term memory, interconnections amongst text concepts, and meaning, when summaries were written immediately, information retrieved was from immediate memory. Because cues that were generated in immediate memory might not be present at later retrieval, basing metacomprehension judgments on those cues led to inaccurate metacomprehension. The most accurate metacomprehension was observed when summaries were written at a delay because using this strategy, cues that were diagnostic of later comprehension were generated providing diagnostic cues on which to base metacomprehension judgments. The delayed summarization was especially effective because the cues available at a delay were similar to those that would be available on the comprehension test, which was also given after a delay. These findings support the view that metacomprehension accuracy is dependent on using a learning strategy that generates cues on which to base metacomprehension judgments that are diagnostic of future comprehension.

The findings from Thiede and Anderson (2003) and others (Dunlosky et al., 2005; Thiede et al., 2003) also highlight a disconnect in the way the concept of "effectiveness" of learning strategies is discussed in the metacomprehension literature. In the metacomprehension literature, a learning strategy is called "effective" if it improves metacomprehension accuracy via the monitoring process. However, it is frequently found that these "effective" learning strategies do not result in better comprehension accuracy over less effective learning strategies. Therefore, this term "effective" could create confusion when interpreting the term from either a metacomprehension or comprehension accuracy standpoint.

This disconnect is more than a failure of semantics, however. Rather, it highlights a failure of the literature to fully investigate the interplay between monitoring and control processes during comprehension. The goal in improving metacomprehension accuracy is to improve comprehension accuracy. If the monitoring process is accurate, then appropriate control processes can be implemented; the result of those control processes serve to both improve comprehension and generate cues that improve metacomprehension; improving metacomprehension results in implementation of appropriate control processes, and so on until the goal state of comprehension is achieved and the control process to terminate the comprehension process is implemented. However, most of the literature focuses on improving metacomprehension accuracy without examining the resultant impact of that improvement on control processes. Correcting this oversight in the literature is a key aim of this dissertation, as will be discussed. Until that discussion, the term "effective" will refer to learning strategies that have been shown to be effective in improving metacomprehension, but also have the potential (as demonstrated in the comprehension literature) to improve comprehension. Certainly, there are similar underlying aspects to an effective learning strategy that have the potential to benefit both metacomprehension and comprehension, even though it may not always do both.

In the experiments conducted in this dissertation, the effective learning strategy selected was to have participants write an elaborative summary describing the content of the text as if they were explaining it to a friend. This explanation learning strategy was implemented after a delay that was interjected after reading the text. This learning strategy was selected as the effective learning strategy because if fulfilled the criteria to theoretically improve both metacomprehension and comprehension accuracy. The delayed explanation learning strategy can promote metacomprehension accuracy because by writing a summary after a delay, the gist of the information has had the chance to integrate with long-term knowledge and thus the cues available during the metacomprehension judgment phase should be diagnostic of subsequent comprehension. Additionally, the explanation learning strategy is almost identical to the delayed summarization learning strategy that has been shown to improve metacomprehension accuracy in literature (Thiede & Anderson, 2003; Thiede et al., 2003). It is also similar to the retrieval practice learning strategy that was shown to be beneficial for comprehension by Tan and Eakin (2012).

Assessing Metacomprehension

Calibration. Calibration, also known as *absolute accuracy,* measures the degree to which the mean metacomprehension judgments match the level of comprehension test performance. Calibration shows an individual's ability to gauge the actual level of test performance as a whole. For example, a participant would have perfect calibration if their mean judgment was 60% and their score on the comprehension test was also 60%. If not perfectly calibrated, participants can be either over- or under-confident. Calibration correlations are often tested against zero; insignificant difference from zero is interpreted as a correlation no difference from chance, indicating that participants are not able to differentiate between accurate and inaccurate answers on the test. Higher mean JOLs correlated against a lower comprehension test score would yield over-confident metacomprehension calibration, whereas lower mean JOLs correlated against a higher comprehension test score would yield under-confident metacomprehension calibration. To calculate calibration, the metacomprehension scale used must be comparable to the scale used to represent the test performance. For example, using a scale of 0 – 100 for metacomprehension judgments to represent percentage of predicted performance and a proportion representation (0 – 1) for test performance scores would yield appropriate calibration.

Relative accuracy. Relative accuracy, also known as *resolution*, measures the degree to which an individual's metacomprehension judgment about an item is correlated with their performance on that test item. Relative accuracy represents an individual's ability to distinguish between well-learned items from less-learned items by making higher or lower metacomprehension judgments accordingly. To calculate relative accuracy, participants must make metacomprehension judgments for individual test questions, known as absolute judgments; judgments can be made using any increasing scale where a higher number indicates higher confidence. The nonparametric Goodman-Kruskal gamma (Goodman & Kruskal, 1954) is usually used to calculate the relationship between metacomprehension judgments and test performance, and range from -1.0 for a perfect negative correlation to a +1.0 for a perfect positive relationship. A correlation of zero is interpreted as inability to distinguish performance of one item relative to another. In effect, the gamma correlation is a proportion of congruent items minus incongruent items over all items. Therefore, the more congruent a participant's judgments are—a high judgment made for a question answered correctly, and a low judgment made for a question answered incorrectly—the higher and more positive their gamma value will be.

Multi-level modeling analysis. Although the Goodman-Kruskal gamma is the traditional method of calculating relative accuracy, there are problems with calculating gamma when either the metacomprehension judgments or comprehension outcome for a participant are constant. For instance, if a participant gave DJOLs of 50 to every question, or were either accurate or inaccurate for on all of the questions, a gamma could not be calculated for the participant. A mixed-effects method of analysis has been applied to metamemory data to measure accuracy, as well as sensitivity and recall effects by Murayama, Sasaki, Yan, & Smith (2014). This mixed-effects method of multi-level modeling is able to determine accuracy across participants within a condition without dropping any participant due to constant values. In effect, it is calculating accuracy across all of the participants, rather than for each participant directly (Murayama et al., 2014, p. 14). Another benefit of this analysis is that it can test for and model both participant and question level effects, such as varying degrees of test question difficulty. This feature is especially beneficial in studies where tight control over the stimuli is difficult, such as for multiple-choice or free recall questions. Finally, a mixed-effects model can be used even when there is missing data. In the experiments conducted in this dissertation, relative metacomprehension accuracy was calculated using the mixed-effects multi-level modeling approach.

Metacomprehension Paradigm

The first paradigm to investigate metacomprehension was introduced by Glenberg and Epstein (1985). The general procedure had participants read 15 single-paragraph expository texts—130 to 260 words long—from a wide range of topics. After reading, participants made a confidence judgment[1] about their comprehension of the text using a scale of 1 (very low confidence) to 6 (very high confidence). Half of the participants made confidence judgments immediately after reading each text; half of the participants read all 15 texts before making confidence judgments for each text in the same order that they read them. After making their confidence judgments, participants were tested on their comprehension using inference verification statements; statements that were in agreement with the general theme of the text, but were not explicitly stated in the text. Participants had to decide whether the inference statement could be accurately deduced from the text. The test included either an accurate or inaccurate statement for each text.

Glenberg and Epstein (1985) found that the mean proportion correct on the inference verification test for both conditions averaged at .68. However, when they

[1] Although Glenberg and Epstein (1985) called their metacomprehension assessments confidence judgments, they were actually more like immediate and delayed JOLs.

calculated metacomprehension calibration, they found that participant's confidence judgments were not indicative of their actual comprehension level for the texts. Not only did the mean correlation values for both conditions not differ from zero, when tested individually almost half of the participants showed a negative correlation, regardless of whether the confidence judgments were immediate or delayed. The findings from this first experiment, and confirmed by subsequent research (Maki, 1998b; Maki & Serra, 1992; Thiede & Anderson, 2003; Thiede, et al., 2003), indicated that participants were poor at metacomprehension.

Although the proportion correct on the inference verification test was higher than chance, it should be taken into consideration that there was only one true/false inference verification statement per text. Additionally, the inference statements were predominately inferences that could have been directly connected to the explicit idea and topic of the text. Hence, making single inference verification was relatively easy, and therefore may not have been the optimal comprehension measure; especially given that Weaver (1990) demonstrated that multiple test questions improved metacomprehension accuracy, presumably because they are more representative of the text. Additionally, considering the parameters of the inference verification tests, the proportion correct on the inference verification tests should have been closer to perfect (1.0), especially when participants were only required to infer an explicit true/false connection to the core idea of a short expository text.

One of the reasons participants had poor metacomprehension in the first experiment in the Glenberg and Epstein (1985) study could have been that participants did not have experience with the type of inference verification statements used on the comprehension test. Therefore, in the second experiment, half of the participants were provided with three practice trials to familiarize them with the procedure and the types of inference verification statements they would encounter on the comprehension test; half of the participants did not have that opportunity of familiarization. The practice trials were included with the goal of improving calibration by allowing participants to more aptly adjust their confidence judgments to match their comprehension for the texts, now that they knew the types of inference verification statements they would encounter. The rest of the procedure followed the first experiment.

Consistent with the findings of the first experiment, despite familiarization with the statements, participants were still unable to differentiate texts they comprehended from those they did not (Glenberg & Epstein, 1985). The mean proportion correct on the inference verification test averaged at .75, regardless of whether participants received the opportunity of familiarization with the inference verification statements. Similarly, for the measure of metacomprehension calibration, participant's confidence judgments were still not indicative of their comprehension of the text, even with familiarization. Not only did the mean correlation values for both conditions not differ from zero, when tested individually, one third of the participants showed a negative correlation regardless of whether they were familiarized.

One possible reason participants were still not able to calibrate their metacomprehension judgments with their test performance, despite experience with the inference verification statements, is because they only made a single rating for each text. Inherent in the word "calibration" is a sense of ongoing, changing regulation of metacomprehension with changing comprehension so that ratings match one's current comprehension level of the text. In fact, the Nelson & Narens (1990) model speaks to the continuous, dynamic relationship between the meta and object levels of metacomprehension and comprehension, respectively. Collecting only one rating at one point after comprehension may not accurately reflect the outcome of this dynamic process. It is unlikely that participants would be capable of accurate calibration without experiencing this dynamic process. (Although, as will be discussed, they should have experienced this process during reading of the text.) By only giving participants one opportunity to give a metacomprehension judgment, there was little opportunity for them to learn to adjust their ratings and to become more calibrated.

The final experiment conducted by Glenberg and Epstein (1985) addressed the issue of the lack of the opportunity to calibrate their confidence judgments. To allow for more accurate calibration, a slightly different procedure that included two additional confidence judgments and an additional inference verification statement for each text was used. After reading, for each text, participants made an initial confidence judgment, answered an inference verification statement, and made a second confidence judgment of their performance on the initial inference verification test. After that, participant made a third confidence judgment and answered one more inference verification statement.

The main result of interest for this experiment is participant's ability to recalibrate their confidence judgments on the final confidence judgment. The mean proportion correct on the inference verification test remained consistent for both the inference verification tests with an average of .78. On the contrary, the mean calibration for the final confidence judgments was significantly higher than the mean calibration of the initial confidence judgments. Findings showed this procedure resulted in a modest improvement in calibration of the final confidence judgment even when performance remained consistent for both the initial and final inference verification tests (Glenberg & Epstein, 1985). By staggering the confidence judgment and inference verification test phases, participants had the opportunity to more accurately predict and adjust their overall performance on the final inference verification test, by basing their judgments on prior experience.

The seminal study by Glenberg and Epstein (1985) did much to provide both a paradigm for the investigation of metacomprehension and evidence for the importance of experience with the comprehension test toward achieving calibrated metacomprehension. However, a key aim of the experiments conducted in this dissertation was not addressed in Glenberg and Epstein (1985). Although calibrated metacomprehension was achieved, it did not translate into better comprehension performance. Presumably, comprehension performance remained consistent because participants were not given the opportunity to go back and do something about the texts that were not comprehended. Therefore, Glenberg and Epstein (1985) only impacted the monitoring portion of the metacomprehension process. Because participants were not able to use the information from accurate metacomprehension to implement changes to their learning strategies, the control process could not be implemented and measured in their study. Therefore, the benefit of calibrated metacomprehension was confined to the ability to accurately differentiate texts they comprehended from those they did not (Glenberg & Epstein, 1985), rather than also to benefit comprehension due to the implementation of control processes on comprehension.

Overview of Dissertation Experiments

As discussed thus far, a student ideally should be able to accurately monitor their current state of learning so that appropriate and effective control processes are implemented such that comprehension accuracy benefits. However, left to their own devices, students fail to select learning strategies that lead to optimal comprehension (Karpicke et al., 2009; McCabe, 2011; Tan & Eakin, 2012), perhaps due to a failure to base monitoring on factors that are diagnostic of future comprehension (Thiede, 1999; Thiede et al., 2003). If the researcher assigns participants an effective learning strategy, metacomprehension improves, showing the impact of using effective learning strategies on monitoring (Anderson & Thiede, 2008; Thiede & Anderson, 2003; Thiede et al., 2003). Glenberg and Epstein (1985) provided a paradigm for investigating metacomprehension, and demonstrated that although monitoring can be improved across trials, improving monitoring alone does not influence comprehension outcomes.

A likely explanation as to why participants do not appear to be able to improve their comprehension accuracy, despite multiple trials, is related to the dynamic relationship between the monitoring and control processes. A common factor across most metacomprehension experiments is that participants are never allowed to revisit the texts to demonstrate how their updated monitoring influence learning strategy selection via the control process. The process of metacomprehension is dynamic (Nelson & Narens, 1990). Experimentally operating at just the monitoring (Maki, 1998a; Maki & Serra, 1992) or just the control process (Anderson & Thiede, 2008; Thiede & Anderson, 2003; Thiede, et al., 2003) produces a static process, and does not allow for the reflection of monitoring on control and control on monitoring in a truly dynamic way. Thiede and Anderson (2003) showed that effective learning strategies, such as delayed summarization, can improve metacomprehension accuracy over less effective learning strategies; using the effective learning strategy improves the diagnosticity of the cues generated of comprehension, and predictions based on these cues are more accurate than those based on less diagnostic cues.

However, achieving improved metacomprehension is an empty goal if better-calibrated metacomprehension does not also improve comprehension. Indeed, the very motivation for improving metacomprehension should be to achieve accurate monitoring so that the control process can implement appropriate learning strategies to improve comprehension. The main aim of the experiments conducted in this dissertation was to investigate whether improved metacomprehension accuracy via the monitoring process also impacted decisions about learning strategies via the control process, such that comprehension also improved.

The procedure in the dissertation experiments follows that of Glenberg and Epstein (1985), up to a point. The general procedure of the experiments consisted of three trials with of three phases each: a) reading comprehension, b) metacomprehension judgment, and c) comprehension test plus confidence rating. In the first trial, participants read an expository text that was approximately 1200 words on a novel topic. They were assigned to either use an ineffective (i.e., keyword-identification) learning strategy while reading the test, or to use an effective (i.e., delayed explanation) learning strategy after a five-minute delay between reading and doing the summarization. Then, during the metacomprehension judgment phase, participants made absolute DJOLs; each comprehension test question was presented and participants judged the degree to which they would be able to answer the question on the upcoming comprehension test. DJOLs were made using a scale of 0 (certain not to remember) to 100 (certain to remember). Finally, the comprehension test was administered; the test consisted of ten four-alternative forced-choice test questions; participants had to select the best answer to each question. Immediately after providing their answer, participants made a CJ, using the same 0 – 100 scale about their confidence in the answer they selected.

This novel *multi trial metacomprehension paradigm* extends prior research by allowing participants the opportunity to experience the whole sequence of the three phases—reading comprehension, metacomprehension judgment, and comprehension test—three separate times. Allowing multiple trials follows the procedure of Glenberg and Epstein (1985) in giving participants the opportunity to modify their metacomprehension judgments with the goal of improving metacomprehension accuracy; that finding was anticipated. However, the procedure expands on Glenberg and Epstein (1985) by also allowing participants to experience differentially effective learning strategies during the first two trials, and gives the opportunity for control processes to be implemented during the third trial. Allowing multiple trials also follows the procedure of Thiede and Anderson (2003) by allowing participants to show the effect of monitoring accuracy improvement on control processes by having them select which text to restudy, but expands upon that procedure by actually allowing participants to experience the outcome of that control process decision and make metacomprehension predictions about the impact of that learning strategy use on comprehension.

The first aim, explored by Experiment 1, tested whether repeated trials improved metacomprehension accuracy as compared to a control condition, a comparison lacking in prior studies (Thiede & Anderson, 2003; Thiede et al., 2003; Maki, 1998a). Following the procedure above, Experiment 1 consisted of three trials; during the comprehension phase of all three trials, participants were assigned to read a text using either delayed explanation or keyword-identification learning strategy. An addition single trial condition was added during which half of the participants read a text using the delayed explanation and half used the keyword-identification learning strategy. Metacomprehension and comprehension accuracy were predicted to be higher in Trial 3 for multi trial conditions as compared to the single trial condition. Metacomprehension and comprehension accuracy were also predicted to be higher in Trial 3 for the effective learning strategy as compared to the ineffective learning strategy condition.

The second aim, explored by Experiment 2, tested whether the impact of improved metacomprehension accuracy with multiple trials of an effective learning strategy translated into selection of the effective learning strategy so that comprehension was also improved. Experiment 2 tested the impact of improved metacomprehension accuracy on the control process of learning strategy selection. Participants studied the text using either the assigned effective or ineffective learning strategy in Trial 1. In the second trial, participants studied the text using the assigned learning strategy that was opposite to the one they used in Trial 1. For example, if participants were assigned to use delayed explanation in the first trial, they were assigned to use the keyword-identification learning strategy in the second trial. The purpose of this manipulation is to allow participants to experience and contrast the effectiveness of each learning strategy on metacomprehension. The third trial has the critical manipulation to examine how monitoring impacts control processes in metacomprehension. Participants were allowed to select between the two learning strategies to study the text a final time. Based on the findings from Experiment 1, metacomprehension accuracy should have increased across the multiple trials, leading to the potential for the selection of the effective learning strategy by the control process. The likelihood of selecting the effective strategy in these multi mixed trial conditions will be compared to two repeated learning strategy (i.e., effective-effective, ineffective-ineffective) multi same trial conditions, as well as to a control, single trial condition for which participants will select between the two learning strategies without any prior experience with them.

Taken together, these two experiments extended the current literature on metacomprehension, as well as provided a first test of the multi trial metacomprehension paradigm. A complete description of the aims, hypotheses, and experimental methods and procedures will be discussed. First, however, a more detailed review of the metacomprehension literature will be presented in the next chapter.

CHAPTER II

REVIEW OF METACOMPREHENSION LITERATURE

Following the initial paradigm introduced by Glenberg & Epstein (1985), current metacomprehension studies usually include three experimental phases: a) reading comprehension, b) metacomprehension judgment, and c) comprehension test. During reading comprehension, participants read texts as if they were studying for an exam, either using a learning strategy of their choice or one assigned to them by the researchers. The metacomprehension judgment phase follows during which participants make predictions about their future performance on an upcoming comprehension test. Finally, participants take a comprehension test about details from the text they read, sometimes followed by confidence judgments (CJs) about their answers. The goal of metacomprehension research has predominately been to determine how to improve metacomprehension accuracy (Thiede & Anderson, 2003; Thiede, et al. 2003; Maki, 1998b; Maki & Serra, 1992; Weaver, 1990; Weaver & Bryant, 1995). Rather than manipulate factors at all three phases, researchers have typically examined factors at each of the three phases independently to determine ways to improve metacomprehension accuracy by improving comprehension, improving metacomprehension, or by improving test anticipation. The review of prior research that follows is organized by the three phases.

Reading Comprehension Phase

The focus on improving metacomprehension accuracy in the literature has led to research designed to determine the factors improving learning of the text information during the reading comprehension phase. Although the metacomprehension judgment phase, to be discussed second, represents the overt measure of metacomprehension, both the monitoring and control processes of metacomprehension theoretically are at work during the comprehension phase (Nelson & Narens, 1990). Most of the literature has focused on monitoring processes during comprehension (Thiede & Anderson, 2003; Thiede, et al., 2003), while keeping control processes constant by using researcher-assigned learning strategies. However other research has allowed participants to self-select learning strategies, thereby examining the impact of monitoring on control (Tan & Eakin, 2012). The literature examining control processes in comprehension will follow this discussion of monitoring processes.

Monitoring comprehension.

During text comprehension, the monitoring process assesses on-going comprehension, providing continuous updates to the meta level representation of the object level of comprehension. In typical metacomprehension literature, the monitoring of comprehension assessed via metacomprehension judgments are made after reading the text. Although not a direct assessment of the on-going dynamic process discussed earlier during reading of the text, this post-reading comprehension assessment is still a valid assessment of comprehension, especially when absolute metacomprehension judgments are made for each individual test questions and not just a global judgment of the whole text. Certainly, the impact of metacomprehension judgments on control processes could occur at any time during reading comprehension, but judgment are not limited to that phase. As discussed, the most accurate metacomprehension judgments occur when comprehension is assessed after a delay (Maki, 1998a). Therefore, metacomprehension assessments can occur at any stage of the comprehension process.

It has been established that people are generally poor at monitoring the effects of learning strategies on comprehension (Anderson & Thiede, 2008; Linderholm, Wang, Therriault, Zhao, & Jakiel, 2012; Maki, 1998a; McCabe, 2011; Thiede & Anderson, 2003; Thiede et al., 2003). Research has operationalized this monitoring failure as the failure to select learning strategies that lead to optimal comprehension (Hartlep & Forsyth, 2000; McCabe, 2011, Karpicke et al., 2009; Winne & Jamieson-Noel, 2002)

McCabe (2011) conducted a series of experiments to examine whether undergraduates possessed basic knowledge about the benefits of several learning strategies. In the first experiment, participants had to make judgments about the comprehension benefits of learning strategies when presented with scenarios based on findings from past research. Each scenario describes a hypothetical study where Group A was assigned a learning strategy such as practicing retrieval or self-summarization whereas Group B was assigned a learning strategy such as rereading or summarization by others. After reading each scenario, participants made a judgment about which group would perform better using a likert-scale of 1 (Group A would result in higher test scores) to 7 (Group B would result in higher test scores), with a mean rating of 4 (test scores would be similar for both Group A and B). Participants gave significantly higher ratings to rereading than to retrieval practice and self-summarization, indicating that they thought rereading or summarization by others were the better learning strategies. Both retrieval practice (e.g., Rawson & Dunlosky, 2011; Roediger & Karpicke, 2006) and self-summarization (Bloom & Lamkin, 2011; Chi et al., 1994) have previously been shown to be effective for optimal comprehension; however, the participants in this study did not appear to know it. The findings from McCabe (2011) can be concluded that students are not likely to be aware of the benefits of these learning strategies on comprehension. Additionally, these findings suggest the underlying reason as to why studies of metacomprehension accuracy have shown participants to be so poor at assessing comprehension is that students do not know the benefit of these learning strategies to begin with and they are not given the opportunity to experience the benefits by having multiple trials.

In a follow up experiment, McCabe (2011) investigated whether students' awareness of the benefits of some learning strategies could be improved by undergoing training during class lectures throughout the semester. Four groups of students ranging from introductory psychology students to graduate students received different levels of in-depth lectures regarding the learning strategies, including self-summarization, retrieval practice, rereading, and summarization by others. Two groups of participants were enrolled in two sections of Introductory to Psychology (IP) classes; one of these two groups did not receive any lecture about learning strategies (non-lecture IP), the other group (lecture IP) heard a lecture about improving academic performance that included a discussion about the degrees of effectiveness of several learning strategies included in the scenarios, including retrieval practice and self-summarization. A third group of participants consisted of mid-level students enrolled in Human Learning and Memory who had learned about the effect of learning strategies on memory throughout the course.

A fourth group of participants were graduate students enrolled in a special reading seminar on Cognition and Education. Throughout the course, these graduate students read and discussed in detail research articles about the learning strategies represented in each of the scenarios.

Two weeks following the last lecture relevant to the experiment manipulation, all four groups of students served as participants in an experiment using the materials and procedure from the first experiment. Now, only students in the non-lecture IP group gave significantly higher ratings to rereading than to retrieval practice; they gave similar ratings to the participants from the first experiment. All remaining three groups rated retrieval practice significantly higher than rereading. This finding as obtained both when compared to participants from non-lecture IP group and participants in the first experiment. Similar findings were obtained in comparisons between self-summarization and a summary written by others. Participants who received information via lectures regarding the effectiveness—and ineffectiveness—of specific learning strategies on comprehension selected the more advantageous learning strategy out of the two by giving it a higher rating, leading them to endorse retrieval practice over rereading, and self-summarization over a summary written by others.

The findings from the second experiment from McCabe (2011) are encouraging in that they suggest that providing training about what kinds of learning strategies improve comprehension could increase their use, which would also improve their metacomprehension accuracy, which would also lead to more appropriate control processes to be implemented. However, concluding that from McCabe (2011) would be a mistake because, although participants in this study learned how to correctly endorse one learning strategy over another, it does not translate into actual selection of the effective learning strategy during their own study. Karpicke, et al. (2009) surveyed students about their own study habits and reported that students were more inclined to use learning strategies that were less advantageous for comprehension, such as rereading, and highlighting, than they were to use more advantageous learning strategies, such as retrieval practice. Additionally, confirming results by McCabe (Exp. 1; 2011), when asked to choose one of two forced-choice learning strategies, students were more likely to select rereading over practicing retrieval.

The findings from McCabe (Exp. 1; 2011) and Karpicke et al. (2009) suggest that students do not know which learning strategies will benefit comprehension when reading texts. However, it could also be the case that students tend to choose strategies that require less effort to implement. Linderholm et al. (2012) manipulated the amount of cognitive effort participants put forth when studying texts. Some participants were told that they had to meet a high accuracy criterion on the comprehension test, and would have to stay in the experiment until they achieved that criterion. Linderholm et al. (2012) presumed that participants in this high-criterion condition were inclined to put in more cognitive effort when studying the texts than those in the low-criterion condition. Although the learning strategies used were not measured directly, the results for this experiment showed that participants in the high-criterion condition also had higher metacomprehension accuracy (+.20) than those in the low-criterion condition (-.08). Mean comprehension scores did not differ significantly between the high-criterion (M = .53) and low-criterion (M = .55) conditions. Linderholm et al. (2012) stated that the increased cognitive effort translated into using more effective learning strategies during comprehension.

Understanding why students do not choose optimal learning strategies is important. Thiede et al., (2003) found that using effective learning strategies could potentially boost metacomprehension accuracy up to +.70. Participants read six expository texts under three learning strategy conditions. After reading all six texts, one third of the participants generated five keywords that captured the gist of each text (delayed-keyword generation condition). A third of the participants also generated five keywords, but did so immediately after reading each text (immediate-keyword generation condition). The last third of the participants did not generate key words (no-keyword generation condition). All participants then made metacomprehension judgments; participants were cued with the title of each text and made global JOLs about how well they thought they understood that text using a scale of 1 (very poorly) to 7 (very well). Finally, participants were given a comprehension test for each of the texts. Metacomprehension accuracy was calculated using Goodman-Kruskall gamma correlations between metacomprehension ratings and accuracy on the comprehension test. Metacomprehension accuracy was highest in the delayed-keyword generation condition; accuracy did not differ between the immediate-keyword, and no keyword generation condition.

Another learning strategy shown to increase metacomprehension accuracy is summarization (Thiede & Anderson, 2003). Participants read six expository texts under three learning strategy conditions. A third of the participants wrote summaries for each text after reading all six; they were assigned to the delayed-summary condition. A third of the participants also wrote summaries, but did so immediately after reading each text (immediate-summary condition). The last third of the participants were assigned to the no-summary condition. During the metacomprehension judgment phase, participants were cued with the title of each text and made global JOLs about how well they thought they understood that text using a scale of 1 (very poorly) to 7 (very well). After rating their comprehension, participants were given a comprehension test for each of the texts. Comprehension accuracy did not vary with the learning strategy condition ($M = .71$, $SE = .04$ for the no-summary condition; $M = .78$, $SE = .04$ for the immediate-summary condition; $M = .74$, $SE = .04$ for delayed-summary condition). However, gamma correlations between metacomprehension ratings and accuracy on the test showed that metacomprehension accuracy was highest in the delayed-summary condition, and the immediate-summary condition did not differ from the no-summary condition. Writing summaries increased metacomprehension accuracy, but only when written after a delay. Thiede and Anderson (2003) suggested that because the timing of the metacognitive judgments (Maki, 1998a; Nelson and Dunlosky, 1991) and the timing of the test (Maki, 1998a) has been shown to influence the accuracy of the metacomprehension judgments; the timing of summary writing was also crucial to metacomprehension accuracy.

Both the Linderholm et al. (2012) and Thiede and Anderson (2003) studies found that using effective learning strategies improved metacomprehension, but comprehension accuracy remained the same across learning strategies. This finding demonstrates that what is changing when effective learning strategies are implemented is what cues are available to the monitoring process during comprehension. When effective learning strategies are used, the kinds of cues generated are those that are also going to be associated with comprehension. This accessibility viewpoint (Koriat, Lichtenstein, & Fischhoff, 1980; Koriat, 1993) is the predominant theory regarding the basis of metacomprehension judgments, as will be discussed next.

Theories about monitoring.

The finding that effective learning strategies improve metacomprehension accuracy raises the question about what the differences are in terms of what is being monitored when effective versus ineffective learning strategies are used. Current theories propose that metacomprehension judgments are inferential, rather than based on direct access to the cognitive process of comprehension.[2] Sometimes the information used to make inferences is the same information that will subsequently impact comprehension accuracy, but at other times it is not diagnostic of comprehension. Maki (1998b) demonstrated that the kind of information used to make inferences about future comprehension accuracy impacts metacomprehension accuracy. Using a manipulation formerly used by Koriat et al., (1980), Maki (1998b) asked participants to give reasons for why they thought they would get an answer right or wrong on the comprehension test. Koriat et al. (1980) had shown that when participants were instructed to list reasons they would get answers either right or wrong, their ability to accurately assess their future test performance increased. This increased performance was shown especially when they had to list reasons they would get the answers wrong, a finding supported by confirmation bias research (e.g., Nickerson, 1998). Koriat et al. (1980) attributed this increased accuracy to the increased quality of cues that were generated when searching for reasons

[2] In theoretical terms, the meta level does not have direct access to the object level (Nelson & Narens, 1990). Therefore, the meta level representation is based on inferences made based on information from the monitoring process.

why they would—or would not—get answers right. Maki (1998b) suspected that monitoring accuracy improved with more effective learning strategies because the kinds of cues generated while using effective learning strategies were more diagnostic of future comprehension accuracy. Monitoring based on these cues was also, therefore, more accurate (Maki, 1998b).

To further explore the accessibility hypothesis (Koriat et al., 1980; Koriat, 1993), after reading a text, Maki (1998b) instructed participants to list reasons they would either perform well or poorly on a future comprehension test. After producing reasons, participants made metacomprehension judgments prior to taking the comprehension test. Participants who had to list reasons they would get the answers wrong produced more accurate metacomprehension judgments than both participants who had to list reasons they would get the answers right and participants who did not have to list any reasons. Maki (1998b) concluded that the act of listing reasons itself forced participants to more carefully consider the basis of their predictions as compared to when they made predictions without having to produce reasons. This method was particularly effective when listing reasons they would get the answers wrong because not only were cues that were diagnostic of remembering generated, but also cues that were diagnostic of not remembering. Listing reasons generated the kinds of cues that were diagnostic of future memory—or lack thereof—producing more accurate monitoring, leading to better metacomprehension accuracy.

Generation of diagnostic cues could also explain the improved monitoring accuracy in both Thiede et al. (2003; discussed in Chapter I) and Thiede and Anderson (2003). When participants generated five key words either immediately after reading each text or at a delay, they were also generating cues on which they could base their monitoring assessments. Because these cues would also be associated with comprehension accuracy, these cues were diagnostic of later comprehension, producing accurate metacomprehension. The same explanation serves the finding that monitoring accuracy improved with summarization. During summarization, whether immediate or delayed, cues would have been generated on which monitoring judgments could be based and which also impacted comprehension accuracy. By providing these cues on which to base monitoring, using effective learning strategies improves monitoring accuracy over ineffective learning strategies, leading to both better metacomprehension and comprehension.

Using ineffective learning strategies can actually produce cues that are not diagnostic of comprehension, providing an explanation for why ineffective learning strategies, including participant-selected learning strategies, produce poor monitoring accuracy. As Karpicke et al. (2009) noted, students are most likely to use ineffective learning strategies, like rereading. The kinds of cues produced by rereading are things like domain familiarity with the text, which is not associated with metacomprehension accuracy. Glenberg et al. (1987) explored the possibility that judgments are made based on the readers' familiarity with the texts and not the amount of information learned from the text. One group of participants made familiarity ratings for the central theme of the text. The other group participants made judgments of future test performance, similar to a DJOL. The DJOLs were more correlated with the familiarity judgments than they were with actual test performance.

In another experiment, Glenberg and Epstein (1987) showed evidence of an interaction between domain familiarity and expertise when making metacomprehension judgments. They tested groups of music and physics majors; both groups read texts within both the music and physics domains. Each group scored higher on the tests related to their domain. Participants gave higher judgments for questions about texts within their domain of expertise, leading to inaccurate metacomprehension, especially with the participant's own domain. Taken together, these findings demonstrate that domain familiarity is one factor that serves as a basis of metacomprehension judgments (Glenberg et al., 1987). For this reason, most metacomprehension studies test and control for participants' prior domain knowledge (Glenberg & Epstein, 1987; Griffin, Jee, & Wiley, 2009; Jee, Wiley, & Griffin, 2006; Tan & Eakin, 2012).

Apart from domain familiarity, the *ease-of-processing hypothesis* suggests that text fluency also serves as a basis of metacomprehension judgments. Indeed, text fluency does influence both metacomprehension and comprehension accuracy (i.e., Rawson & Dunlosky, 2002; Rawson, Dunlosky, & Thiede, 2000; Dunlosky, Baker, Rawson, & Hertzog, 2006). Rawson and Dunlosky (2002) investigated the influence of text fluency on metacomprehension by manipulating causal coherence, or the degree to which each subsequent sentence is causally related to the initial sentence. In the first experiment, participants read a total of 24 sentence pairs: eight high-coherence, eight moderate-coherence, and eight low-coherence pairs. After reading each pair, participants immediately made an IJOL about the likelihood of remembering the second sentence if presented with the first on a later test using a scale of 0 (definitely will not be able to recall) to 100 (definitely will be able to recall). The next pair was presented and the study-then-IJOL procedure continued for all 24 pairs. Participants then took a cued-recall test on which they had to recall the second sentence when cued with the first sentence. There was a significant difference in comprehension; participants performed better for both the high- (M = .61, SE = .03) and medium-coherence (M = .62, SE = .03) than the low-coherence (M = .47, SE = .03) sentence pairs. Metacomprehension sensitivity showed that participants significantly gave the highest ratings for high-coherence sentence pairs (M = 54, SE = 3), followed by medium- (M = 49, SE = 3), and low-coherence sentences (M = 40, SE = 3). This result shows that the metacomprehension predictions were based on fluency, providing support for the ease-of-processing hypothesis.

In a follow up experiment, Rawson and Dunlosky (2002) investigated whether results from their earlier experiment with paired sentences extended into longer expository texts. A total of eight texts were used in this experiment; each text was modified to create a coherent and an incoherent version. Ambiguous key terms were replaced with more commonly known terms to create coherent versions of the texts. Incoherent texts were created by rearranging clauses in the sentences to increase difficulty of connecting the information in the text. During the reading comprehension phase, each participant first read eight texts, four coherent and four incoherent. Then, participants were prompted to make a global JOL about how they would be able to answer a question about the text on a test using a scale of 0 (definitely won't be able) to 100 (definitely be able). Participants then read the next text and made a judgment after reading it. After reading and making judgments for all eight texts, participants took comprehension tests for each text in the order that they were read. Although not statistically significant, comprehension test performance showed a numerical improvement for coherent texts (*M* = .45, *SE* = .04) over incoherent texts (*M* = .41, *SE* = .03). However, metacomprehension sensitivity was significantly higher for coherent (*M* = 49, *SE* = 3), than for incoherent texts (*M* = 40, *SE* = 3). Although metacomprehension accuracy was not reported, the finding of metacomprehension sensitivity, but not comprehension, that varied with coherence suggests that basing predictions on text fluency would produce inaccurate metacomprehension.

Basing predictions on factors such as fluency (Rawson & Dunlosky, 2002) and domain knowledge (Glenberg et al., 1987) are not diagnostic of comprehension and leads to inaccurate metacomprehension. Accurate metacomprehension relies on basing monitoring on information that is diagnostic of comprehension (Koriat et al., 1980; Maki, 1998b), and whether or not that information available is dependent on the learning strategy, for one (Thiede & Anderson, 2003; Thiede et al., 2003), and opportunity to adjust metacomprehension assessments as more or less diagnostic information becomes available (Glenberg & Epstein, 1985).

Control of comprehension.

Most metacomprehension research discussed thus far demonstrates how people are generally poor at making accurate metacomprehension judgments. However, certain factors have been shown to improve metacomprehension accuracy such as using effective learning strategies. The aim of this section is to explore why these learning strategies that are known to improve monitoring are being classified as effective to begin with. As comprehension theories have suggested, learning from texts involves the integration of incoming information to existing knowledge in long-term memory (i.e., Kintsch & van

Dijk, 1978). Therefore, an effective learning strategy can be operationalized as a learning strategy that promotes this process of comprehension. As is evident from the review of metacomprehension literature thus far, at times learning strategies are only "effective" at improving metacomprehension; their effect is not on comprehension (e.g., Thiede & Anderson, 2003; Thiede et al., 2003; Maki, 1998b). However, there is evidence that some learning strategies are more effective than others at achieving accurate comprehension.

Bretzing and Kulhavy (1979) examined the effectiveness of four different learning strategies—summarization, paraphrase, verbatim, and letter-search—on narrative text comprehension for a group of high school students. Participants were assigned to one of five different learning conditions to use while reading a 2,000-word text about a made up African tribe. Participants assigned to the summarization group were instructed to write a three-line summary of the main theme for each page of the text. Participants assigned to the paraphrase condition were instructed to write up to three lines of notes describing the main theses for each page of the text in their own words. Participants in the verbatim condition were instructed to identify and copy verbatim three of the most important sentences from each page. Participants in the letter-search condition were only instructed to copy all the words that began with a capital letter. A control group of participants did not have to take any notes while reading the text. After the self-paced reading of the text, participants took a 25-question comprehension test. Participants assigned to the summarization and paraphrasing conditions answered the most questions correctly (M = 13.94 and M = 10.55, respectively). Participants assigned to the verbatim condition ranked next (M = 11.11, followed by the control group (M = 11.11); the letter-search condition participants had the lowest score (M = 9.44).

These findings from Bretzing and Kulhavy (1979) provide evidence that learning strategies do vary in their effectiveness toward comprehension accuracy. Allowing students to generate their own information promotes the consolidation of new information from the text with prior knowledge from long-term memory (Chi et al., 1994), which led to a coherent representation of the text, which comprehension theorists say is required for accurate comprehension (e.g., Kintsch & van Dijk, 1978; Gernsbacher, 1990; and Zwaan et al., 1995). Both summarization and paraphrasing met this criterion, and resulted in more accurate comprehension than the learning strategies that did not require this generation and integration process.

Additionally, the self-explanation learning strategy has also shown to lead to good comprehension. Chi, et al., (1994) had eight grade students read about the human circulatory system. Half of the students were assigned the self-explain condition; half was assigned to the rereading condition. Students in the self-explain condition were instructed to explain out loud to themselves what the text meant after reading each line of the text. Students assigned to the rereading condition read the text twice and were not prompted to self-explain. The self-explain group of students showed a significantly greater gain of knowledge (M = 32%) as compared to the rereading group of students (M = 22%).

The benefit from the self-explanation learning strategy is similar to those of summarization and paraphrasing as shown by Bretzing and Kulhavy (1979). Taken together with the findings from Chi et al., (1994), these findings demonstrate that the act of generating and explaining information that has just been obtained from the text to oneself appears to promote the integration process during comprehension (e.g., Kintsch & van Dijk, 1978). The experiments conducted in this dissertation used two learning strategies, one shown to benefit comprehension accuracy more than the other. The effective learning strategy selected was delayed explanation. Explanation is the process of elaborating the text to oneself in one's own words. Explanation meets the criterion set by comprehension theorists of promoting integration and organization of the text information (e.g., Kintsch & van Dijk, 1978; Gernsbacher, 1990; and Zwaan et al., 1995), and also requires self-generation of information. Thiede and Anderson (2003) showed that delayed summarization was better than immediate summarization or no summarization; therefore, a delay was implemented in the dissertation experiments. In addition, the explanation learning strategy used in the dissertation experiments included instructions that were in line with the self-explanation learning strategy (e.g., Chi, et al., 1994; McNamara, 2004). Using a self-explanation style of summarization was hypothesized to lead to generation of cues that will benefit comprehension accuracy, as compared to the ineffective learning strategy. The additional benefit of this cue generation is that these cues will serve as the basis of metacomprehension monitoring, and because these cues will be diagnostic of comprehension, basing monitoring on these cues will also benefit metacomprehension accuracy.

The ineffective learning strategy selected was to identify and write down vocabulary words while reading the text. This keyword-identification learning strategy was ineffective both because the final comprehension test does not focus on definitions of key terms and because the learning strategy was used during study, rather than after reading the texts (see discussion about timing in the Metacomprehension Judgments Phase section, next).

Metacomprehension Judgment Phase

In the studies discussed in the literature review of the reading comprehension phase, several different types of metacomprehension judgments were used. Some studies used global metacomprehension judgments by asking participants to make a single judgment about future comprehension of a complete text (Dunlosky et al., 2005; Glenberg & Epstein, 1985; Glenberg et al., 1987). Other studies provided the questions to be used on the comprehension test and asked participants to make absolute judgments about the degree to which they thought they would remember the answer to the question posed (Maki, 1998a; 1998b; Koriat et al., 1980). Sometimes a likert-type scale was used (Glenberg & Epstein, 1985; Maki, 1998b; Thiede & Anderson, 2003; Thiede et al., 2003) and sometimes a continuous scale of scale of 0 (very unlikely to recall) to 100 (very likely to recall) was used (Dunlosky et al., 2005). Although different scales have been used, a comparison of the effectiveness and accuracy of each type of scale at measuring metacomprehension has not been done. In the experiments conducted in this dissertation, participants made only absolute metacomprehension judgments using a 0 – 100 probability scale. There are two assumed (but not empirically tested) benefits of using this scale over yes/no or likert-type scales. First, participants have a familiarity with probability scales, and there is not a learning curve in terms of understanding how to use this type of scale. Second, participants have shown that they can use this scale to make fine-graded metacomprehension judgments; therefore, the scale is potentially a more sensitive measure of the range of metacomprehension judgments than a dichotomous or categorical scale. For these two reasons, the 0 to 100-probability scale was used in the experiments conducted in this dissertation.

One key factor that has been shown to impact metacomprehension accuracy is the timing of the metacomprehension judgment phase. As discussed, the metacomprehension literature has relied heavily on the metamemory literature to inform its procedures. In the metamemory literature, the timing of the metamemory judgment has been shown to directly impact metamemory accuracy. Nelson and Dunlosky (1991) showed the superior accuracy of judgments made after a delay, DJOLs, as compared to judgments made immediately after study, IJOLs. Nelson and Dunlosky (1991) explained why a delay was crucial toward making more accurate metamemory judgments by proposing the *two-store hypothesis*, which stated that metamemory judgments could be based on information from both short-term and long-term memory. When making an IJOL, participants based their judgments on what was currently available in both short-term and long-term memory, whereas when making a DJOL, participants only had information from long-term memory available on which to base their judgments. Because the cues generated from short-term memory are not diagnostic of retrieval from long-term memory, IJOLs based partly on those cues were less accurate than DJOLs, which were based only on cues generated from long-term memory.

Interestingly, the robust finding for superior monitoring accuracy of DJOLs over IJOLs did not translate into similar benefits on monitoring of text comprehension. Findings from early metacomprehension studies suggested that students are not able to accurately predict their comprehension even after a delay has occurred (Glenberg & Epstein, 1985; Glenberg, et al., 1987). The experiments conducted in this dissertation used DJOLs because doing so provided the best opportunity for accurate metacomprehension, as theoretically, if not empirically, suggested.

Comprehension Test Phase

Metacomprehension researchers have used varying types of tests to measure comprehension, including fill-in-the-blank tests (Maki, Foley, Kajer, Thompson, & Willert, 1990), free recall tests (Maki & Swett, 1987), and multiple-choice questions (Maki & Berry, 1984; Weaver, 1990; Weaver & Bryant, 1995). The importance of appropriate tests to measure comprehension is crucial because participants are making metacomprehension judgments based on their expectation about the comprehension test they will receive. Participants should only be expected to make accurate metacomprehension judgments when their expectation of the future comprehension test matches the actual comprehension test administered to them.

Although there has been no conclusive emerging pattern that identifies which test type effectively improves metacomprehension accuracy, Pressley, Ghatala, Woloshyn, and Pirie (1990) investigated whether participants exposed to different test types after reading the text would impact metacomprehension and comprehension accuracy. After reading each text, participants had to answer either a short-answer essay question or a multiple-choice question. After providing an answer, participants had to self-assess their own answer to the test question. If participants were satisfied with their answer, they were instructed to move on to read the next text. However, if participants were not satisfied with their answer, they reread the text. Overall, participants who were assigned to answer short-answer questions were better at assessing their own answers and therefore were better at making decisions about whether to restudy or move on to the next text. These findings can suggest that short-answer questions produce cues on which to base metacomprehension that are diagnostic of comprehension as compared as multiple-choice questions. Indeed, the benefit of self-generation of cues on memory is well founded in the literature (e.g., Gardiner & Hampton, 1985). Therefore, generating cues while answering essay questions, and then monitoring these cues when making metacomprehension judgments, would produce more accurate judgments. The findings from this study are parallel with the findings on the impact of summarization on metacomprehension accuracy (Thiede & Anderson, 2003; Thiede et al., 2003). In this study, the essay test served as a kind of summary of the text, and the summarization benefited metacomprehension.

However, most current metacomprehension studies use multiple-choice tests as the measure of comprehension accuracy (Maki & Serra, 1992; Weaver, 1990; Weaver & Bryant, 1995; Thiede & Anderson, 2003; Thiede et al., 2003). It could be the case that multiple-choice tests are used because they are the most ecologically valid, given that most general psychology classes also are tested using multiple-choice exams. In addition, using this type of exam allows for tighter experimental control over the information provided by participants. Also, it is important that participants have some knowledge about the type of test they will receive, prior to reading texts. Therefore, participants in the dissertation experiments were informed about the nature of the comprehension test in order to control for text expectancy effects.

Test expectancy has been shown to affect both monitoring and control processes (Glenberg & Epstein, 1985; Glenberg et al., 1987; Maki, Mikkelsen, & Gerlach, 1988; Pressley, et al., 1990; Thiede, Wiley, & Griffin, 2011; Weaver, 1990; Weaver & Bryant, 1995). To explicitly investigate the effect of test expectancy on metacomprehension accuracy, Thiede et al. (2011) manipulated test expectancy by having participants read an initial set of texts and immediately answer either a detail-oriented or inference-oriented test question. These intermediate test questions were meant to manipulate test expectancy for the final test. In the next phase, participants studied six new texts. After reading each text, participants immediately made a global JOL on how well they would perform on a future test, and then took comprehension tests for all six texts. The type of test was manipulated such that half of the time the test was congruent with the intermediate test format (e.g., a detail-oriented intermediate test and a detail-oriented final test) and half of the time the intermediate and final tests were incongruent. Both comprehension and metacomprehension accuracy was significantly higher when the final test matched the test expectancy established by the intermediate test. This finding supports the hypothesis that test expectancy influences metacomprehension judgments. Additionally, because comprehension accuracy also increased when the test type matched the anticipated test type, it can be concluded that test expectancy also guides participants in their selection of learning strategies to use while studying texts.

Knowing not just what to expect on the test, but actually having experienced the test could explain the better metacomprehension accuracy for CJs over JOLs, a finding known as the post-diction superiority effect (e.g., Maki & Serra, 1992; Pierce & Smith, 2001). Accessibility to diagnostic cues during the metacomprehension judgment phase could explain this effect; testing provides cues that occur during comprehension and basing predictions on these cues could produce more accurate CJs than JOLs. Pierce and Smith (2001) compared the accuracy of global JOLs, given prior to a test, to CJs, provided after a test. Participants read a series of four texts. After reading each, participants were given a 16-question test broken into four sets of four questions each for each text. When cued with the text title, participants predicted how many questions out of the first set of four questions they predicted they would answer correctly. Then, after answering the four questions, participants made global CJs about how many questions they thought they had answered correctly. The procedure continued for the four sets of questions and repeated the same procedure for all four texts. The procedure in this experiment is comparable to the delayed condition from Maki (1998a). Pierce and Smith (2001) found a robust post-diction superiority effect; CJ accuracy was significantly higher than DJOL accuracy (see also Maki and Serra, 1992).

Considering the factors that have been shown to impact metacomprehension judgments in the literature reviewed to this point, the experiments conducted in this dissertation used multiple-choice comprehension tests, because it is a more ecologically valid test for undergraduate students. Additionally, because most recent metacomprehension research used multiple-choice comprehension tests (Thiede & Anderson, 2003; Thiede et al., 2003; Maki, 1998a; 1998b; Weaver, 1990; Weaver & Bryant, 1995), the findings from the experiments in this dissertation will be more comparable to the literature. The types of questions that were generated were inferential question that required participants to come to a conclusion based on the information presented in the text as well as the understanding the gist of the text. At each trial, the comprehension test consisted of a total of ten four-alternative forced-choice test questions. In addition, to control for test expectancy amongst participants, they were informed to expect a multiple-choice comprehension test prior to reading the text and when making predictions during the metacomprehension judgment phase.

Conclusion

One major criticism of the metacomprehension studies reviewed in this chapter is that none of these studies used procedures with repeated trials. In order to investigate the dynamic relationship between the monitoring and control process during comprehension; theoretically, after participants experience the whole comprehension process, their monitoring process is updated and will in turn impact the control process for future trials. In the next chapter, a small number of metacomprehension studies using some version of the repeated trials procedure will be reviewed. A new paradigm that should theoretically allow for improved monitoring accuracy as well as allow for changes in learning strategy selection to be implemented by the control process will be introduced in the next chapter.

CHAPTER III

THE MULTI TRIAL METACOMPREHENSION PARADIGM

As discussed thus far, the literature is consistent in demonstrating that metacomprehension monitoring can be less than accurate, but that using effective learning strategies can improve monitoring accuracy (Thiede & Anderson, 2003; Thiede et al., 2003). A key criticism of the metacomprehension literature reviewed thus far is that the procedures used rarely allow participants to demonstrate the impact of improved monitoring on control processes, such as updated learning strategy selection. However, a small number of studies have used repeated trials to allow participants the opportunity to update their control processes on subsequent trials, based on monitoring accuracy on previous trials. Using repeated trials is beneficial toward understanding how the monitoring process impacts the control process based on the dynamic relationship between the two during comprehension. Using repeated trials, changes in control process when informed by the monitoring process can be measured by having participants make decisions about control processes such as allocation of more study time to less learned materials (i.e., Metcalfe, 2009; Son & Metcalfe, 2000) and by selecting effective over ineffective learning strategies on future trials (Tan & Eakin, 2012).

Previous Research Using Repeated Trials

To investigate participants' ability to identify texts to which they would allocate more study time on future trials, Thiede and Anderson (2003; Exp. 2) had participants study longer expository texts than in their first experiment (reported previously in Chapter II). Participants studied the texts using immediate, delayed, or no summarization learning strategies. They then made global JOL predictions for each text, after which they took comprehension tests. After completing the test, participants selected the texts they wanted to restudy, if given the opportunity. Findings in the second experiment paralleled those found in the first experiment; summarization improved metacomprehension accuracy, but only after a delay. Apparently, just as for the first experiment, the cues generated during delayed summary were more diagnostic of comprehension and therefore served as more reliable cues for metacomprehension judgments as compared to cues generated during immediate summary or no summary (Thiede & Anderson, 2003). Additionally, it was found that when metacomprehension accuracy improved, participants were able to identify the less learned from the well-learned text by selecting the less-learned texts for restudy. Although participants did not actually restudy the text, they did show ability to assess texts that would have benefitted from more time, given a restudy trial. This study informed the creation of the multi trial metacomprehension procedure used in the experiments conducted in this dissertation. The benefit of repeated trials in Thiede & Anderson (2003) was that participants had the opportunity to experience the entire comprehension trial: the reading comprehension, metacomprehension judgment, and comprehension test phases. Doing so improved their metacomprehension accuracy. They were then able to use their improved monitoring accuracy to inform their control process to identify texts that were not as well learned as the others.

Ideally, in addition to being able to allocate future study time effectively (i.e., Thiede & Anderson, 2003; Exp. 2), participants should also be able to also select effective learning strategies based on their updated monitoring process to benefit not only metacomprehension accuracy but also comprehension for future study trials. Tan and Eakin (2012) conducted a study that not only allowed participants to experience the entire comprehension trial multiple times but also allowed participants to self-select their own learning strategy to study texts on the initial and final trial. In doing so, participants in Tan and Eakin (2012) were able to use their updated monitoring after experiencing multiple comprehension trials to inform their learning strategy selection on a final trial.

In the initial trial of Tan and Eakin (2012), participants were given an option to select one of three learning strategies—two ineffective (i.e., rereading and highlighting) and one effective (i.e., retrieval practice with restudy opportunity)—to study texts. During the first trial, participants were allowed to select which learning strategy they wanted to use to study the texts. If rereading was selected, participants were allowed to read the text twice. If highlighting was selected, participants highlighted portions of the text while they reading. If retrieval practice with restudy opportunity was selected, participants were presented with the text once and then were given an opportunity to practice retrieval by typing out everything they remembered from the text, a form of free recall of the text. Then they were allowed to restudy the text. Participants then made DJOLs about future performance for each test question using a scale of 0 (not certain to recall) to 100 (certain to recall), and then answered the same questions on a comprehension test. This entire procedure represented one trial.

After the initial trial, each participant completed three training trials during which they received experience using each of the three learning strategies, regardless of which learning strategy they initially selected in Trial 1. This process allowed all participants to experience the effect of all three learning strategies on comprehension. Finally, participants completed a final trial on which they were allowed to select the learning strategy to use while studying new texts.

Participants were categorized into separate groups post hoc based on their learning strategy selection in the initial and final trial. The discussion of results will focus on the group of participants who switched from an ineffective learning strategy (rereading or highlighting) in the initial trial to the effective learning strategy (retrieval practice with restudy opportunity) in the final trial. These participants showed an increase in their comprehension test scores at the final trial as compared to the initial trial. However, contrary to findings from Thiede and Anderson (2003), metacomprehension accuracy dropped significantly ($p < .05$) from the initial ($G = +.82$, $SE = .05$) to the final trial ($G = +.35$, $SE = .11$). This suggests that although participants in Tan and Eakin (2012) selected the effective learning strategy that benefited their final comprehension test scores, they did not accurately monitor that benefit.

Taken together, both Thiede and Anderson (2003) and Tan and Eakin (2012) used multiple trials, but they differed in the impact of repeated trials on metacomprehension accuracy. Metacomprehension accuracy improved in Thiede and Anderson (2003), but not in Tan and Eakin (2012). Another difference is that comprehension accuracy improved in Tan and Eakin (2012), but not in Thiede and Anderson (2003). One key difference between these two studies is that for Tan and Eakin (2012), participants were able to make adjustments to their learning strategy after updating their monitoring. The benefit to comprehension of selecting the effective learning strategy on the final trial was clear. However, it is not clear that this change in learning strategy selection between the initial and final trials was due to increased metacomprehension accuracy. Tan and Eakin (2012) might not have obtained this typical benefit of repeated trials because of the way learning strategy was manipulated in their study. Participants in Thiede and Anderson (2003) used the same learning strategy across trials, whereas participants in Tan and Eakin (2012) actually had more experience with ineffective than effective learning strategies because the three training trials consisted rereading, highlighting and retrieval practice. In particular, those who selected an ineffective learning strategy on the initial trial and switched to select the effective one on the final trial had three trials using ineffective and only one using the effective learning strategy prior to the final trial. In addition to repeated trials, effective learning strategies are also critical toward accurate metacomprehension, and that factor was missing from the Tan and Eakin (2012) study.

An additional factor leading to the disparate findings between Thiede and Anderson (2003) and Tan and Eakin (2012) regards the timing of the metacomprehension judgment phase. Thiede and Anderson (2003) had participants read all texts before making predictions for each of the text, whereas Tan and Eakin (2012) had participants make predictions for each text immediately after reading the text. Maki (1998a) conducted a study to investigate the effects of different timing manipulations on metacomprehension and comprehension accuracy improvement using repeated trials. Maki manipulated the timing between the reading comprehension, metacomprehension judgment, and comprehension test phases to more closely create conditions comparable to the immediate and delayed conditions used in Nelson and Dunlosky (1991). The general procedure for each text required participants to study the text, make a global JOL on how well they would perform in a future test using a likert-scale of 1 (very well) to 6 (very poorly), and finally take the test. All participants completed all read-rate-test task procedure for a total of six texts; however, the timing of rating each text was manipulated creating four different conditions: a) P_iT_i, b) P_iT_d, c) P_dIT_d, and d) P_dNIT_d (P stands for prediction, T for test, i or immediate and d for delay). Table 1 depicts the sequence of the task procedures for all four experimental conditions in Maki (1998a).

Table 1

Sequence of task procedure by experimental conditions in Maki (1998a)

P_iT_i	P_iT_d	P_dIT_d	P_dNIT_d
Read Text 1	Read Text 1	Read Text 1	Read Text 1
Rate Text 1	Rate Text 1	Read Text 2	Read Text 2
Test Text 1	Read Text 2	Rate Text 1	Rate Text 1
Read Text 2	Rate Text 2	Test Text 1	Rate Text 2
Rate Text 2	Test Text 1	Rate Text 2	Test Text 1
Test Text 2	Test Text 2	Test Text 2	Test Text 2
Text 3 – 6: Read-Rate-Test	Text 3 – 6: Read-Rate, Test	Text 3 – 6: Read, Rate-Test	Text 3 – 6: Read, Rate, Test

The first condition was the immediate prediction followed by immediate test condition, P_iT_i. Participants read the text and made a prediction about how well they would perform on a multiple-choice question test using a scale of 1 (very well) to 6 (very poorly). Immediately after making the prediction, participants took the comprehension test. Participants repeated this read-rate-test procedure for all six texts. The second condition was the immediate prediction followed by a delayed test condition, P_iT_d, where participants read a text and immediately made a prediction for the text. Participants repeated this read-rate procedure for all six texts before taking tests for each of the six texts. The third condition was the delayed predictions followed by delayed intervening test condition, P_dIT_d. Participants first read all six texts, and then made predictions about and took a comprehension test for each text. Participants repeated this rate-test procedure for all six texts. The fourth condition was the delayed predictions followed by non-intervening (NI) delayed test condition, P_dNIT_d. Participants did each task in blocks. Participants first read all six texts, then made predictions for all texts, and finally took a comprehension test for all the texts.

Maki (1998a) stated that the comparison between the P_iT_d condition and the P_dNIT_d condition were analogous to the comparison between the IJOL and DJOL conditions, respectively, in Nelson and Dunlosky (1991). However, not only did this comparison not produce better metacomprehension accuracy for the P_dNIT_d condition, the only condition that produced metacomprehension accuracy greater than chance was the P_iT_i condition ($M = .48$). Further, the lowest metacomprehension accuracy was obtained in the P_dNIT_d condition, which was predicted to have the highest metacomprehension accuracy. Maki's (1998a) findings were consistent with those of Glenberg et al. (1987) in finding no evidence for a delayed effect in metacomprehension.

It should be noted that, although the P_dNIT_d condition showed the lowest metacomprehension accuracy, studies that used a similar procedure (i.e., Thiede & Anderson, 2003; Thiede et al., 2003) were able to improve participants' metacomprehension accuracy by assigning participants effective learning strategies, such as delayed-summarization or delayed-keyword generation techniques, suggesting that this procedure can only improve metacomprehension accuracy if effective learning strategies are used. In fact, Maki's (1998a) P_iT_i condition was the only condition during which participants could use their experience from the first text to inform their learning strategies on each subsequent text. Therefore, this was the only condition for which not only monitoring could improve with experience (as occurred in Glenberg and Epstein, 1985) but also this improved monitoring could have impacted learning strategy selection on subsequent trials. If more effective learning strategies were used in subsequent trials, both metacomprehension accuracy and comprehension accuracy would have been higher. Although Maki (1998a) did not report accuracy across trials, on the last trial, metacomprehension and comprehension accuracy were highest for this condition.

Additionally, participants in the P_iT_i condition went through the entire study-predict-test procedure for each text without interruption from other texts. This factor is another reason that metacomprehension and comprehension accuracy were best for this condition. By experiencing the whole process of read-rate-test paradigm, the experience from the monitoring process of the previous trial (text) allows the control process to make any adjustments before beginning a new trial with a new text. For this reason, the proposed new multi trial metacomprehension paradigm will follow this procedure.

The Multi Trial Metacomprehension Paradigm

The multi trial metacomprehension paradigm was proposed to resolve the issues identified in the review of other studies using repeated trials (Glenberg & Epstein, 1985; Thiede & Anderson, 2003; Tan & Eakin, 2012). Using this new paradigm, participants will experience all phases of the comprehension process within a single trial. In the reading comprehension phase, participants will read the text. After reading the text, during metacomprehension judgment phase, participants will make a global JOL rating about the overall information they had learned from the text. Additionally, participants will also make DJOLs for each individual test questions. Finally, in the comprehension test phase, participants will take a comprehension test. This entire process is one trial within the multi trial metacomprehension paradigm (see Figure 3). For this paradigm, these complete trials can be repeated, depending on the research question at hand.

The procedure in the new paradigm provides solutions for the problems highlighted above. First, because participants will experience all phases of the comprehension process in one trial, metacomprehension accuracy can improve across trials (Experiment 1). Second, the timing of the metacomprehension judgment phase within the new paradigm is modeled after the condition for which Maki (1998a) found improved metacomprehension accuracy. Third, learning strategies in the dissertation experiments using this new paradigm will equate the experience gained with effective versus ineffective learning strategies across the trials, rather than load training toward ineffective learning strategies as was done in Tan and Eakin (2012). Finally, this paradigm allows for the potential influence of improved metacomprehension accuracy on the control process by allowing participants to self-select learning strategies in the final trial (Experiment 2).

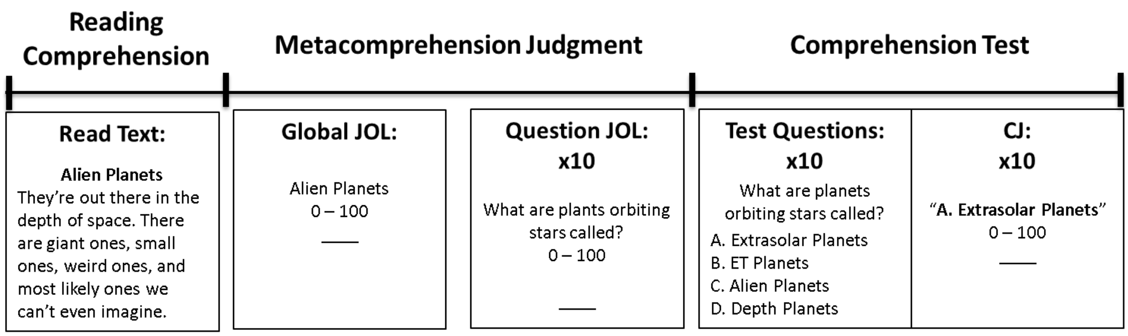

Figure 3. Schematic of the comprehension process within a single trial in the Multi Trial Metacomprehension Paradigm

Schematic for one trial in the multi trial metacomprehension paradigm.

In the multi trial metacomprehension paradigm, this entire trial is repeated multiple times using the same text. Additional experimental manipulations can be implemented at any trial within the paradigm, as was done in the experiments conducted in this experiment.

Pilot Experiment

The aim of this pilot experiment was to test the impact of the multi trial metacomprehension paradigm on metacomprehension accuracy. This pilot experiment investigated whether the multi trial metacomprehension procedure led to both metacomprehension and comprehension accuracy improvement across trials.

Method

Participants and Design

A total of 219 Mississippi State University undergraduate students were recruited via the Psychology Research Program SONA-system website. Participants had to be at least 18 years old and have English as their native language. Participants received research credits for their participation. The design of the experiment was a single factor within-subjects design.

Materials

Text. A single 1,225-word text entitled "Digitized Signals as the Future of the Black Box", taken from ReadWorks.org, an online comprehension texts database, was used in this pilot experiment (see Appendix B). This particular text was selected using several criteria. Due to the fact that this text was to be repeated four times throughout the experiment, the selected text had to be long enough to contain sufficient information to be learned with multiple reading trials, but not too long to reduce interest and effort mid-experiment. The length criterion was at least 1,200 words. The text was also selected based on its topic; the selected text had to be on a topic that was not well known amongst the undergraduate students. Feedback about prior knowledge of the text topic was obtained from undergraduate research assistants, and selected to fulfill the second criterion.

Comprehension test questions. Four separate sets of comprehension test questions were developed. Each set of comprehension test consisted of six open-ended inferential-type questions. An answer rubric was created for each of the comprehension test questions for scoring purposes.

General Procedure

After giving consent, the experiment was presented on a PC computer and programmed using the EPrime 2.0 software (Psychology Software Tools, Pittsburgh, PA). The EPrime 2.0 program presented all the instructions and the experimental tasks on the computer monitor. The experiment began with a practice phase that familiarized participants with the whole procedure of a trial.

Following practice, the experiment began. There were four experimental trials, each consisting of three phases: a) reading comprehension, b) metacomprehension judgment, and c) comprehension test. For the reading comprehension phase, participants were instructed to read the text as if studying for an exam. For the metacomprehension judgment phase, after reading the text, participants first made a global JOL about the overall degree to which they thought they learned the information in the text in order to remember it on a later test. Participants made the prediction using a scale of 0 (will not remember at all) to 100 (will remember completely) after being cued with the title of the text they had read earlier. Then, participants were presented with each comprehension test question and made DJOLs about the likelihood of recalling the answer for each question on a later test. Participants made these individual DJOLs using a scale of 0 (certain not to remember) to 100 (certain to remember) when cued by each of the six questions.

Finally, during the comprehension test phase, participants were given a test that consisted of the same six questions about which they made DJOLs. For each question, participants attempted to recall the answer to the question based on what they had learned from the text. Immediately after answering each test question, participants also made a CJ about the degree to which they felt that the answer they had just provided was correct. Participants made their CJs using a scale of 0 (not confident at all) to 100 (extremely confident). Figure 4 shows the procedure for one trial. Participants completed four trials in total.

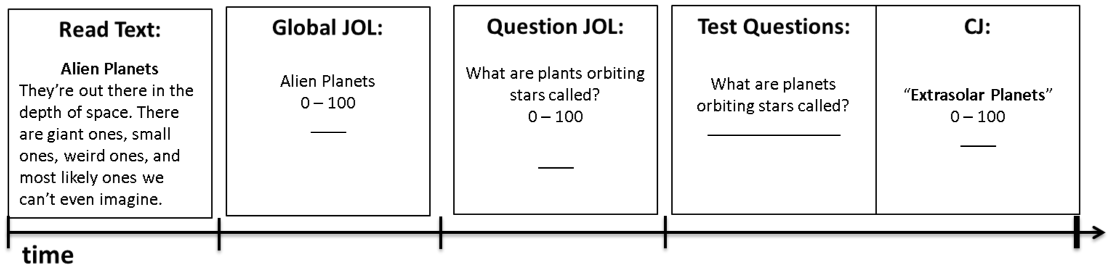

Figure 4. Procedure timeline for Trial 1 of the Pilot Experiment

Timeline schematic of the procedure for Trial 1

Results and Discussion

Comprehension Test Performance

Before reporting results on metacomprehension accuracy, the relationship between comprehension test performance and metacomprehension judgments, descriptive means for each dependent measures across four trials will be reported and discussed. Participants who scored lower than 25% across all four trials were removed from the following analyses. Therefore, the following analyses are conducted based on 88 participants.

A repeated measures Analysis of Variance (ANOVA) showed a significant main effect of Trials on comprehension accuracy, $F\,(3,261) = 10.98$, $p < .01$, $\eta^2 = .11$. Post-hoc tests revealed significant accuracy improvement across all trials except for the difference in Trial 1 ($M = .46$, $SE = .02$) and Trial 2 ($M = .43$, $SE = .02$). All means for comprehension test scores are presented in Table 2 and depicted in Figure 5. Interestingly, this improvement in comprehension test scores across trials generally was not found in metacomprehension literature (i.e., Thiede & Anderson, 2003). However, given that participants read the same text for all four trials, it should be not surprising that overall comprehension test scores improved across trials. It does hint at the possibility of improvement in comprehension test scores if effective learning strategies are selected, or assigned, on future trials. Another possible explanation could be that this pilot experiment employed a different procedural manipulation than a typical metacomprehension paradigm. Contrary to most metacomprehension studies that are set up similar to the P_dNIT_d procedure, this pilot study employed a procedure similar to the P_iT_i explored in Maki (1998a).

Table 2

Mean comprehension accuracy and metacomprehension judgments by trial in Pilot Experiment

Trial #	Comp ACC	Global JOL	DJOL	CJ
1	.46 (.02)	73.96 (1.98)	72.09 (1.77)	73.24 (1.98)
2	.43 (.02)	80.24 (1.96)	72.06 (1.98)	71.84 (2.01)
3	.48 (.02)	82.77 (1.86)	80.95 (1.96)	79.81 (1.73)
4	.55 (.02)	77.80 (1.96)	83.26 (1.64)	81.62 (1.78)

Note. Values in parentheses are standard errors of the means. Descriptive analyses conducted based on n = 88.

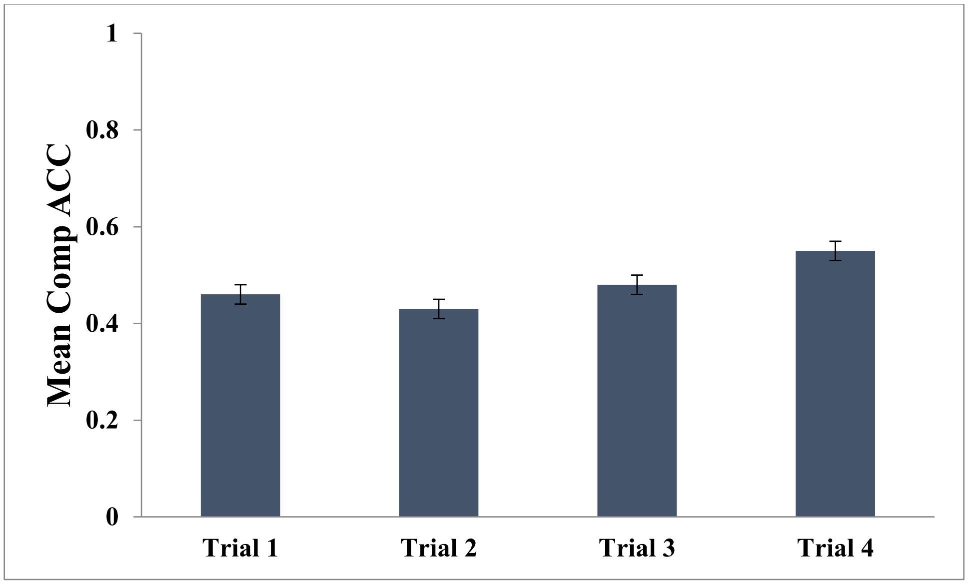

Figure 5. Comprehension accuracy across trials in Pilot Experiment

Mean comprehension accuracy across all trials. Error bars represent standard errors.

Metacomprehension Judgments

Global JOL. A repeated measures ANOVA conducted showed a significant main effect of Trials on global JOL ratings, F (3,261) = 5.76, $p < .01$, $\eta^2 = .06$. Post-hoc tests revealed a significant and steady increase in global JOL ratings from Trial 1 to Trial 2 to Trial 3 before dropping significantly in Trial 4. All means for global JOL ratings are presented in Table 2. Mean values presented show that participants gave the lowest ratings in Trial 1 (M = 73.96, SE = 1.98), followed by Trial 2 (M = 80.24, SE = 1.96), and the highest ratings in Trial 3 (M = 82.77, SE = 1.86). Interestingly, participants significantly lowered their global JOL ratings in Trial 4 (M = 77.80, SE = 2.96). After multiple restudy attempt of the same text, participants got more confident in their ability to remember information from the text.

The significant drop in global JOL ratings for Trial 4 can be explained by the robust findings of underconfidence with-practice (UWP) effect across multiple experimental manipulations conduced with word-pairs in within metamemory literature (i.e., Koriat, Sheffer, & Ma'ayan, 2002). The UWP effect is in accordance with the cue-utilization model (Koriat, 1997), proposing that JOL assessments are made based on a range of cues available to the individual. For example, a mnemonic cue is the sense of familiarity with a text after repeated exposure to it, such as the familiarity from reading the text for the third time. More specifically, participants would have read the same text four times by the time they made their global JOL assessment at Trial 4. The UWP effect suggests that after practice, there was a shift in reliance of intrinsic (i.e., inherent characteristics of the text such as level of text difficulty) and extrinsic (i.e., assessment of the effectiveness of learning strategies) to that of mnemonic (i.e., the subjective experience of whether the text has been well-learned such as familiarity of the materials or retrieval fluency) cues. This could be a possible explanation as to why participants' global JOL ratings dropped significantly in Trial 4. At the time of the metacomprehension assessment at Trial 4, participants would have had exposure to the same text four times.

DJOL. A repeated measures ANOVA conducted showed a significant main effect of Trials for DJOL magnitude (sensitivity), F (3,261) = 30.49, $p < .01$, $\eta^2 = .26$. Post-hoc tests revealed significant differences in DJOL ratings between Trials 1 and 2 and Trials 3 and 4. All mean DJOLs are presented in Table 2. Mean values show a trend of increasing DJOL ratings across trials with the lowest ratings for Trial 1 (M = 72.09, SE = 1.77) and Trial 2 (M = 72.06, SE = 1.98), and significantly higher ratings for Trial 3 (M = 80.95, SE

= 1.69) and Trial 4 (*M* = 83.26, *SE* = 1.64). Similar to the increment in global JOL ratings, participants gave higher DJOLs on later trials; perhaps because they were very familiar with the text at the time of DJOL assessments at later trials.

CJ. A repeated measures ANOVA conducted showed a significant main effect of Trials for CJ rating magnitude, F (3,261) = 19.67, $p < .01$, $\eta^2 = .18$. Post-hoc tests revealed significant differences in CJs between Trials 1 and 2 and Trials 3 and 4. All means for CJ are presented in Table 2. Mean values presented show increasing CJ magnitude with the lowest ratings for Trial 1 (*M* = 72.24, *SE* = 1.98) and Trial 2 (*M* = 71.84, *SE* = 2.02), and significantly higher ratings for Trial 3 (*M* = 79.81, *SE* = 1.73) and Trial 4 (*M* = 81.62, *SE* = 1.78). Similar to the explanation for the increment in DJOLs across trials, it is not surprisingly that participants made higher CJs at later trials; they would have already been very familiar with the text at the time of assessments at later trials.

Summary of descriptive means. For most of the dependent variables tested in this pilot experiment, the descriptive means comparisons generally showed a steady increase in metacomprehension sensitivity across trials. Interestingly, judgments made at Trial 4 occasionally showed a significant drop (i.e., Global JOL) or no significant increase from its previous trial (i.e., DJOL, and CJ). This preliminary comparison of mean judgments across trials hints at the possibility the benefit of multi trial metacomprehension paradigm may peak after three trials.

Metacomprehension Accuracy

Metacomprehension accuracy was operationalized as the correlation between participant's metacomprehension judgments and comprehension test performance across separate trials. The Goodman-Kruskal gamma correlation was calculated between these measures for each individual participant. The means of these intra-individual correlations was then computed across participants within each trial.

Gamma correlation between DJOLs and comprehension accuracy. After calculating Goodman-Kruskal gamma correlations between the DJOLs and comprehension accuracy, a repeated measures ANOVA was computed across trials. No main effect of Trial was found, F (3,156) = 1.26, p > .05, η^2 = .02. However, a comparison of descriptive means across trials showed a steady numerical increase from Trial 1 to Trial 3. Means gamma correlation in Trial 4 showed a large numerical drop, to the level of accuracy at Trial 1. All mean gamma values are presented in Table 3 and depicted in Figure 6.

Although there was not a significant improvement in DJOL accuracy from Trial 1 (*M* = .21, *SE* = .09) to Trial 2 (*M* = .29, *SE* = .09) to Trial 3 (*M* = .41, SE = .09), there was a trend of increased accuracy across trials. This trend suggests that the multi trial metacomprehension paradigm is capable of improving metacomprehension accuracy. There was, however, a drop in accuracy after three trials, with Trial 4 producing the lowest DJOL accuracy (*M* = .20, *SE* = .10). With this drop of accuracy in Trial 4 which is slightly lower than accuracy in Trial 1, it suggests that four trials may be one too many trials for the benefits of the multi trial metacomprehension paradigm to overcome the UPW effect.

Table 3

Mean gamma correlation values by trial in Pilot Experiment

Trial	DJOL x Comp ACC	CJ x Comp ACC
1	.21 (.09)	.40 (.08)
2	.29 (.09)	.47 (.07)
3	.41 (.08)	.52 (.07)
4	.20 (.10)	.29 (.11)

Note. Values in parentheses are standard errors of the means. Gamma correlations comparison conducted based on *n* = 53 for DJOL x Comp ACC, and *n* = 57 for CJ x Comp ACC.

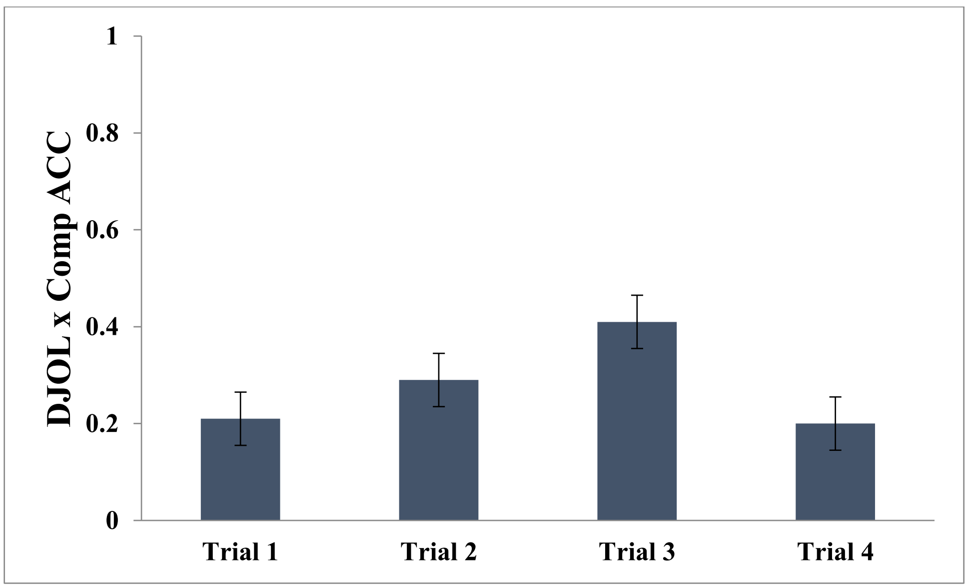

Figure 6. Gamma correlations across trials in Pilot Experiment

Mean gamma correlation across all trials. Error bars represent standard errors.

Gamma correlation between CJs and comprehension accuracy. After calculating Goodman-Kruskal gamma correlations between the CJ ratings and comprehension test scores, a repeated measures ANOVA was computed across trials. No main effect of Trial was found, F (3,168) = 1.39, $p > .05$, $\eta^2 = .02$. However, a comparison of descriptive means across trials showed a numerical increase from Trial 1 to Trial 3. Means gamma correlation in Trial 4 showed a large numerical drop, to a level of accuracy lower than Trial 1. All mean gamma values are presented in Table 3.

A potential reason why no main effect was found with CJ accuracy is because due to the superior CJ accuracy over DJOL accuracy, CJ accuracy for Trial 1 (M = .40, SE = .08), Trial 2 (M = .47, SE = .07), and Trial 3 (M = .52, SE = .07) were already relatively high to begin with. That CJ accuracy for Trial 4 dropped to a (+.29) mimicked the pattern of DJOL accuracy in Trial 4.

Summary of metacomprehension measures. Taking together the results from both DJOL and CJ accuracy and the descriptive means comparison for each of the dependent variables, it can be concluded that there is a numerical benefit for the multi trial metacomprehension paradigm on metacomprehension sensitivity and accuracy across trials. This numerical benefit peaks at Trial 3. Participant's metacomprehension judgments and accuracy were the highest at Trial 3, and either dipped or was maintained at Trial 4.

Conclusions

The main aim for the experiments conducted in this dissertation was to investigate whether improving metacomprehension accuracy via monitoring process impacted decisions on learning strategy via control process such that comprehension was also improved. The main finding from this pilot experiment confirmed that the multi trial metacomprehension paradigm is capable of improving both the sensitivity and accuracy of metacomprehension. This pilot experiment demonstrated that, by allowing participants to experience the comprehension process (via multiple trials), as opposed to experiencing just the comprehension test phase (i.e., Glenberg & Epstein, 1985), participants' metacomprehension accuracy could also be improved across trials.

Additionally, although the findings in this pilot experiment demonstrated the benefits of multiple trials, it also showed that this benefit peaks at the third trial especially for the same text across trials. Furthermore, metacomprehension accuracy peaked at (+.41) for DJOLs and at (+.52) for CJs, which is higher than the typical metacomprehension accuracy found in prior single trial experiments (Pierce & Smith, 2001; Maki & Serra, 1994).

Applying the Multi Trial Metacomprehension Paradigm

The pilot experiment demonstrated the effectiveness of the multi trial metacomprehension paradigm at improving both metacomprehension and comprehension accuracy. Therefore, the experiments conducted in this dissertation extended this novel paradigm to investigate the interaction between the monitoring and control processes during comprehension.

CHAPTER IV

SPECIFIC AIMS

The main goal of this dissertation was to investigate whether improved metacomprehension accuracy via the monitoring process impacted learning strategy selection via the control process such that comprehension was also improved. Two experiments were conducted to achieve this goal. Specific aims and hypotheses for each experiment are described below. A pilot experiment conducted found evidence for the benefit of multi trial metacomprehension procedure in metacomprehension accuracy at the third trial after experiencing the comprehension process twice. The design and results of the pilot experiment are described in Chapter III.

For all of the aims of this study, an *effective learning strategy* is operationalized as a learning strategy that significantly improves metacomprehension across trials and theoretically should improve comprehension, if given a chance to restudy the text and allow control processes to operate on learning strategy selection. *A metacomprehension trial* is operationalized as the entire metacomprehension process from the reading comprehension phase to the comprehension test phase. The multi trial metacomprehension paradigm was used such that the metacomprehension trial was repeated with the goal of improving metacomprehension accuracy to impact control processes such that effective learning strategies are selected, also improving comprehension accuracy.

Aim 1

Aim 1: To determine whether metacomprehension accuracy improved across trials using the multi trial metacomprehension paradigm as compared to a single trial.

The experiment for Aim 1 compared metacomprehension accuracy between single and multiple trials using the multi trial metacomprehension paradigm. Therefore, one factor in the design was Trial Type: single vs. multi. Because the pilot experiment showed that metacomprehension accuracy did not improve until the third trial, the multi trial condition consisted of three trials. Each trial consisted of reading comprehension, metacomprehension judgment, and comprehension test phases.

A second factor, Learning Strategies: ineffective vs. effective, was manipulated in Experiment 1; learning strategy was always researcher-assigned. The ineffective learning strategy was keyword-identification and the effective learning strategy was delayed explanation. Table 4 shows the different conditions for Experiment 1.

Table 4

The experimental conditions for Experiment 1

Condition	Trial 1	Trial 2	Trial 3
Single Ineffective	-	-	Ineffective
Single Effective	-	-	Effective
Multi Trial Ineffective	Ineffective	Ineffective	Ineffective
Multi Trial Effective	Effective	Effective	Effective

Hypothesis 1A directly informed Aim 1: Metacomprehension accuracy was predicted to be higher for the multi trial than the single trial condition. Consistent with Glenberg and Epstein (1985) and the pilot experiment, using repeated trials has shown to provide participants with the opportunity to adjust their metacomprehension judgments, increasing accuracy across trials. Although not directly related to Aim 1, several other hypotheses were also tested:

Hypothesis 1B: Comprehension accuracy was predicted to be higher under the effective than ineffective learning strategy; a main effect of Learning Strategy on comprehension accuracy was predicted. Comprehension accuracy was also predicted to be higher for multi than single trials within each Learning Strategy condition. The two factors were not predicted to interact.

Hypothesis 1C: Metacomprehension accuracy was predicted to be higher for the effective than ineffective learning strategy conditions, overall. Consistent with Thiede and Anderson (2003) and Thiede et al., (2003), effective learning strategies should generate cues that will be diagnostic of future comprehension, producing more accurate metacomprehension for effective than ineffective learning strategies.

Hypothesis 1D: Trial Type and Learning Strategy were predicted to interact; metacomprehension accuracy was predicted to be highest for the Multi Trial Effective condition.

Aim 2

- Aim 2a: To examine the impact of increased metacomprehension accuracy on the control process of learning strategy selection.
- Aim 2b: To determine whether the impact of the control process of learning strategy selection via increased metacomprehension accuracy improved comprehension accuracy across trials.

The experiment for Aim 2 investigated whether improved metacomprehension accuracy by using the multi trial metacomprehension paradigm impacted control-process learning strategy selection on a trial requiring self-selection of the learning strategy.

Metacomprehension accuracy was predicted to improve by using multiple trials, as is predicted to have occurred in Experiment 1. In Experiment 2, however, Trial 3 was modified to allow participants to select which learning strategy they wanted to use to study the text, rather than have it be researcher-assigned. The learning strategy experienced in Trials 1 and 2 was manipulated to examine the impact of experience with effective and ineffective learning strategies on subsequent learning strategy selection. In addition to the Multi Trial Effective and Multi Trial Ineffective conditions from Experiment 1, two mixed-strategy conditions were added in Experiment 2 to investigate whether experience with learning strategies of different levels of effectiveness would contribute to the improvement of metacomprehension accuracy and learning strategy selection on the final trial. A control condition with one self-selected learning strategy trial served as a baseline measure of learning strategy selection. Table 5 shows the different conditions for Experiment 2.

Table 5

Learning strategies assigned at Trials 1 and 2 and to be self-selected at Trial 3 for Experiment 2

Condition	Trial 1	Trial 2	Trial 3
Single Trial	-	-	
Multi Trial Mixed	Ineffective	Effective	Effective
Multi Trial Mixed	Effective	Ineffective	OR
Multi Trial Same	Ineffective	Ineffective	Ineffective
Multi Trial Same	Effective	Effective	

Hypothesis 2A predicted a replication of Experiment 1 in terms of improved metacomprehension accuracy across conditions:

Hypothesis 2A: Metacomprehension accuracy was predicted to be higher for all of the multi trial conditions as compared to the single trial condition, regardless of learning strategy.

Hypothesis 2B directly informed Aim 2a: Hypothesis 2B: The frequency of choosing the effective learning strategy on Trial 3 was predicted to be higher for conditions in which the effective learning strategy was experienced, with the highest frequency being in the Multi Trial Same Effective-Effective condition. Amongst the multi trial conditions, the lowest frequency of selecting the effective learning strategy on Trial 3 was predicted to be found in the Multi Trial Same Ineffective-Ineffective condition.

Hypothesis 2C directly informed Aim 2b: Hypothesis 2C: Comprehension accuracy was predicted to be higher when the effective learning strategy was selected than when the ineffective learning strategy was selected; a main effect of Learning Strategy on comprehension accuracy at Trial 3 was predicted. Comprehension accuracy was also predicted to be higher for the effective than for the ineffective learning strategy in Trial 1 when the learning strategy was assigned by the researcher.

CHAPTER V

EXPERIMENT 1

The aim of this experiment was to determine whether metacomprehension accuracy improved across trials using the multi trial metacomprehension paradigm as compared to a single trial. According to past literature, using repeated trials has been shown to improve metacomprehension accuracy across trials (i.e., Glenberg & Epstein, 1985; Maki, 1998a; Tan & Eakin, 2012). Additionally, effective learning strategies have been shown to improve metacomprehension accuracy (i.e., Thiede & Anderson, 2003; Thiede et al., 2003). Although not directly related to the aim, this experiment also examined the interaction between Trial Type and Learning Strategy toward improving metacomprehension accuracy across trials.

Method

Participants and Design

A total of 140 participants were recruited from Mississippi State University undergraduate students currently enrolled in the General Psychology classes via the Psychology Research Program SONA-system website. Sample size calculation was conducted using the G*Power analysis program (Faul, Erdfelder, Lang, & Buchner, 2007) for a medium effect size of .25, and a power of .80. Participants were at least 18 years old and have English as their native language. Participants received research credit for their participation. Additionally, in order to maintain motivation in a lengthy study, participants were told that they will receive an entry into a raffle drawing for a chance to win a $25 Amazon gift card for each comprehension question they get right.

The design of this experiment was a 2 (Trial Type: single, multi) x 2 (Learning Strategy: ineffective, effective) full factorial between-subjects design. Table 6 shows the four experimental conditions for Experiment 1.

Table 6

The experimental conditions for Experiment 1

Condition	Trial 1	Trial 2	Trial 3
Single Ineffective	-	-	Keyword
Single Effective	-	-	Explanation
Multi Ineffective	Keyword	Keyword	Keyword
Multi Effective	Explanation	Explanation	Explanation

Materials

Text. A single expository text selected from ReadWorks.org, an online comprehension texts database and was used in this experiment (see Appendix B). The text selected was the same text used in the pilot experiment, "Digitized Signals as the Future of the Black Box". The text was selected because of its length and novelty. Because participants read the text multiple times, the selected text had to be long enough to allow for additional information to be learned with each reading. Therefore, the length criterion was set at 1,200 words. The text was also selected based on topic novelty; the selected text was on a topic that is not well known amongst undergraduate students. Undergraduate research assistants in the lab provided feedback indicating that their own prior knowledge of the text topic was minimal to non-existent in fulfillment of the second criterion.

Comprehension test questions. Three separate sets of comprehension test questions for each text were developed, one set per trial. Each set included ten four-alternative multiple-choice questions. In addition to the ten multiple-choice questions per trial, two additional manipulation check multiple-choice questions were also included. These manipulation check questions were included to ensure that participants were paying attention to the task. One question was a simple surface-level question about the text they had just read whereas the other was a simple third grade level general knowledge question. These manipulation check questions served as baseline of attention to the task; participants were required to get at least one of the two manipulation check questions right at each trial in order to be included in the dataset for analysis.

Multiple-choice test questions were selected for this experiment to achieve a more ecologically valid comprehension test measure, as well as to allow for comparison to the bulk of metacomprehension research using multiple-choice tests (Thiede & Anderson, 2003; Thiede et al., 2003; Maki, 1998b; Weaver, 1990; Weaver & Bryant, 1995). The multiple-choice test questions focused on inferential questions rather than definitional questions. The questions tested conclusions drawn based on information presented in the text. Questions were generated on information from the entire text, covering all sections of the text equally. All questions within a trial remained the same for each participant. The order of the test questions was counterbalanced within a trial. Refer to Appendix B for the comprehension test questions at each trial.

Consent form. An IRB-approved and stamped informed consent form informed participants about the general aims and procedures of the experiment. It also explained the minimal risks involved and the benefits for participating in the experiment.

Participants were informed that participation in the experiment is voluntary and that they may choose to cease participation at any time during the experiment and will not be penalized if they choose to do so (see Appendix A).

Post-experiment interview. At the end of the experiment, participants answered some questions about the experiment. Participants were asked to explicitly provide feedback on whether they were aware of the experimental manipulation. For example, participants were asked whether they were aware of their change in ability in answering comprehension test questions and providing metacomprehension judgments across trials.

Debriefing form. A debriefing form was created to give participants more detailed post-experiment information. Participants were informed of the benefit of the effective (delayed explanation) over the ineffective (keyword) learning strategy toward both metacomprehension and comprehension accuracy. Participants were also provided with contact information of the Eakin Memory and Metamemory research lab and the researcher-in-charge in case they have any further questions about the experiment (see Appendix A).

General Procedure

After giving consent, the experiment was presented on a PC computer and programmed using the EPrime 2.0 software (Psychology Software Tools, Pittsburgh, PA). The EPrime 2.0 program presented all the instructions and the experimental tasks on the computer monitor. The experiment began with a practice phase that familiarized participants with the entire procedure of a trial, including how to use the assigned learning strategy properly. Research assistants were trained to monitor participants' progress throughout the experiment to make sure they were using the learning strategy assigned to them correctly.

Following practice, the experiment began. There were three experimental trials, each consisting of three phases: a) reading comprehension, b) metacomprehension judgment, and c) comprehension test.

Reading comprehension phase. For the reading comprehension phase, participants were assigned to use one of two learning strategies, delayed explanation or keyword. Participants were instructed to read the text as if studying for an exam by using the learning strategy assigned to them.

Procedure for the effective learning strategy. The effective learning strategy to be used was the delayed explanation learning strategy. Participants first read the text. The text was presented on timed slides with a total read time of eight minutes, allowing for five minutes of summary writing later. After reading the text, participants proceeded to solve the Tower of Hanoi for five minutes to serve as an interval task. This interval task was meant to induce a delay before participants wrote an elaborative-summary about the text. Participants were instructed to write a summary explaining what the text was about as if telling a story to an audience in their own words. Participants typed their summary into the computer using the computer keyboard. This strategy has previously been shown to promote metacomprehension accuracy (e.g., Thiede & Anderson, 2003[3]) as well as comprehension (e.g., Bretzing & Kulhavy, 1979).

---

[3] Thiede and Anderson (2003) demonstrated the effectiveness of the summarization learning strategy after a delay. This strategy is being used here as the effective learning strategy.

Procedure for the ineffective learning strategy. The ineffective learning strategy to be used was the keyword learning strategy. Participants assigned to the keyword learning strategy were provided with four notecards to write down key terms while they read the text. The text was presented on timed slides with a total read time of 13 minutes, allowing for writing of keywords on the notecards. This learning strategy has previously been shown to be ineffective toward promoting comprehension (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Bretzing & Kulhavy, 1979) and theoretically ineffective toward metacomprehension.

To make sure participants were using the assigned learning strategy appropriately, research assistants checked their summary and notecards before they were allowed to proceed to the next phase of the trial. This manipulation check was put in place to make sure participants were properly using the learning strategy to which they were assigned.

Metacomprehension judgment phase. The second phase was the metacomprehension judgment phase. After reading the text and prior to the start of the metacomprehension judgment phase, participants completed an interval task for 5 minutes (solving Tower of Hanoi puzzles). After completing the interval task, participants were presented with each comprehension test question and made DJOLs about the likelihood of recalling the answer for each question on a later test. Participants made these individual DJOL predictions using a scale of 0 (certain not to remember) to 100 (certain to remember) when cued by each of the twelve questions, including the two manipulation check questions. Participants were informed that the test was a multiple-choice test, but the answer alternatives will not be presented during the metacomprehension prediction.

Comprehension test phase. Finally, during the comprehension test phase, participants took a four-alternative multiple-choice test that consisted of the same twelve questions (including the two manipulation check questions) to which they made DJOLs. For each question, participants selected the answer to the question based on what they learned from the text. Participants were required to make a selection—no blanks were allowed—and guessing was allowed. Immediately after answering each test question, participants also made a CJ about the degree to which they felt that the answer they just provided was correct. Participants made their CJs using a scale of 0 (not confident at all) to 100 (extremely confident). Depending on which condition participants were assigned to, they either completed one or three trials in total. The phases for each trial were identical.

At the end of the experiment, participants answered the questions in the post-experiment questionnaire. Finally, participants were debriefed and assigned research credit. After the comprehension accuracy was scored for the participant, they had the number of tickets equal to their accuracy placed in the drawing.

Figures 7, 8, and 9 show the procedure timeline for the first, second, and third trials, respectively, for the Multi Trial Ineffective condition. Figure 10, 11, and 12 show the procedure timeline for the first, second, and third trial, respectively, for the Multi Trial Effective condition. For the Single Trial condition, participants only completed Trial 3 using either the Effective or Ineffective learning strategy.

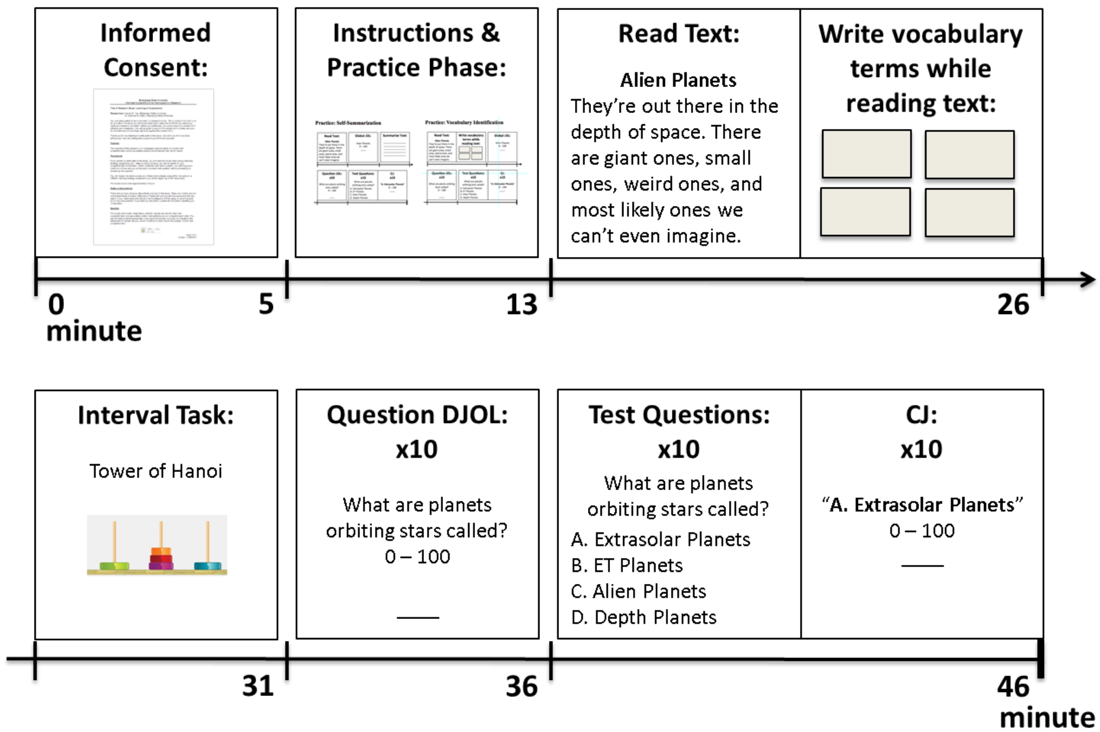

Figure 7. Procedure timeline for Trial 1 in the multi trial ineffective learning strategy condition in Experiment 1

Timeline schematic for Trial 1 of the multi trial keyword condition in Experiment 1.

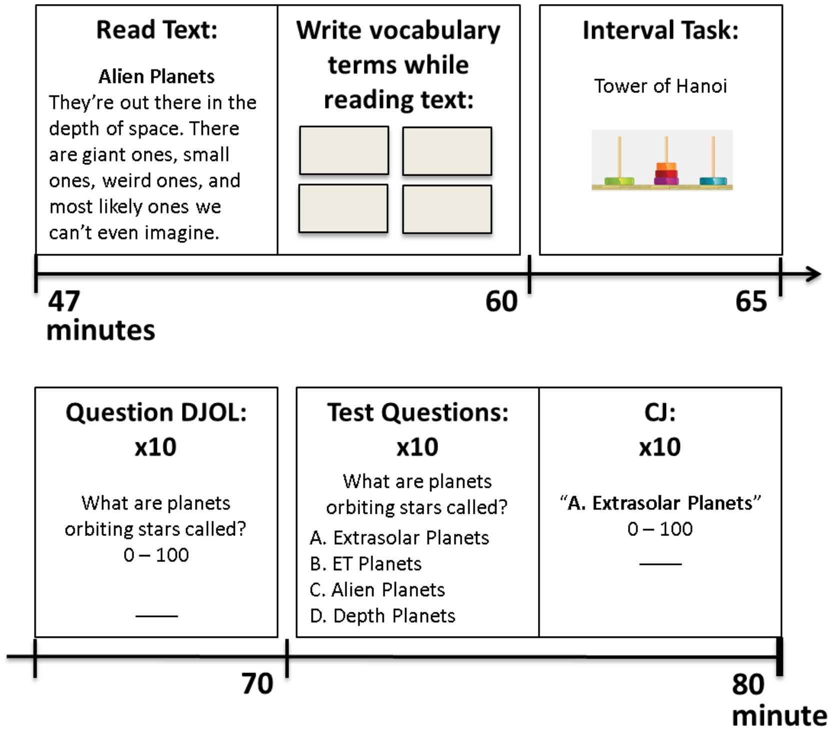

Figure 8. Procedure timeline for Trial 2 in the multi trial ineffective learning strategy condition in Experiment 1

Timeline schematic for Trial 2 of the multi trial keyword condition in Experiment 1.

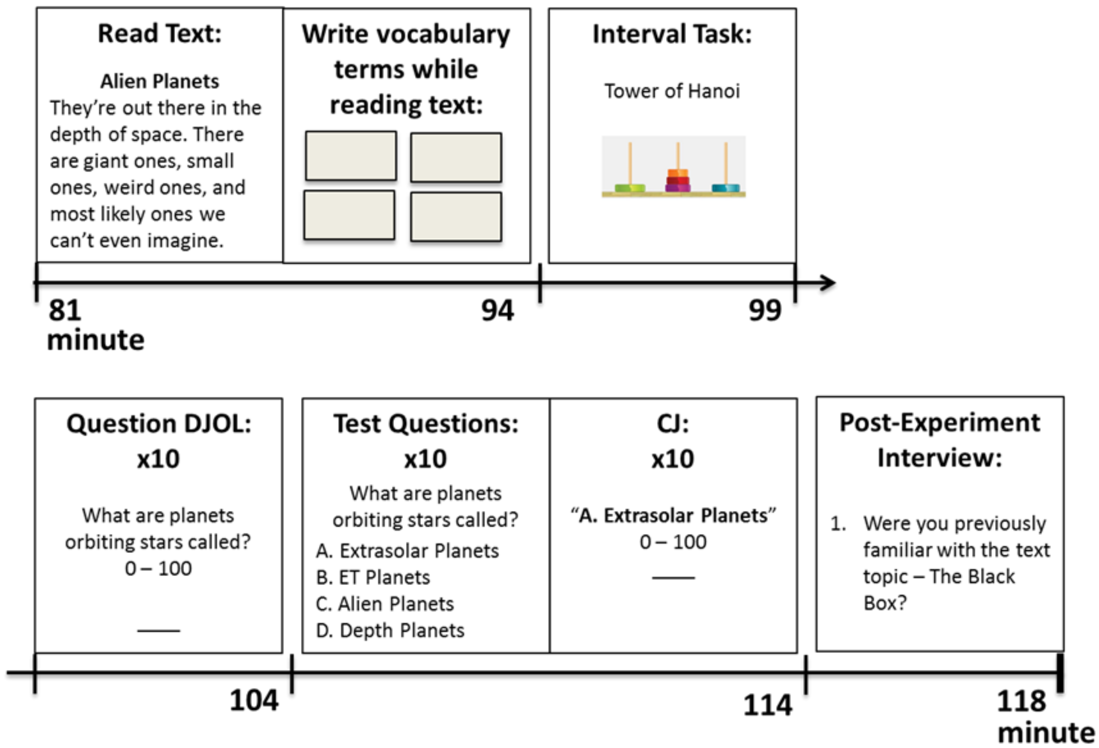

Figure 9. Procedure timeline for Trial 3 in the multi trial ineffective learning strategy condition in Experiment 1

Timeline schematic for Trial 3 of the multi trial keyword condition in Experiment 1.

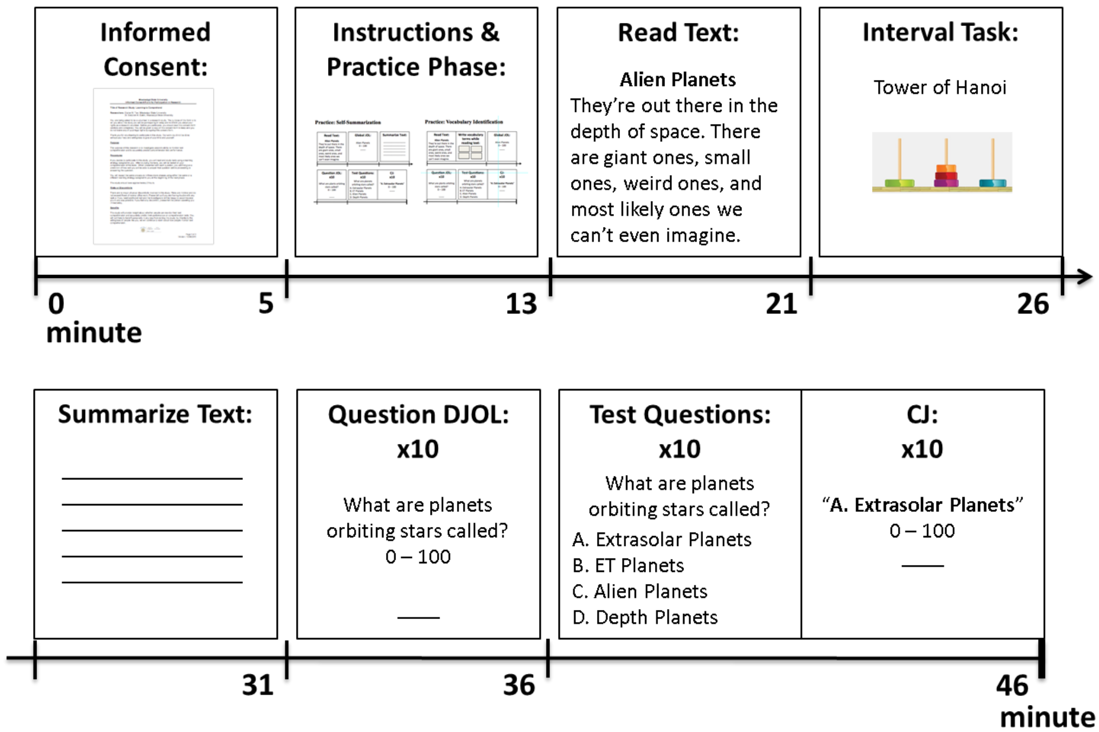

Figure 10. Procedure timeline for Trial 1 in the multi trial effective learning strategy condition in Experiment 1

Timeline schematic for Trial 1 of the multi trial explanation condition in Experiment 1.

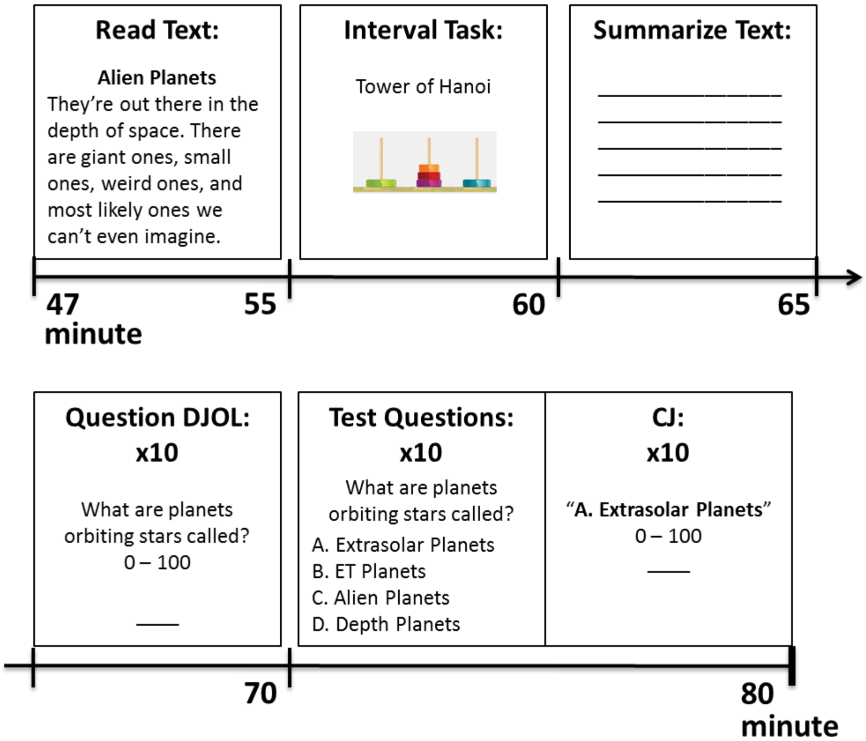

Figure 11. Procedure timeline for Trial 2 in the multi trial effective learning strategy condition in Experiment 1

Timeline schematic for Trial 2 of the multi trial explanation condition in Experiment 1.

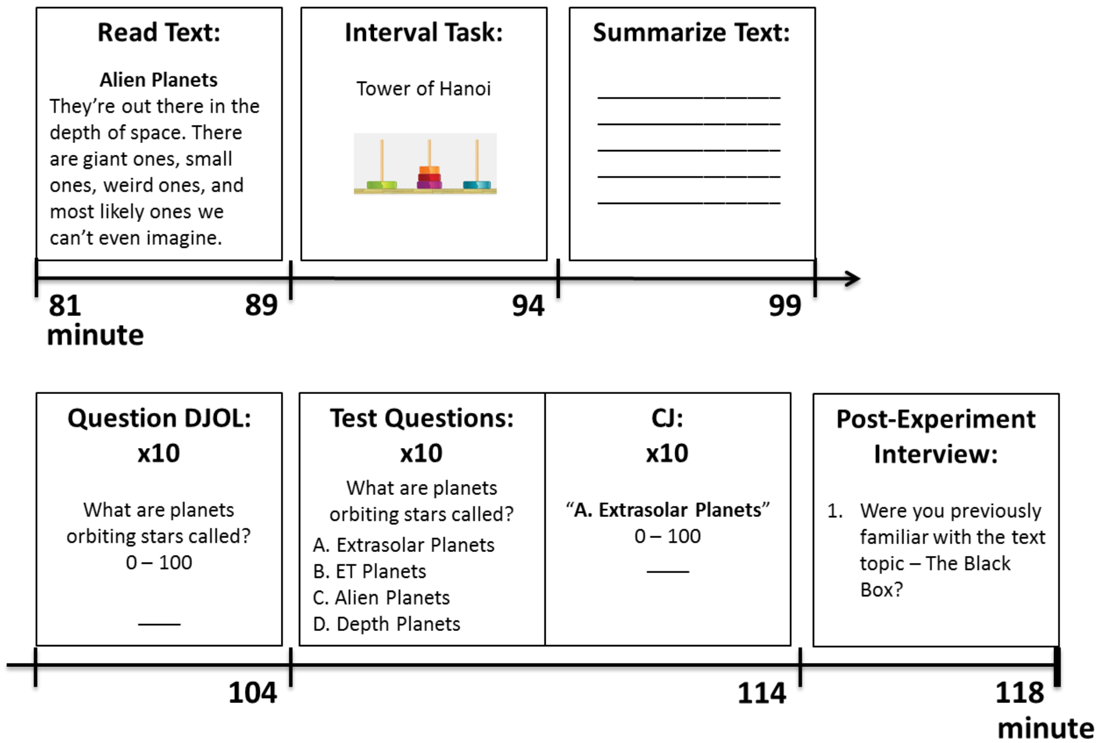

Figure 12. Procedure timeline for Trial 3 in the multi trial effective learning strategy condition in Experiment 1

Timeline schematic for Trial 3 of the multi trial explanation condition in Experiment 1.

Results

The manipulation check questions were first scored in order to determine whether there were participants who should be excluded from analysis for failing to meet this inclusion criterion; all participants answered at least one of the questions correctly. The analysis included all participants who completed the experiment. In preparation of data analyses, all metacomprehension prediction variables (e.g., DJOLs, and CJs) were centralized and converted to z-scores within each participant and each trial. Due to 24 incidences—12 DJOLs and 12 CJs—when participants did not provide any variability when making metacomprehension predictions for questions within a single trial (e.g., when a participant predicted DJOLs of 80 for all questions in Trial 3), z-scores were not tabulated for those participants for that particular trial. In order to not lose those participants, missing z-scores were replaced with a "0" to represent the participant's mean centralized prediction for that trial.

The data were analyzed using generalized linear mixed effects modeling (GLMM) approach when calculating metacomprehension accuracy using the lme4 package (Bates, Maechler, Bolker, & Walker, 2015) in an R statistical computing environment. The GLMM approach used comprehension accuracy, a dichotomous variable, as the dependent variable and metacomprehension predictions (i.e., DJOLs and CJs) as the main predictors. This method of analysis has the advantage of not eliminating participants due to the lack of variability in measures, such as constant metacomprehension predictions, and retaining participants who have missing data (see Murayama et al., 2015). Additionally, this GLMM method of analysis models random effects of participant and items (i.e., comprehension test question). For all GLMM analyses discussed in this section, a maximal random effect structure was used, as suggested by Barr, Levy, Scheepers, and Tily (2013). Participants and questions within trials were modeled as random effects (i.e., both random intercepts and random slopes for each random factor was included in all the comparison model analysis). All GLMM analyses reported in these results successfully converged using the maximal random effect structure with the help of "bobyqa" optimizer. The full syntax and model output for each hypothesis testing, including the random effects model can be found in Appendix C, Tables C1 – C24.

For analyses comparing the magnitude of metacomprehension predictions—DJOLs and CJs are continuous variables—linear mixed effects modeling (LMEM) was used. The same benefits of the GLMM are also true of the LMEM approach. Similarly, all LMEM comparisons discussed in this section included the maximal random effect structure. Both participant and questions were modeled as random effects. All LMEM analyses reported successfully converged with the maximal random effect structure.

This Results section will discuss the findings in terms of the hypotheses delineated in the Aims section, starting with the findings regarding comprehension accuracy. However, before introducing the hypotheses regarding metacomprehension accuracy, the findings regarding metacomprehension sensitivity will be presented because accuracy is the correspondence of sensitivity with comprehension accuracy. All hypotheses were based on comparing participants' comprehension and metacomprehension performances in Trial 3. After all hypotheses have been addressed, additional analyses suggested by the findings, will be discussed that further explore Trial Type and Learning Strategy on metacomprehension accuracy.

Hypotheses Testing for Comprehension Accuracy

Comprehension accuracy. To test for Hypothesis 1B, both Trial Type and Learning Strategy were the main predictor for comprehension accuracy in the GLMM analyses.

Hypothesis 1B. Comprehension accuracy was predicted to be higher under the effective than ineffective learning strategy; a main effect of Learning Strategy on comprehension accuracy was predicted. Comprehension accuracy was also predicted to be higher for multi than single trials within each Learning Strategy condition. The two factors were not predicted to interact.

GLMM logistic regression results for effects of Learning Strategy and Trial Type on comprehension accuracy are presented in Table 7. Comprehension accuracy was measured as the proportion of test questions that were scored as correct on the comprehension test. Hypothesis 1B predicted a main effect of both Learning Strategy and Trial Type on comprehension accuracy. The main effect of Learning Strategy was not significant. Participants in the explanation condition ($M = .68$, $SE = .02$) and keyword condition ($M = .66$, $SE = .02$) had similar comprehension accuracy.

The main effect of Trial Type was significant. Test performance for participants in the multi condition ($M = .73$, $SE = .02$) was significantly different than for those participants in the single condition ($M = .61$, $SE = .02$). Comprehension mean scores for all conditions are presented in Figure 11 below.

Hypothesis 1B was partially supported.

Table 7

GLMM fixed effects results for planned comparisons for Hypothesis 1B in Experiment 1 (DV: comprehension accuracy)

Hypothesis 1B	Predictors	Estimate	*SE*	z-Wald	*p*
Main Effect of Trial Type					
Multi > Single	Intercept	1.41	.39	3.66	< .05
	TT2	- .85	.24	-3.57	< .01**
Main Effect of Learning Strategy Assigned					
E = KW	Intercept	1.02	.38	2.66	< .05
	LS3	- .19	.23	- .83	.41

[2] Trial Type; [3] Learning Strategy Assigned
** $p < .01$ for fixed effects of conditions
Note. Full GLMM output for these analyses can be located in Appendix C.

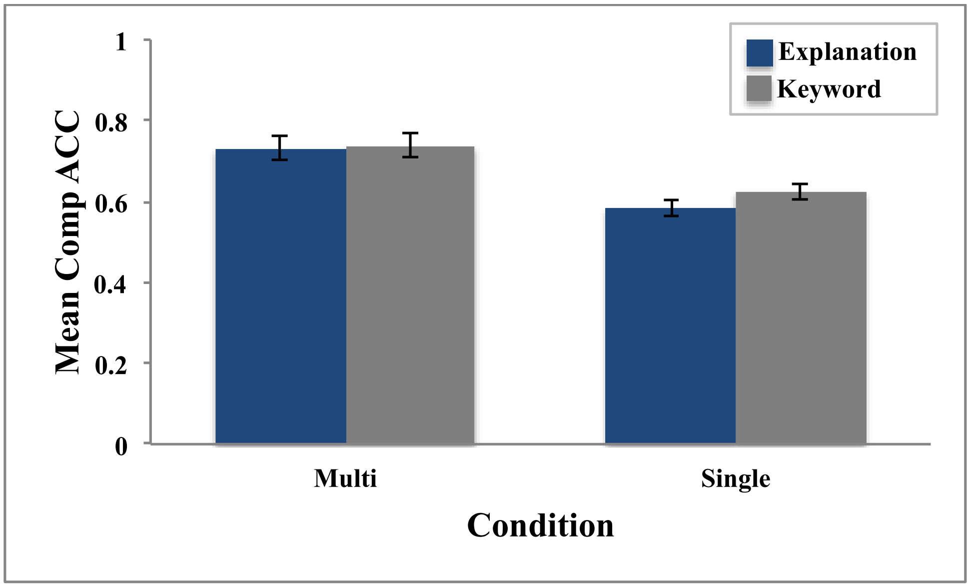

Figure 13. Comprehension accuracy for all conditions at Trial 3 in Experiment 1

Mean comprehension accuracy results for Hypothesis 1B. Error bars represent standard errors.

Delayed judgments of learning (DJOL) sensitivity. The magnitude of DJOL predictions was analyzed using the LMEM approach. The effect of Trial Type and Learning Strategy on the magnitude of DJOLs was computed. Results for the LMEM regression are presented in Table 8. There was a significant main effect of Trial Type; participants who had multiple trials (*M* = 80.31, *SE* = 1.12) gave significantly higher DJOLs than participants who had a single trial (*M* = 58.77, *SE* = 1.06). The main effect of Learning Strategy was not significant. Participants assigned to use the explanation learning strategy (*M*= 69.18, *SE* = 1.11) made similar DJOLs to participants assigned to use the keyword learning strategy (*M* = 69.90, *SE* = 1.13). The interaction between Trial Type and Learning Strategy was not significant. DJOLs did not vary with Learning Strategy within each Trial Type conditions. Mean DJOL sensitivity across condition is depicted in the graph below (see Figure 14).

Table 8

LMEM fixed effects results for planned comparisons for DJOL sensitivity for Trial 3 in Experiment 1 (DV: DJOL)

	Predictors	Estimate	*SE*	*df*	*t*	*p*
Main Effect of Trial Type and Learning Strategy						
Multi > Single	Intercept	80.26	2.96	32	27.14	< .01**
	TT[2]	-21.49	3.47	55	-6.19	< .01**
E = KW	Intercept	59.43	3.92	31	15.16	< .01
	LS[3]	- 1.32	4.49	69	- .30	.80
Interaction Effect: Multi-E as comparison group						
Multi-E = Multi-KW	Intercept	78.92	4.05	35	19.48	< .01
	TT[2]xLS[3]	2.78	4.97	71	.56	.60
Multi-E > Single-E	TT[2]xLS[3]	-19.49	4.45	95	-4.37	< .01**
Multi-E > Single-KW	TT[2]xLS[3]	-20.81	4.81	65	-4.33	< .01**
Interaction Effect: Multi-KW as comparison group						
Multi-KW > Single-E	Intercept	81.69	3.37	61	22.05	< .01
	TT[2]xLS[3]	-22.26	4.98	61	-4.47	< .01**
Multi-KW > Single-KW	TT[2]xLS[3]	-23.58	5.07	51	-4.65	< .01**
Interaction Effect: Single-E as comparison group						
Single-E = Single-KW	Intercept	58.11	3.34	55	17.41	< .01
	TT[2]xLS[3]	1.32	4.41	85	.30	.76

[2] Trial Type; [3] Learning Strategy
** $p < .01$ for fixed effects of conditions
Note. Full LMEM outputs for these analyses can be located in Appendix C.

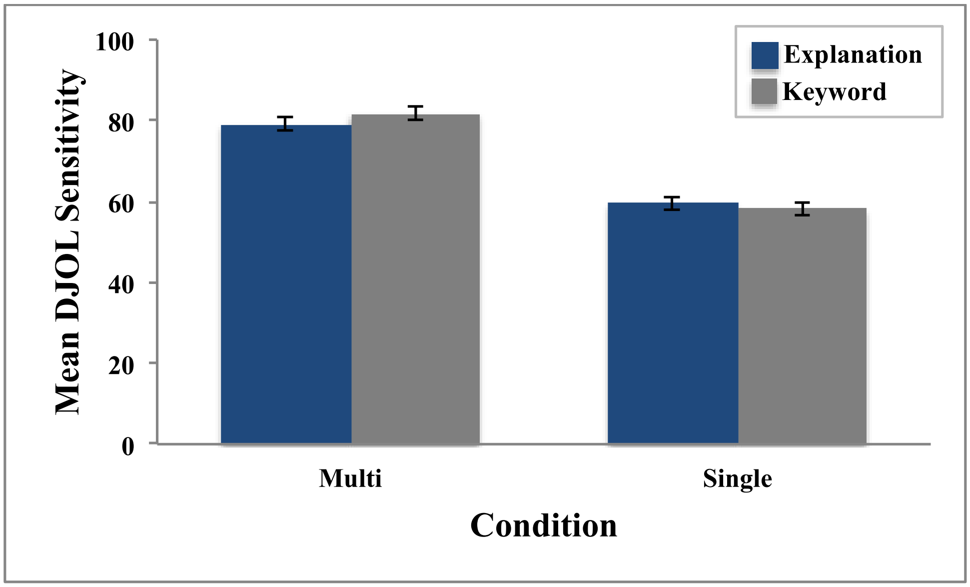

Figure 14. Delayed judgments of learning for all conditions at Trial 3 in Experiment 1

Mean DJOL sensitivity for all conditions at Trial 3. Error bars represents standard errors.

Hypotheses Testing of Metacomprehension Accuracy

The following section will address the findings for each of the hypotheses regarding metacomprehension accuracy.

Metacomprehension accuracy. To measure metacomprehension accuracy, comprehension accuracy was entered as the dependent variable in the GLMM analyses. Centralized DJOLs were the predictor variable. Therefore, metacomprehension accuracy was the degree to which DJOLs predicted comprehension. Additionally, depending on the hypothesis being tested, the effects of Trial Type and Learning Strategy were also included as predictor variables with metacomprehension accuracy as the dependent variable.

The GLMM produces the results based on logistic regression. Therefore, metacomprehension accuracy for each comparison condition is represented as the amount of change in the slope relative to the dependent variable. When metacomprehension accuracy is the dependent variable in the following results, change in metacomprehension accuracy is explained by the degree to which one condition varies as compared to the comparison condition. Therefore, when reporting the mean and standard error values for metacomprehension accuracy for these conditions, the estimate value for the comparison group will also be reported to provide relative comparison. Using the estimates generated from the logistic regression output, line graphs were created to represent the differences in slope change relative to the dependent variable amongst the comparison conditions.

Hypothesis 1A. Metacomprehension accuracy was predicted to be higher for the multi trial than the single trial condition.

To test for Hypothesis 1A, metacomprehension accuracy was compared between multiple and single trials. The multi trial condition served as the comparison group. Results for the GLMM logistic regression is presented in Table 9 and depicted in Figure 15. There was no significant difference in metacomprehension accuracy between multi and single Trial Types. Participants in the multi (M = .32, SE = .15) and single conditions (M_{change} = -.08, SE = .16) did not significantly differ in their metacomprehension accuracy slopes. Therefore, Hypothesis 1A was not supported.

Table 9 *GLMM fixed effects results for planned comparisons for Hypotheses 1A, 1C, and 1D in Experiment 1 (DV: comprehension accuracy)*

Hypothesis 1	Predictors	Estimate	*SE*	*z*-Wald	*p*
H1A: Multi = Single	$DJOL_{Est.}$	.32	.15	2.17	< .05
	DJOL[1] xTT[2]	- .08	.16	- .48	.63
H1C: E > KW	$DJOL_{Est.}$	.43	.14	2.99	< .01
	DJOL[1]xLS[3]	- .33	.16	-2.10	.04[*]
H1D: Multi-E as comparison group					
Multi-E = Multi-KW	$DJOL_{Est}$	.52	.25	2.14	< .05
	DJOL[1]xTT[2]xLS[3]	- .38	.29	-1.29	.20
Multi-E = Single-E	DJOL[1]xTT[2]xLS[3]	- .16	.26	- .62	.54
Multi-E = Single-KW	DJOL[1]xTT[2]xLS[3]	- .44	.26	-1.70	.09
H1D: Single-E as comparison group					
Single-E = Multi-KW	$DJOL_{Est}$	.36	.16	2.29	< .05
	DJOL[1]xTT[2]xLS[3]	-.22	.24	- .90	.37
Single-E = Single-KW	DJOL[1]xTT[2]xLS[3]	-.28	.21	-1.34	.18

[1] DJOL z-scores; [2] Trial Type; [3] Learning Strategy
[*] $p < .05$ for fixed effects of conditions
Note. Full GLMM output for these analyses can be located in Appendix C.

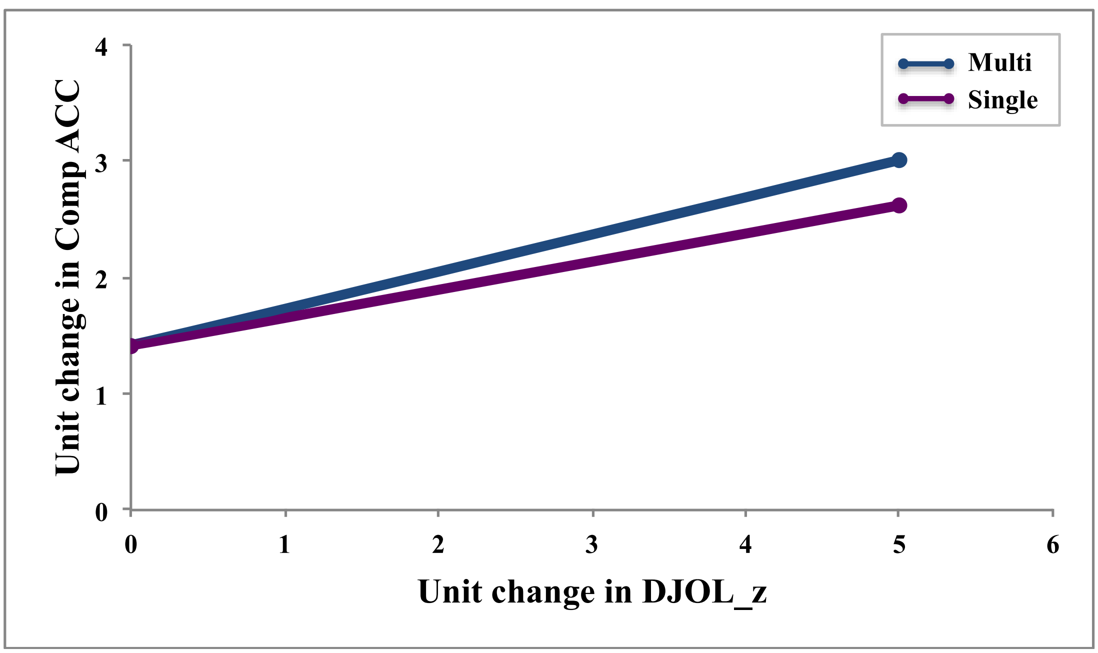

Figure 15. Metacomprehension accuracy for Trial Type at Trial 3 in Experiment 1

Logistic regression comparison for DJOL accuracy between multi and single trial conditions with multi conditions serving as comparison group for Hypothesis 1A.

Hypothesis 1C. Metacomprehension accuracy was predicted to be higher for the effective than ineffective learning strategy conditions, overall.

To test Hypothesis 1C, metacomprehension accuracy was compared between the effective and ineffective learning strategies. The explanation learning strategy served as the comparison group. Results for the GLMM logistic regression is presented in Table 9 and depicted in Figure 16. There was a significant difference in metacomprehension accuracy between the explanation and keyword learning strategies. Participants assigned to the explanation learning strategy (M = .43, SE = .14) were significantly more accurate than participants assigned to the keyword learning strategy (M_{change} = -.33, SE = .16) in their metacomprehension accuracy slopes, as indicated by the negative slope value. Therefore, Hypothesis 1C was supported: metacomprehension accuracy was higher for the effective than ineffective learning strategy.

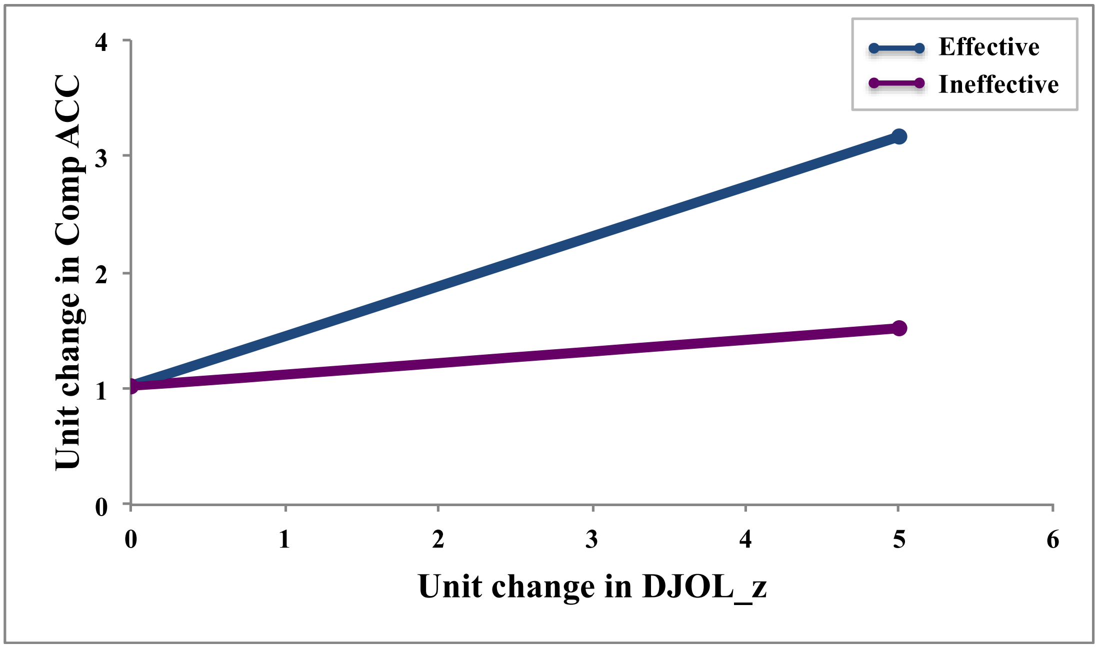

Figure 16. Metacomprehension accuracy for Learning Strategy at Trial 3 in Experiment 1

Logistic regression comparison for DJOL accuracy between effective and ineffective learning strategy conditions with effective learning strategy serving as comparison group for Hypothesis 1C.

Hypothesis 1D. Trial Type and Learning Strategy were predicted to interact; metacomprehension accuracy was predicted to be highest for the Multi Trial Effective condition.

To test for Hypothesis 1D, the interaction effect of Trial Type and Learning Strategy on metacomprehension accuracy was analyzed. Multiple planned comparisons were conducted to compare different combination comparisons. Results for the GLMM logistic regression is presented in Table 9 and depicted in Figure 17. Each condition served as comparison group to every other condition. There was not a significant interaction effect on metacomprehension accuracy between Trial Type and Learning Strategy $p > .05$ for all comparisons.

Therefore, Hypothesis 1D was not supported.

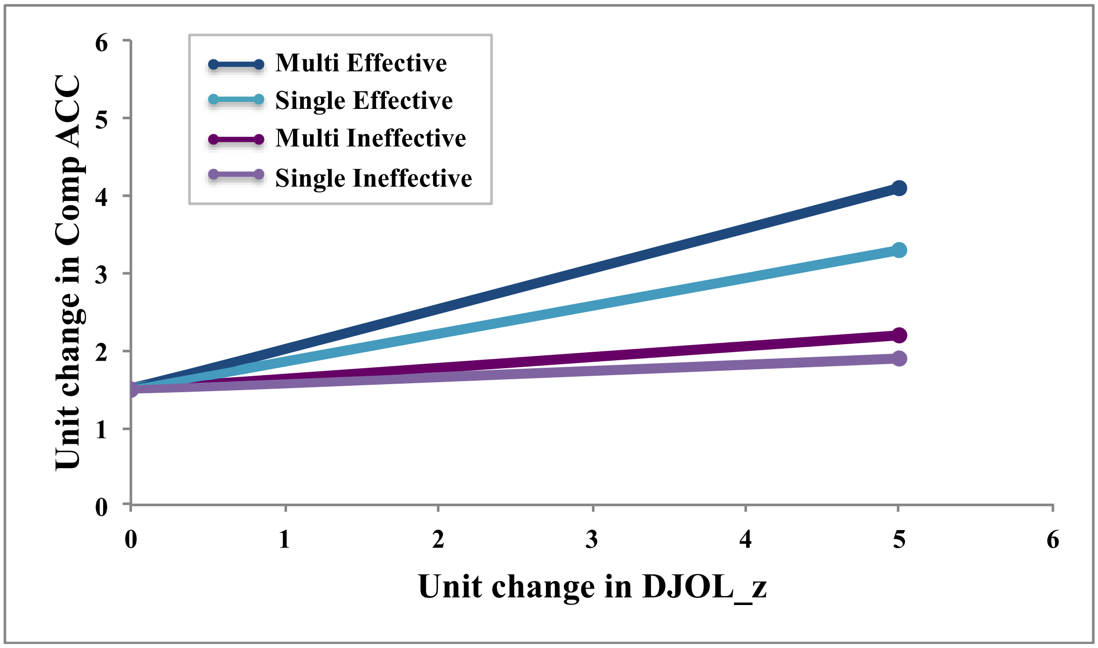

Figure 17. Metacomprehension accuracy for Trial Type and Learning Strategy at Trial 3 in Experiment 1

Logistic regression comparison for DJOL accuracy amongst all conditions with multi effective condition serving as comparison group for Hypothesis 1D.

Summary of Hypotheses Testing Results

For comprehension accuracy, there was a significant main effect of Trial Type, but no effect of Learning Strategy. DJOLs magnitude tracked the pattern of comprehension accuracy; there was a significant main effect of Trial Type, but no effect of Learning Strategy. The main effect of Trial Type on metacomprehension accuracy was not significant; there were no difference between multi and single trial conditions. Although metacomprehension accuracy was not better for multiple than single trials, comprehension accuracy was better for multiple trials.

Metacomprehension accuracy did vary significantly with Learning Strategy. Participants assigned to the explanation learning strategy had a significantly more positive slope change than participants assigned to the keyword learning strategy, indicating that the participants assigned to the explanation learning strategy were more accurate. Interestingly, metacomprehension accuracy did not interact with Trial Type and Learning Strategy. Although metacomprehension accuracy improved when the effective learning strategy was used, that benefit was not observed when Trial Type was included as a factor. There was a benefit of using the effective learning strategy on metacomprehension accuracy over the ineffective strategy, but not a benefit of having multiple trials. This finding for metacomprehension accuracy was the opposite of the finding for comprehension accuracy; comprehension accuracy was better for multiple trials and did not vary with Learning Strategy. In effect, metacomprehension and comprehension accuracy were dissociated with regard to the Trial Type and Learning Strategy factors.

Confidence Judgments

Although specific hypotheses were not generated for CJs, the sensitivity and accuracy of CJs is reported here.

Confidence judgment sensitivity. The magnitude of CJ predictions was analyzed using the LMEM approach. The effects of Trial Type and Learning Strategy on the magnitude of CJ predictions were computed. Results for the LMEM regression are presented in Table 10. Participants who had multiple trials (M = 86.79, SE = 1.02) gave significantly higher CJs than participants who had a single trial (M = 69.69, SE = .91).

The main effect of Learning Strategy was not significant. Participants assigned to the explanation learning strategy (*M* = 78.35, *SE* = .96) gave similar CJs as participants assigned to the keyword learning strategy (*M* = 78.13, *SE* = .98). The interaction between Trial Type and Learning Strategy was not significant. Mean CJ sensitivity for each condition is depicted in the graph below (see Figure 18).

Table 10

LMEM fixed effects results for planned comparisons for CJ sensitivity for Trial 3 in Experiment 1(DV: CJ)

	Predictors	Estimate	*SE*	*df*	*t*	*p*
Main Effect of Trial Type and Learning Strategy						
Multi > Single	Intercept	86.74	2.42	30	35.17	$< .01^{**}$
	TT^2	-17.05	2.67	64	-6.40	$< .01^{**}$
E = KW	Intercept	77.64	2.87	25	27.10	< .01
	LS^3	- .81	3.24	59	- .25	.80
Interaction Effect: Multi-E as comparison group						
Multi-E = Multi-KW	Intercept	85.50	3.23	36	26.49	< .01
	TT^2xLS^3	2.57	3.86	105	.07	.50
Multi-E > Single-E	TT^2xLS^3	-14.31	3.63	103	-3.94	$< .01^{**}$
Multi-E > Single-KW	TT^2xLS^3	-17.32	4.17	41	-4.16	$< .01^{**}$
Interaction Effect: Multi-KW as comparison group						
Multi-KW > Single-E	Intercept	88.07	2.98	63	29.55	< .01
	TT^2xLS^3	-16.88	3.73	98	-4.52	$< .01^{**}$
Multi-KW > Single-KW	TT^2xLS^3	-19.88	3.87	68	-5.14	$< .01^{**}$
Interaction Effect: Single-KW as comparison group						
Single-KW = Single-E	Intercept	71.20	3.24	26	21.95	< .01
	TT^2xLS^3	-3.01	3.93	44	- .77	.45

[2] Trial Type; [3] Learning Strategy
** $p < .01$ for fixed effects of conditions
Note. Full LMEM output for these analyses can be located in Appendix C.

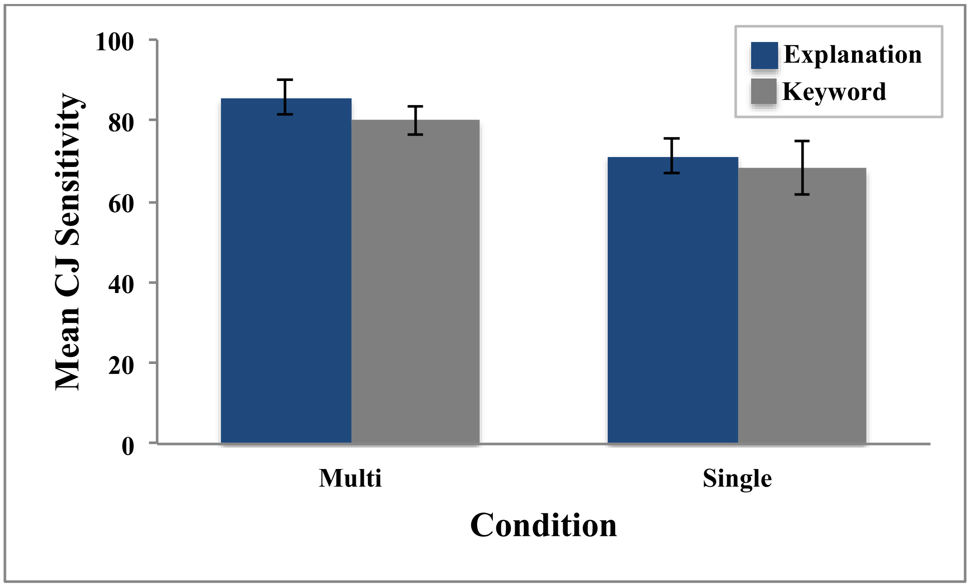

Figure 18. Confidence judgments for all conditions at Trial 3 in Experiment 1

Mean CJ sensitivity for all conditions at Trial 3. Error bars represent standard errors.

Confidence judgment accuracy. The accuracy of CJs was also calculated using the GLMM approach. Centralized CJs served as predictors for comprehension accuracy to determine CJ accuracy. The effects of Trial Type and Learning Strategy on CJ accuracy were analyzed. Results for the GLMM logistic regression are presented in Table 11 and depicted in Figure 19 and 20. There were no significant main or interaction effects.

Table 11

GLMM fixed effects results for planned comparisons for CJ accuracy for Trial 3 in Experiment 1 (DV: comprehension accuracy)

	Predictors	Estimate	*SE*	*z*-Wald	*p*
Main Effect of Trial Type and Learning Strategy					
Multi = Single	$CJ_{Est.}$	.55	.17	3.36	< .01
	CJ[1] xTT[2]	.14	.16	.83	.41
E = KW	CJ_{Est}	.06	.19	3.55	< .01
	CJ[1]xLS[3]	- .03	.18	- .16	.87
Interaction Effect: Multi-E as comparison group					
Multi-E = Multi-KW	$CJ_{Est.}$	.56	.27	2.07	< .01
	CJ[4]xTT[2]xLS[3]	.03	.29	.10	.92
Multi-E = Single-E	CJ[4]xTT[2]xLS[3]	.16	.26	.65	.52
Multi-E = Single-KW	CJ[4]xTT[2]xLS[3]	.10	.26	.36	.72
Interaction Effect: Multi-KW as comparison group					
Multi-KW = Single-E	$CJ_{Est.}$	.59	.18	3.31	< .01
	CJ[4]xTT[2]xLS[3]	.14	.22	.62	.53
Multi-KW = Single-KW	CJ[4]xTT[2]xLS[3]	.07	.23	.30	.77
Single-E as comparison group					
Single-E = Single-KW	$CJ_{Est.}$	.72	.18	4.00	< .01
	CJ[4]xTT[2]xLS[3]	- .06	.22	- .31	.76

[4] CJ z-scores; [2] Trial Type; [3] Learning Strategy
[*] $p < .05$ for fixed effects of conditions
Note. Full GLMM output for these analyses can be located in Appendix C.

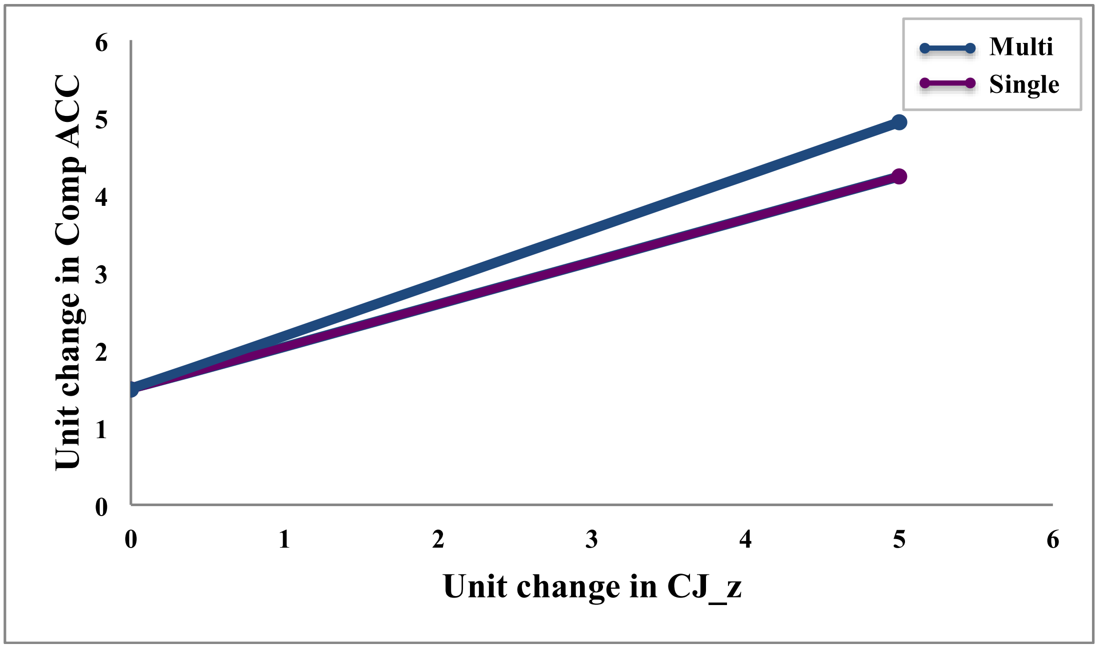

Figure 19. CJ accuracy for Trial Type at Trial 3 in Experiment 1

Logistic regression comparison for CJ accuracy between multi and single trial conditions with multi conditions serving as comparison group.

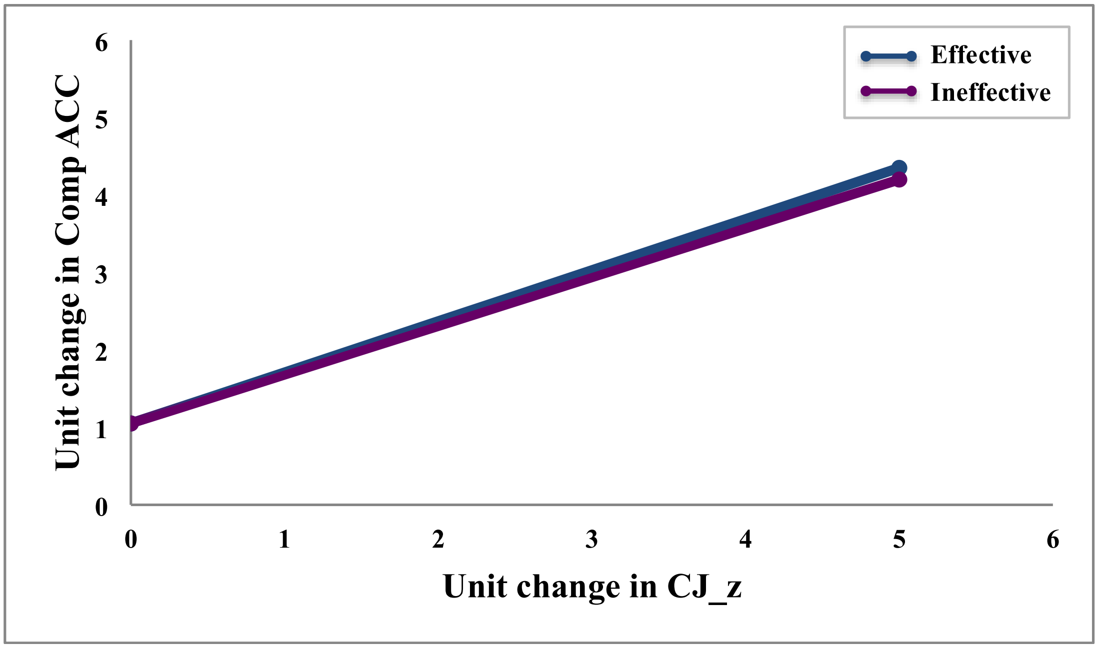

Figure 20. CJ accuracy for Learning Strategy at Trial 3 in Experiment 1

Logistic regression comparison for CJ accuracy between effective and ineffective learning strategy conditions with effective learning strategy serving as comparison group.

Metacomprehension Accuracy for Multi Trial Conditions Only

The dissociation between metacomprehension and comprehension accuracy prompted additional investigation. Because all of the hypothesis testing focused a comparison of results between multiple trials and single trials just on Trial 3, the degree to which the results were changing across the multiple trials was not tested. The analysis of Trial Type presented thus far tested a between-subjects comparison between participants who had a single trial to the third trial of those who had multiple trials. Any benefits of the multi trial condition could have been obscured by just examining the end result of having multiple trials. Another way to examine the effect of multiple trials is to conduct a within-subjects analysis comparing performance on Trial 3 to performance on each previous trial just for those participants in the multi trial conditions. Conducting within-subjects comparisons allows for a direct test of the hypothesis that improved metacomprehension accuracy across trials also improved comprehension accuracy across trials.

Results from the GLMM logistic regression for the multiple planned comparisons are presented in Table 12 and depicted in Figures 21 - 23. The main effect of Trial on metacomprehension accuracy was tested. There was no significant main effect of Trial on metacomprehension across trials. Post hoc planned comparisons showed that metacomprehension accuracy was similar for all three trials, $p < .05$.

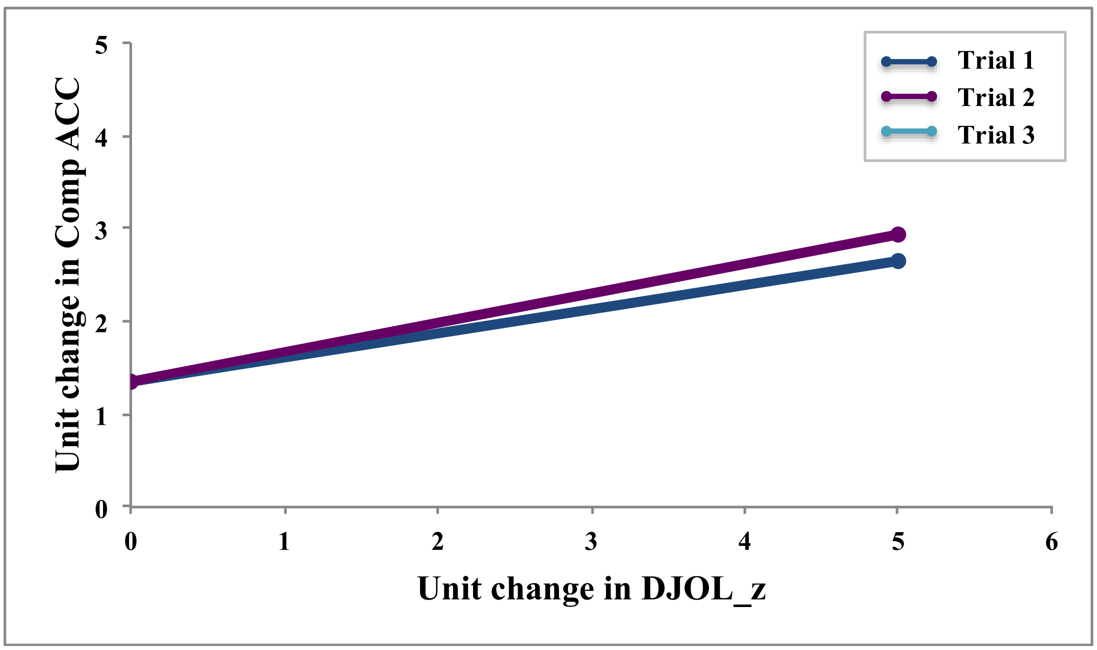

Figure 21. Metacomprehension accuracy across Trials for multi trial conditions in Experiment 1

Logistic regression comparison for DJOL accuracy across trials with Trial 1 serving as comparison group.

Table 12

GLMM fixed effects results for planned comparisons for metacomprehension accuracy within multi trial conditions in Experiment 1 (DV: comprehension accuracy)

	Predictors	Estimate	*SE*	*z*-Wald	*p*
Main Effect of Trial: Trial 1 as comparison group					
Trial 1 = Trial 2	$DJOL_{Est.}$	.26	.13	1.97	< .05
	$DJOL^1 xT^5$	.06	.16	.36	.72
Trial 1 = Trial 3	$DJOL^1 xT^5$	- .01	.17	- .03	.97
Main Effect of Trial: Trial 2 as comparison group					
Trial 2 = Trial 3	$DJOL_{Est.}$	.31	.13	1.97	< .05
	$DJOL^1 xT^5$	- .06	.16	- .39	.70
Main Effect of Trial for Explanation Learning Strategy					
Trial 1 = Trial 2	$DJOL_{Est.}$	.48	.22	2.22	< .05
	$DJOL^1 xT^5$	- .41	.24	-1.74	.08
Trial 1 = Trial 3	$DJOL^1 xT^5$	.00	.25	.03	.98
Trial 3 = Trial 2	$DJOL_{Est.}$	.50	.20	2.54	< .05
	$DJOL^1 xT^5$	- .02	.23	- .01	.99
Main Effect of Trial for Keyword Learning Strategy					
Trial 2 > Trial 1	$DJOL_{Est.}$	.52	.16	3.23	< .01
	$DJOL^1 xT^5$	- .51	.22	-2.27	$< .05^{*}$
Trial 2 > Trial 3	$DJOL^1 xT^5$	- .50	.23	-2.17	$< .05^{*}$
Interaction Effect: Multi-KW @ Trial 2 as comparison condition					
KW@T2 < E@T1	$DJOL_{Est.}$	.52	.16	3.23	< .01
	$DJOL^1 xT^5$	.89	.30	2.94	$< .01^{**}$
KW@T2 < E@T3	$DJOL^1 xT^5$	.89	.32	2.81	$< .01^{**}$

[1] DJOL z-scores; [5] Trial

* $p < .05$, ** $p < .01$ for fixed effects of conditions

Note. Only results that yielded at least a significant $DJOL_{Est}$ value were included in Table 12 because an insignificant $DJOL_{Est}$ would mean that the grand mean of the DJOL accuracy was not better than chance.

Note. Full GLMM output for these analyses can be located in Appendix C.

To test for interactions between Trial and Learning Strategy, several planned comparison analyses were conducted comparing each learning strategy within each trial and then comparing each trial within each learning strategy. The GLMM comparison of Trial for the keyword condition yielded higher metacomprehension accuracy for Trial 2 as compared to that of Trials 1 or 3; presented in Figure 22. Although the comparison of the explanation to the keyword learning strategy in Trial 2 did not differ significantly, there was a significant interaction when the Trial 2 keyword condition served as the comparison group for the explanation condition for Trials 1 and 3; presented in Figure 23. Metacomprehension accuracy was better for the explanation learning strategy for Trials 1 and 3, even though Trial 2 produced the best metacomprehension accuracy of the keyword condition trials.

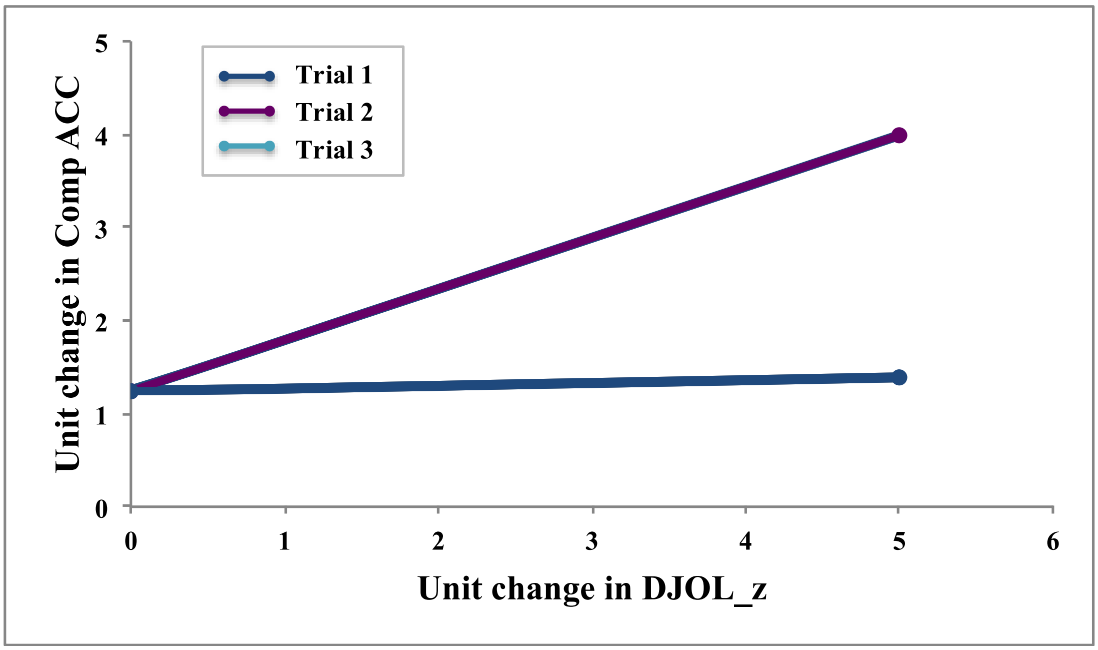

Figure 22. Metacomprehension accuracy across trials for multi keyword conditions in Experiment 1

Logistic regression comparison for DJOL accuracy across trials within multi keyword conditions with Trial 1 of multi keyword serving as comparison group.

Having multiple opportunities to study and make predictions about the text benefited participants using the keyword strategy, but that benefit was limited. It only occurred on the second trial; no additional benefit was obtained with a third trial. In addition, although this improvement in metacomprehension accuracy occurred, it was still not as high as for those who used the explanation learning strategy. Those participants using explanation had higher metacomprehension accuracy at Trials 1 and 3 than those participants using the keyword learning strategy, even taking the Trial 2 improvement into consideration.

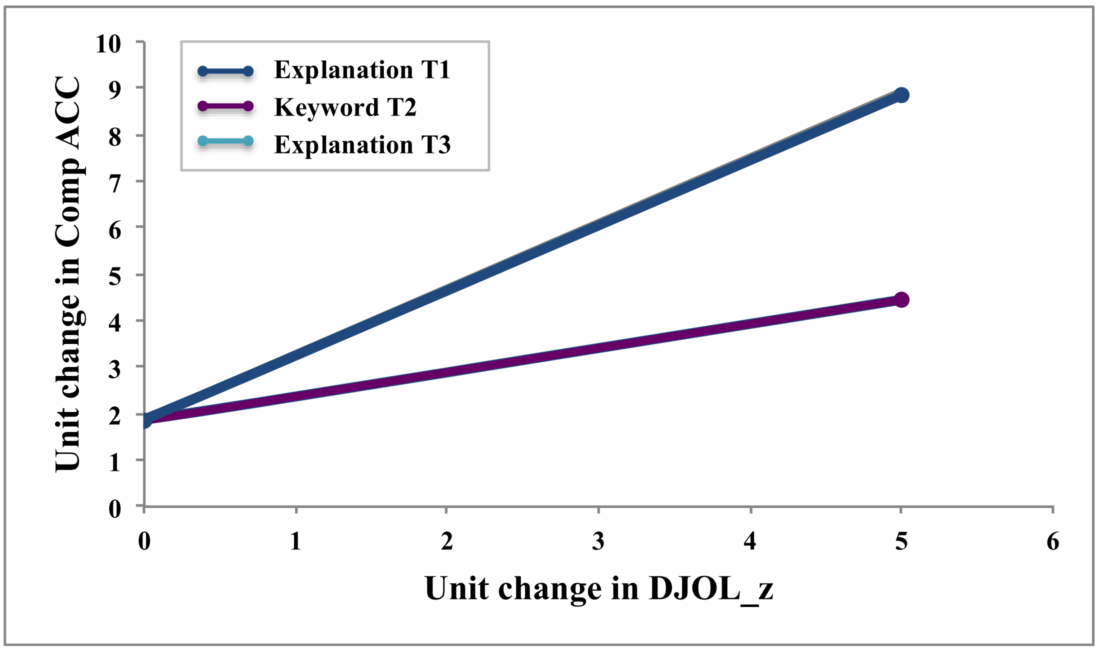

Figure 23. Metacomprehension accuracy across trials for Trials and Learning Strategy in Experiment 1

Logistic regression comparison for DJOL accuracy for with Trial 2 of the multi keyword condition serving as comparison group.

In summary, experiencing multiple trials only improve metacomprehension accuracy for participants who used the ineffective learning strategy. Having multiple trials did not add to the benefit already obtained by using the effective learning strategy. Apparently, any benefit was gained just by using the effective learning strategy to begin with. This main finding could potentially provide an explanation as to why effective learning strategies improve metacomprehension accuracy in the literature even when multiple trial procedures were not used. Using the effective learning strategy from the start allowed for cues to be generated that were diagnostic of comprehension, leading to accurate metacomprehension.

Participant-level Metacomprehension Analysis

The finding that metacomprehension accuracy was best for Trials 1 and 3 for the explanation learning strategy, but best on Trial 2 for the keyword learning strategy demonstrated that testing the results only on Trial 3 obscured subtler effects of Trial Type and Learning Strategy on metacomprehension accuracy. Aggregating across all participants could have done the same to the effect of metacomprehension accuracy on comprehension accuracy in the previous analysis. For the explanation learning strategy, metacomprehension accuracy in the aggregate did not differ across trials. However, on closer examination the participant-level data showed that some participants improved across trials, some got worse, and some stayed the same. Given the theoretical stance that improving metacomprehension should also lead to an improvement of comprehension accuracy, comprehension and metacomprehension accuracy was compared between a subset of participants whose metacomprehension accuracy improved across trials and a subset whose metacomprehension accuracy did not improve as further testing of this stance.

Using participant coefficient values generated from the GLMM analysis that accounted for random effects of participant and question, the relative metacomprehension accuracy change of slope was used to create change scores between each trial for each participant. Using this change value, a new post-hoc categorical variable was created—

Metacomprehension Accuracy Improvement—for each two change phases (T1-T2 and T2-T3). Participants were categorized in either the "improved" or "unimproved" group based on the difference value for each change phase. Table 13 presents the descriptive statistics for the number of participants in the newly coded categories for each change phase.

Table 13

Descriptive statistics for newly formed category based on participant's change in metacomprehension (DJOL) accuracy between trials in Experiment 1

N = 62	Trial 1 – Trial 2	Trial 2 – Trial 3
Improved	45	31
Unimproved	17	31

As presented in Table 13, 70% of participants showed an improvement in metacomprehension accuracy from Trials 1 to 2. Half of the participants showed improvement in metacomprehension accuracy from Trials 2 to 3. Aggregating the data across trials, as was done in the hypothesis testing analyses, obscured these findings. Univariate ANOVAs were conducted to measure whether Metacomprehension Accuracy Improvement—improved versus unimproved—from Trials 1 to 2 impacted comprehension accuracy at Trial 2, and to measure whether Metacomprehension Accuracy Improvement from Trials 2 to 3 impacted comprehension accuracy at Trial 3.

Metacomprehension Accuracy Improvement from Trials 1 to 2 significantly impacted comprehension accuracy at Trial 2, F (1,60) = 78.86, p < .01, partial $\eta^2 = .57$. Pairwise comparison showed that comprehension accuracy was higher at Trial 2 for participants with improved metacomprehension accuracy from Trials 1 to 2 as compared to participants with unimproved metacomprehension accuracy from Trials 1 to 2. Mean comparisons are presented in Table 14 and depicted in Figure 24.

Because metacomprehension accuracy is a correlation between metacomprehension and comprehension, using it as a predictor of comprehension accuracy could inflate the findings. An additional Pearson product-moment correlation of the change in metacomprehension accuracy between Trials 1 and 2 and comprehension accuracy at Trial 2 was conducted. There was a significant positive relationship between the two factors, $r(62) = .91$, $p < .01$. These correlation findings provide strong support for the initial Univariate ANOVA analysis; participants' comprehension accuracy paralleled their metacomprehension accuracy. When metacomprehension accuracy improved, so did comprehension accuracy. When metacomprehension accuracy did not improve—or even declined—across trials 1 and 2, comprehension accuracy followed suit.

Table 14

Mean comparisons for comprehension accuracy by metacomprehension accuracy improvement between Trials 1 and 2 and Trials 2 and 3 in Experiment 1

	MetaCompACC	Mean Comp ACC
T1-T2	Improved	.79 (.01)
	Unimproved	.57 (.02)
T2-T3	Improved	.86 (.02)
	Unimproved	.63 (.02)

Note. Standard Error values are in parenthesis beside the mean values.

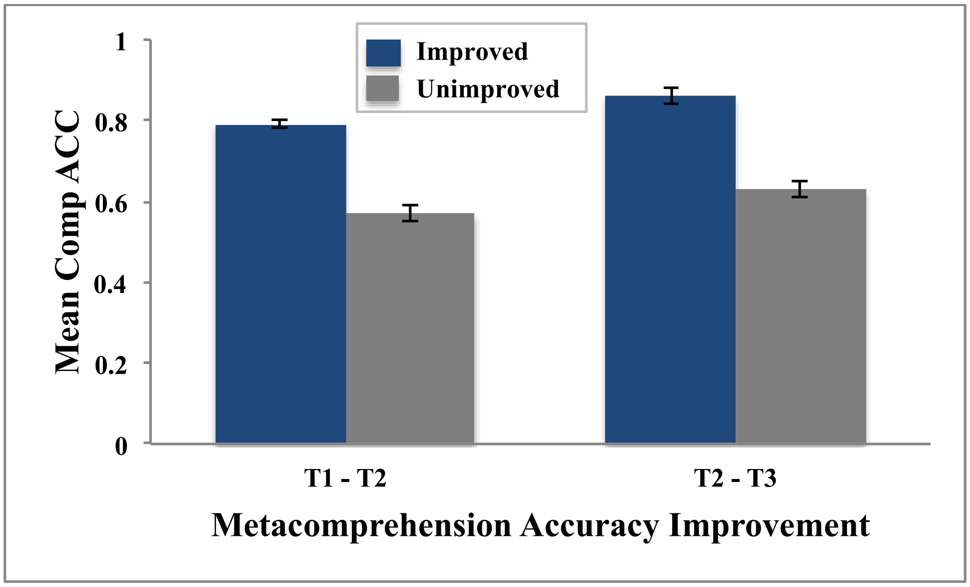

Figure 24. Metacomprehension accuracy improvement from previous trial on comprehension accuracy at current trial in Experiment 1

Mean comprehension accuracy for improved vs. unimproved metacomprehension accuracy across trials. Error bars represents standard errors.

There was also a significant main effect of Metacomprehension Accuracy Improvement from Trial 2 to Trial 3 on comprehension accuracy at Trial 3, F (1,60) = 74.67, $p < .01$, partial $\eta^2 = .55$. Pairwise comparison showed comprehension accuracy was higher at Trial 3 for participants with improved metacomprehension accuracy from Trials 2 to 3 as compared to participants with unimproved metacomprehension accuracy from Trials 2 to 3. Mean comparisons are presented in Table 14 and depicted in Figure 24.

The Pearson's correlation of the change in metacomprehension accuracy between Trials 2 and 3 with comprehension accuracy at Trial 3 found that the two were strongly correlated, $r(62) = .97$, $p < .01$. Consistent with the previous findings, these additional correlation findings provide strong support that when participants showed any direction of change—improved or unimproved—in their metacomprehension accuracy from trials 2 to 3, their comprehension accuracy also followed suit.

Comprehension performance at Trial 2 was two test questions higher on average (*M* = .22) for participants who showed improvement in metacomprehension accuracy from Trial 1 to Trial 2 as compared to participants who did not improve in metacomprehension accuracy from Trial 1 to Trial 2. Similarly, comprehension performance at Trial 3 was two test questions higher on average (*M* = .23) for participants who showed improvement in metacomprehension accuracy from Trials 2 to 3 as compared to participants who did not improve.

The categorization of participants' change values in metacomprehension used thus far in this participant-level analysis is a between-subjects comparison; participants categorized as improved from trials 1 to 2 may not be the same participants as those categorized as improved from trials 2 to 3. A separate Univariate ANOVA analysis was conducted to compare participants who showed metacomprehension accuracy improvement across trials (n = 31) and those who did not (n = 31) on comprehension accuracy at Trial 3. There was also a significant main effect of Metacomprehension Accuracy Improvement Across Trials on comprehension accuracy at Trial 3, F (1,60) = 74.67, $p < .01$, partial $\eta^2 = .55$. Pairwise comparison showed comprehension accuracy to be higher at Trial 3 for participants with improved metacomprehension accuracy across trials (*M* = .86, *SE* = .02) as compared to participants with unimproved metacomprehension accuracy across trials (*M* = .63, *SE* = .02). Mean comparison is depicted in Figure 25.

Pearson's correlations were conducted to provide additional support of the Univariate ANOVA analysis. A Pearson's correlation between the change in metacomprehension accuracy from Trials 1 to 3 (across trials) and comprehension accuracy at Trial 3 found that the two were strongly correlated, $r(62) = .97$, $p < .01$. This additional correlation finding provides strong support that participants' comprehension accuracy varied significantly with their metacomprehension accuracy improvement—or lack thereof—across trials.

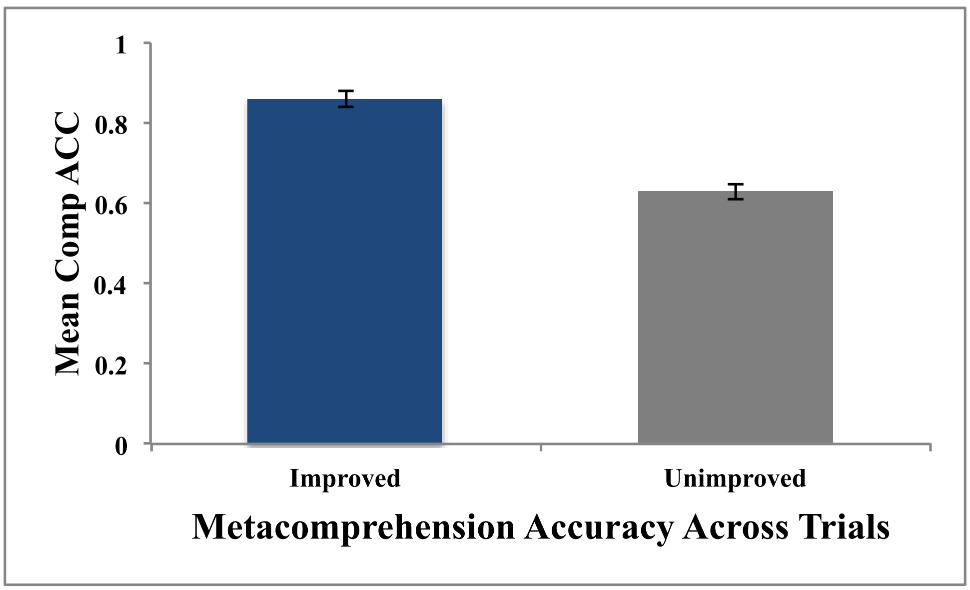

Figure 25. Metacomprehension accuracy improvement across trials on comprehension accuracy at Trial 3 in Experiment 1

Mean comprehension accuracy at Trial 3 for improved vs. unimproved metacomprehension accuracy across all trials. Error bars represents standard errors.

The findings from the analysis at the participant level support the predicted finding that improvement in metacomprehension accuracy would also improve comprehension accuracy. In combination with the findings from the hypotheses testing, these findings partially fulfilled the goal of Aim 1. Although having multiple trials did not improve metacomprehension accuracy for every participant, those for whom it did also had better comprehension accuracy. This finding supports the use of the multi trial metacomprehension paradigm toward improving both metacomprehension and comprehension accuracy. The findings from this experiment set the basis for Aim 2. Aim 2, explored in Experiment 2, investigated the impact of improved metacomprehension accuracy on the control process via the selection of learning strategy so that comprehension accuracy improved as well.

Post-Experiment Interview Questionnaire

After participants completed all sections of the experiment, they answered some questions about the experiment using the Post-Experiment Interview. Participants were instructed to answer as honestly as possible about their experience with the experiment. Each of the post-experiment interview questions were analyzed using the appropriate statistical test in order to determine whether there was a difference in the answers provided by participants across the different conditions. No participants were eliminated based on their responses on the post-experiment interview, and the only significant difference in responses between the multiple and single trial conditions were that people in the single trial conditions rated the test questions as more difficult than those in the multiple trials conditions. Details of the analysis of the post-experiment interview questions are presented in Appendix D.

Discussion

The aim of Experiment 1 was to determine whether metacomprehension accuracy improved across trials using the multi trial metacomprehension paradigm as compared to a single trial paradigm. In addition to investigating the benefit of multiple over single trials on metacomprehension accuracy, this experiment also aimed to measure the impact of presumably effective versus ineffective learning strategies on metacomprehension accuracy by having participants use either the delayed explanation or keyword learning strategy. Comprehension accuracy was also compared between multiple and single trials and between the two learning strategies. At Trial 3, comprehension accuracy was better for multiple than for single trials. Comprehension was equally accurate for the two learning strategies. DJOL sensitivity showed that the magnitude of metacomprehension predictions tracked the main effect pattern of Trial Type and Learning Strategy on comprehension accuracy; DJOLs were higher for the multiple than for the single trials and similar for the two learning strategies.

Although multiple trials were beneficial to comprehension accuracy when compared to single trials, using the effective learning strategy did not contribute to comprehension accuracy, even when participants had the opportunity to restudy the text. This finding is surprising because the effective learning strategy used was identified as such by the comprehension literature (i.e., Bretzing & Kulhavy, 1979; Chi et al., 1994). However, the finding was consistent with prior research that also failed to find an improvement in comprehension when using "effective" learning strategies (i.e., Thiede & Anderson, 2003; Thiede et al., 2003). Therefore, consistent with prior research, although comprehension was not better for the “effective” strategy, metacomprehension accuracy was.

Although having multiple trials improved comprehension, it did not have the same effect on metacomprehension; metacomprehension accuracy was equal for the multiple and single trial conditions. Metacomprehension and comprehension accuracy were dissociated for the Trial Type factor. Evidence for an improvement in metacomprehension accuracy with multiple trials might have been overpowered by the effect of effective versus ineffective learning strategies. Consistent with past literature on the benefits of effective learning strategies on metacomprehension accuracy (Anderson & Thiede, 2008; Thiede & Anderson, 2003; Thiede et al., 2003), metacomprehension accuracy was higher for the effective than the ineffective learning strategy. When it comes to improving metacomprehension accuracy, the improvement due to using the effective learning strategy appear to have trumped the benefit of multiple trials. Metacomprehension accuracy obtained at Trial 1 for the explanation learning strategy was already high, leaving less room for improvement with multiple trials for that learning strategy.

In order to further investigate this hypothesis, additional analyses were conducted only on the multi trial conditions. This analysis showed that there was a significant improvement of metacomprehension accuracy at Trial 2 as compared to Trials 1 and 3 for participants assigned to the ineffective learning strategy; metacomprehension for participants assigned to the ineffective learning strategy benefited somewhat from multiple trials. However, although metacomprehension accuracy at Trial 2 was highest for participants assigned to the ineffective learning strategy, it was still not higher than the metacomprehension accuracy at Trial 1 and Trial 3 for participants assigned to the effective learning strategy. This finding further supports the assumption that metacomprehension accuracy for the participants who used the effective learning strategy peaked at Trial 1. Although the ineffective learning strategy was able to produce higher metacomprehension accuracy at Trial 2, it was still not as effective as producing accurate metacomprehension as the effective learning strategy at Trial 1.

Because not all participants improved their metacomprehension accuracy, a subset of participants who showed improvement in metacomprehension accuracy across the trials was identified. For this subset of participants whose metacomprehension accuracy improved across trials, comprehension accuracy also improved. Comprehension accuracy did not improve for a comparison group of participants whose metacomprehension did not improve. These findings were independent of the learning strategy that was used; for the participants whose metacomprehension accuracy improved across trials, the strategy they used was not a significant factor. Because of the potential for inflating the relationship between metacomprehension and comprehension accuracy due to the fact that metacomprehension accuracy includes comprehension accuracy in its calculation Pearson's correlations were conducted between the change in metacomprehension across trials and comprehension accuracy on that particular trial. In every case, this correlation was high, providing strong support for the conclusion that participants' comprehension accuracy was influenced by their metacomprehension accuracy. Apparently the critical benefit of multiple trials is only evident if the effect of multiple trials is to improve metacomprehension, and apparently, improving metacomprehension is more crucial than the learning strategy used.

Although the multi trial metacomprehension paradigm had multiple trials, the paradigm did not allow for improved metacomprehension accuracy to impact control processes in such a way that also improved comprehension accuracy. Participants did have an opportunity to restudy the text, but they were constrained to using the learning strategy assigned to them to do so. Theoretically, the dynamic relationship between the object and the meta level allows for the improved metacomprehension accuracy to inform learning strategy decisions such that an effective learning strategy can be selected. However, the manipulation of this experiment denied participants the opportunity for their improved monitoring process to inform their control process of strategy selection on a new trial.

In Experiment 2, additional conditions were added that alternated the two learning strategies between Trials 1 and 2 as a comparison to the conditions from Experiment 1 that assigned the same learning strategy to both trials. The alternating conditions allowed participants to experience the effects of both learning strategies on comprehension before they had to select the learning strategy they wanted to use to at Trial 3. Allowing participants to select the learning strategy at Trial 3 was a direct measure of the implementation of the control process of learning strategy selection. The impact of metacomprehension accuracy and prior experience with the learning strategies on this selection was examined in Experiment 2.

CHAPTER VI

EXPERIMENT 2

The aim of this experiment was to examine the impact of improved metacomprehension accuracy on the control process of learning strategy selection. Specifically, this experiment investigated whether improved metacomprehension accuracy via the monitoring process, achieved by using the multi trial metacomprehension paradigm and learning strategy experience, affects control processes so that the effective learning strategy was more likely to be selected on the final trial.

Additionally, because the ultimate goal of optimal strategy selection is improved comprehension accuracy, this experiment also investigated whether improved control process learning strategy selection via improved monitoring accuracy also benefited comprehension accuracy.

Method

Participants and Design

A total of 168 participants were recruited from Mississippi State University undergraduate students currently enrolled in Introductory Psychology classes via the Psychology Research Program SONA-system website. Sample size calculation for five conditions was conducted using the G*Power analysis program (Faul, Erdfelder, Lang, & Buchner, 2007) for a medium effect size of .25, and a power of .80. Participants were at least 18 years old and have English as their native language. Additionally, participants must not have participated in the first experiment. Participants received research credits for their participation. Similar to Experiment 1, participants received the opportunity to participate in a raffle drawing for a chance to win a $25 Amazon gift card to promote motivation for participation.

The design of this experiment was a 2 (Learning Strategy: ineffective, effective) x 2 (Learning Strategy Order: mixed, same) full factorial between-subjects design with an additional single trial control condition. Table 15 shows the five conditions for Experiment 2.

Table 15

Learning strategies assigned at Trial 1 and 2 and to be self-selected at Trial 3 for Experiment 2

Condition	Trial 1	Trial 2	Trial 3
Single Trial	-	-	
Multi Mixed I-E	Keyword	Explanation	Explanation
Multi Mixed E-I	Explanation	Keyword	OR
Multi Same I-I	Keyword	Keyword	Keyword
Multi Same E-E	Explanation	Explanation	

Materials

Text. The same expository text selected for Experiment 1 was used in this experiment (see Appendix B). Participants who had previously participated in Experiment 1 were not permitted to sign up for this experiment.

Comprehension test questions. The same three separate sets of twelve four-alternative multiple choice comprehension test questions, including two additional manipulation check questions for the text used in Experiment 1 were used in this experiment (see Appendix B). Identical to Experiment 1, all participants were administered the same twelve test questions at each trial but the order of the questions was counterbalanced within the trial.

Consent form. Identical to Experiment 1, an IRB approved stamped informed consent form informed participants about the general aim and procedure of the experiment. Additionally, the consent form explained the minimal risk involve and the benefits for their participation in the experiment. Participants were informed that their participation in this experiment is voluntary; therefore they may choose to cease participation at any time during the experiment and will not be penalized if they choose to do so (see Appendix A).

Post-experiment interview. Participants were asked to provide feedback on whether they were aware of the experimental manipulation in this experiment. Additionally, participants were asked if they were aware of the reason why they selected whichever learning strategy they did in the final trial. Other example questions include their awareness of the change in their ability in making memory judgments as well as answering the test questions for the different trials with learning strategies assigned to them or selected by them.

Debriefing form. A debriefing form was created that explains the nature of the experiment in detail. Participants were informed about the dynamic relationship of the monitoring via metacomprehension judgments and the control processes via self-selection of learning strategy on the final trial. Participants were also informed of the benefit of the effective (delayed explanation) over the ineffective (keyword) learning strategy on their metacomprehension and comprehension accuracy. Participants were also provided with contact information of the Eakin Memory and Metamemory research lab and the researcher-in-charge in case they have any further questions about the experiment (see Appendix A).

General Procedure

After providing consent to participate, the experiment began with instructions for the practice phase presented on a PC computer programmed using the EPrime 2.0 software (Psychology Software Tools, Pittsburgh, PA). All instructions and experimental tasks were presented within the EPrime 2.0 program via the computer monitor. The practice phase familiarized participants with the entire procedure of a single trial, using both the learning strategies. Research assistants were trained to monitor participants' progress throughout the experiment and to make sure they were using the assigned learning strategy properly.

Following the practice phase, the experiment began. Identical to Experiment 1, there were three experimental trials, each consisting of three phases: a) reading comprehension, b) metacomprehension judgment, and c) comprehension test. However, in this experiment, only the first two trials used assigned learning strategies. In Trial 3, participants were allowed to self-select one of the two learning strategy options to study the text. Depending on which condition participants are assigned to, they may use the same or a mix of the learning strategies for the first two trials. Additionally, the control group only had a single trial where participants were allowed to self-select the learning strategy to study the text.

Reading comprehension phase. For the first two trials, participants were assigned to use one of two learning strategies, delayed explanation or keyword. Participants were instructed to read the text as if studying for an exam by using the learning strategy assigned to them. In the third trial, however, participants had the opportunity to self-select one of the learning strategy option.

Procedure for the effective learning strategy. Identical to Experiment 1, the effective learning strategy selected was the delayed explanation learning strategy. Participants first read the text and to induce delay before writing a summary, they were presented with an interval task—Tower of Hanoi—for five minutes. The text was presented on timed slides with a total read time of eight minutes, allowing for five minutes of summary writing later. After the interval task, participants were instructed to write a summary explaining about the text as if they were telling a story about the text using their own words. Participants typed their summary into the computer using the computer keyboard.

Procedure for the ineffective learning strategy. As in Experiment 1, the ineffective learning strategy selected was the keyword learning strategy. If assigned to this learning strategy, participants were provided with four notecards to write down vocabulary terms while they read the text. The text was presented on timed slides with a total read time of 13 minutes, allowing for writing of keyword on the notecards.

In order to make sure that participants were using the assigned learning strategy at each trial, research assistants kept track of the condition participants were assigned to. Additionally, research assistants also checked their summary and notecards before participants were allowed to proceed to the next phase of the trial. This manipulation check was put in place to make sure participants were properly using the learning strategy they were assigned to at the beginning of each trial.

Metacomprehension judgment phase. This phase comes after reading the text. After reading the text and prior to the start of the metacomprehension judgment phase, participants completed an interval task for 5 minutes by solving Tower of Hanoi puzzles. After completing the interval task, participants made individual DJOLs when cued by each of the twelve questions, including the two manipulation-check questions, using the scale of 0 (certain not to remember) to 100 (certain to remember). Participants were informed that they would have a multiple-choice comprehension test, but the alternatives will not be shown during the DJOL.

Comprehension test phase. During the comprehension test phase, participants took a test consisting of ten four-alternative multiple-choice questions. These questions were the same twelve questions (including the two manipulation check questions) participants made DJOLs on in the metacomprehension judgment phase. For each question, participants selected the answer from the four alternative options based on what they had learned from the text. Participants were required to select an answer—they cannot leave an answer blank—and guessing was allowed. Immediately after selecting an answer, participants also made a CJ about the degree to which they felt that the answer they selected was correct. Participants made their CJs using a scale of 0 (not confident at all) to 100 (extremely confident).

Similar to Experiment 1, at the end of the experiment, participants answered some questions in relation to the experiment in the post-experiment questionnaire. Finally, before participants were excused, they were debriefed and assigned research credit. Based on their comprehension scores, they were given raffle tickets equal to the number of questions they got correct.

Figure 26 and 27 show the general procedure for Trials 1 and 2 using the keyword learning strategy and delayed explanation learning strategy, respectively. Trial 3 followed the same procedure except that participants self-selected the learning strategy they wanted to use to study the text. For the Single Trial control condition, participants only performed one trial and selected their own learning strategy from between the two, after they are given instructions about how to do both.

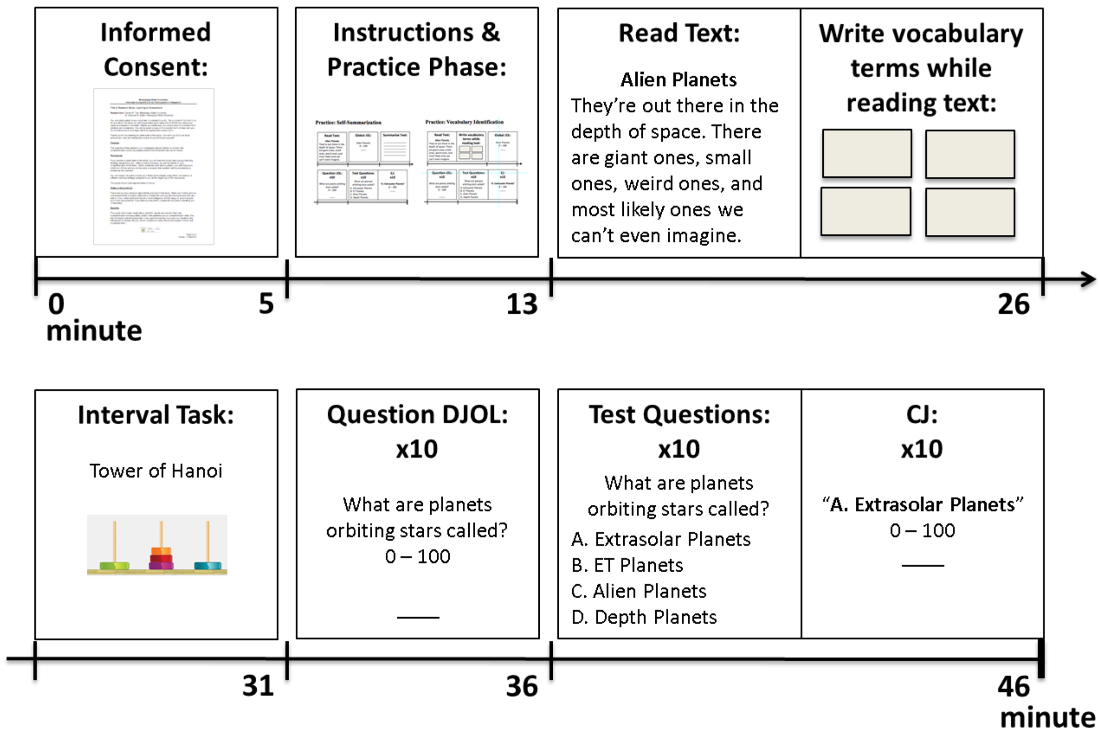

Figure 26. Procedure timeline for Trials 1 and 2 for the keyword-identification learning strategy in Experiment 2

Timeline schematic for Trials 1 and 2 for the keyword learning strategy group. Trial 3 differed only in that participants selected which learning strategy to use.

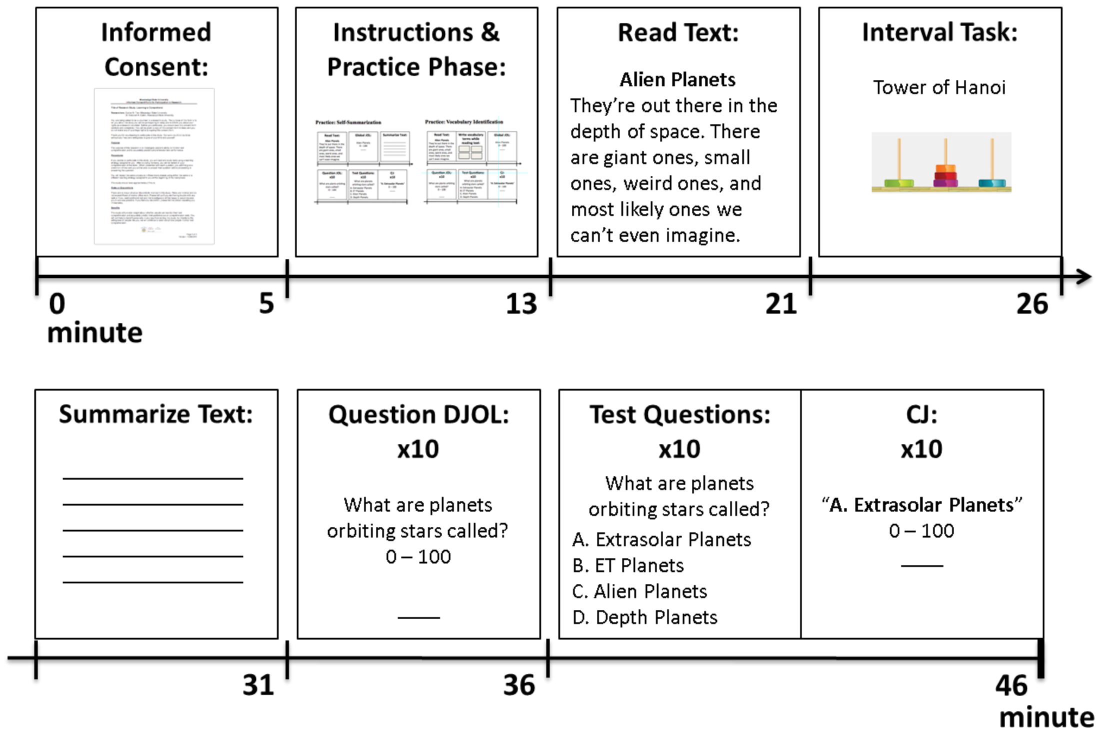

Figure 27. Procedure timeline for Trials 1 and 2 for the delayed explanation learning strategy in Experiment 2

Timeline schematic for Trials 1 and 2 for the delayed explanation group. Trial 3 differed only in that participants selected which learning strategy to use.

Results

The manipulation check questions were first scored in order to determine whether there were participants who should be excluded from analysis for failing to meet the inclusion criterion that participants answer at least one of the questions correctly. None of the participants failed this criterion; the analysis includes all participants who completed the experiment. Identical measures of centralization of all metacomprehension predictions variables (e.g., DJOLs, and CJs) to Experiment 1 were taken in preparation of data analyses. All metacomprehension predictions were converted to z-scores within each participant and each trial. Following the conversion of these predictions, there were 24 incidences—12 DJOLs and 12 CJs—when participants made the same prediction for all question within a trial were given a "0" score to represent the mean value of their predictions. These incidences were replaced with a "0" to represent the participant's mean centralized prediction for that trial.

The data were analyzed using the lme4 package (Bates et al., 2015) in an R statistical computing environment. Generalized linear mixed modeling (GLMM) approach was used when calculating metacomprehension accuracy, a dichotomous variable For all GLMM analyses discussed in this section, a maximal random effect structure was used. Participants and questions within trials were modeled as random effects. All GLMM analyses reported in these results successfully converged using the maximal random effect structure with the help of "bobyqa" optimizer. The full syntax and model output for each hypothesis testing, including the random effects model can be found in Appendix E, Tables E1 – E33.

However, when comparing continuous variables, the magnitude of metacomprehension predictions, the linear mixed effects modeling (LMEM) was used instead. Similarly, all LMEM comparisons discussed in this section included the maximal random effect structure. Both participants and question were modeled as random effects. All LMEM analyses reported successfully converged with the maximal random effect structure.

This Results section is structured similar to the Experiment 1 Results section. The findings in terms of the hypotheses delineated in the Aims section will be discussed, starting with comprehension accuracy, and followed by metacomprehension sensitivity. Then the findings regarding metacomprehension accuracy will be discussed. All hypotheses were based on comparing participants' comprehension and metacomprehension performances at Trial 3, unless otherwise indicated. Additionally, how metacomprehension accuracy impacted learning strategy selection at Trial 3 will be discussed. The results regarding learning strategy selection will be reported first.

Learning Strategy Selection at Trial 3

Hypothesis 2B. The frequency of selecting the effective learning strategy on Trial 3 was predicted to be higher for the multi trial conditions than the single trial conditions, overall. In addition, the frequency of choosing the effective learning strategy was predicted to be higher for conditions in which the effective learning strategy was experienced, with the highest frequency being in the Multi Trial Same Effective-Effective condition. Amongst the multi trial conditions, the lowest frequency of selecting the effective learning strategy on Trial 3 was predicted to be found in the Multi Trial Same Ineffective-Ineffective condition.

To test for Hypothesis 2B, the frequency of learning strategy selected at Trial 3 for each experimental conditions were first tabulated. A graph depicting the frequency of learning strategy selection is presented in Figure 28. Multiple Chi-square tests of independence were performed to examine the relation between different conditions and participant's learning strategy selection at Trial 3.

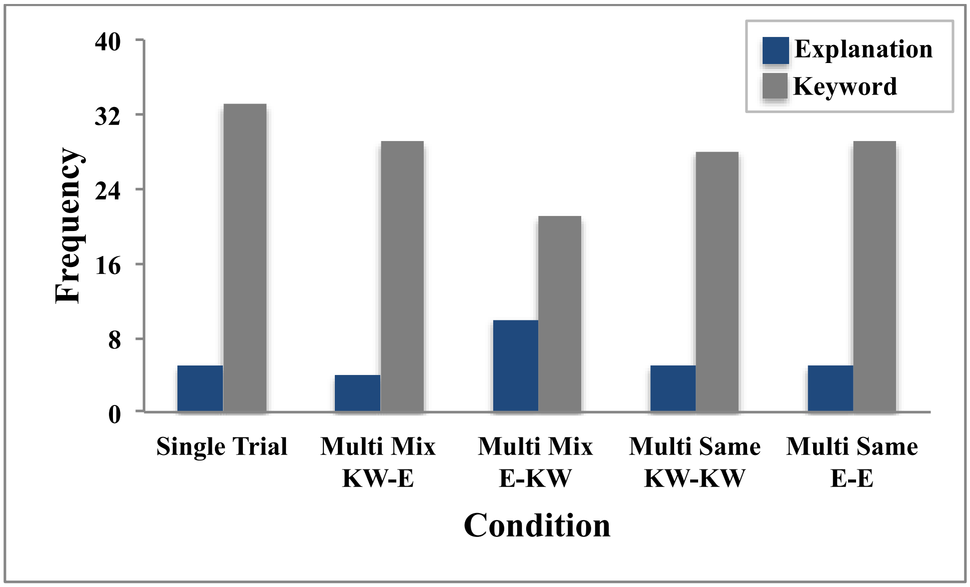

Figure 28. Frequency of learning strategy selection at Trial 3 in Experiment 2

Frequency of learning strategy selection at Trial 3 for Hypothesis 2B

To examine the relation between Trial Type and learning strategy selection at Trial 3, counts from all multi trial conditions were combined and compared against the single trial using a Chi-square test of independence. The difference between Trial Type was not significant, X^2 (1, N = 169) = .55, $p > .05$. The pattern of learning strategy selection between multi and single conditions was not significant. Participants in the multi trial conditions were equally as unlikely to select the explanation learning strategy as participants in the single condition. The single trial acted as a baseline of strategy selection and showed that participants selected the keyword learning strategy six times more often than the explanation learning strategy. The count for this Chi-square comparison is presented in Table 16.

Table 16

Frequency of learning strategy selection by Trial Type in Experiment 2

	Keyword	Explanation
Single Trial	33	5
Multi Trials	107	24

To examine the relation between experience of learning strategies at Trials 1 and 2 and learning strategy selection at Trial 3, counts for each multi trial conditions were compared using a Chi-square test of independence. The difference between the multi trial conditions was not significant, X^2 (3, N = 131) = 5.39, $p > .05$. There was no significant difference in the pattern of learning strategy selection between all the multi conditions. Participants in the multi conditions were all more likely to select the keyword learning strategy than they were the explanation learning strategy. The count for this Chi-square comparison is presented in Table 17.

Therefore, Hypothesis 2B was not supported.

Table 17

Frequency of learning strategy selection amongst the multi conditions in Experiment 2

	Keyword	Explanation
Multi Mix KW-E	29	4
Multi Mix E-KW	21	10
Multi Same KW-KW	28	5
Multi Same E-E	29	5

Although not explicitly predicted, three additional Chi-square tests of independence were conducted to compare between different combinations of multi trial conditions. The first comparison was between the conditions that experienced the effective learning strategy to those that experienced only the ineffective learning strategy. The second comparison was between the Multi Mix conditions and the final comparison was between the Multi Same conditions.

To examine the relation between the presence of experience of the effective learning strategy at Trials 1 and 2 and learning strategy selection at Trial 3, combined counts obtained from Table 17 for the Multi Mix KW-E, Multi Mix E-KW, and Multi Same E-E (*M* for keyword = 79; *M* for explanation = 19) were compared against counts for the Multi Same KW-KW (*M* for keyword = 28; *M* for explanation = 5) condition using a Chi-square test of independence. The difference between the conditions where participants experienced the effective learning strategy and the condition where participants only experienced the ineffective learning strategy was not significant, X^2 (1, N = 131) = 2.96, $p > .05$. Participants were equally likely to select the explanation learning strategy regardless of whether they had previously experienced that effective learning strategy.

To examine the relation between the order of learning strategies at Trials 1 and 2 and learning strategy selection at Trial 3, counts for the Multi Mix KW-E and Multi Mix E-KW were compared using a Chi-square test of independence. The difference between the Multi Mix E-KW and Multi Mix KW-E was marginally significant at $p = .05$, X^2 (1, N = 64) = 3.79. Participants in the Multi Mix E-KW condition were numerically more likely to select the explanation learning strategy than participants in the Multi Mix KW-E condition. The count for this Chi-square comparison is presented in Table 18.

Table 18

Frequency of learning strategy selection between multi mix conditions in Experiment 2

	Keyword	Explanation
Multi Mix KW-E	29	4
Multi Mix E-KW	21	10

Finally, to examine the relation between the learning strategies assigned at Trials 1 and 2 and learning strategy selection at Trial 3, counts for the Multi Same KW-KW and Multi Same E-E were compared using a Chi-square test of independence. The difference between the Multi Same conditions was not significant, X^2 (1, N = 67) = .00, $p > .05$. The pattern of learning strategy selection between both Multi Same conditions was almost identical. Participants in the Multi Same conditions were equally unlikely to select the keyword learning strategy than they were the explanation learning strategy. The count for this Chi-square comparison is presented in Table 19.

Table 19

Frequency of learning strategy selection between multi same conditions in in Experiment 2

	Keyword	Explanation
Multi Same KW-KW	28	5
Multi Same E-E	29	5

In summary, the proportion of participants who selected the keyword learning strategy was significantly higher than those who selected the delayed explanation strategy; this was the case across all conditions. However, when comparing just between the Multi Mix conditions, the Multi Mix E-KW condition produced twice more participants who selected the explanation learning strategy than the Multi Mix KW-E condition.

Hypotheses Testing for Comprehension Accuracy

Comprehension accuracy. To test for Hypothesis 2C, both Trial Type and Learning Strategy were the main predictor for comprehension accuracy in the GLMM analyses. In addition to Trial 3, comprehension accuracy at Trial 1 was also predicted to measure the baseline of the effectiveness of the learning strategies.

Hypothesis 2C. Comprehension accuracy was predicted to be higher when the effective learning strategy was selected than when the ineffective learning strategy was selected; a main effect of Learning Strategy on comprehension accuracy at Trial 3 was predicted. Comprehension accuracy was also predicted to be higher for the effective than for the ineffective learning strategy at Trial 1 when the learning strategy was assigned by the researcher.

GLMM logistic regression results for effects of Learning Strategy at Trials 3 and 1[4] are presented in Table 20. Hypothesis 2C predicted a main effect of Learning Strategy at Trials 3 and 1, a similar comprehension pattern for both trials. The main effect of Learning Strategy was not significant at either trial. At Trial 3, participants who selected explanation (M = .70, SE = .03) had similar comprehension accuracy as participants who selected the keyword (M = .74, SE = .01) learning strategy. Note, however, that only 15% of the participants chose explanation on Trial 3. At Trial 1, comprehension accuracy for

[4] Only multiple trial conditions were included in this analysis because the questions for the single conditions were different than those for the first trial of the multiple trial conditions.

participants assigned to explanation (*M* = .69, *SE* = .02) was similar to participants assigned to the keyword (*M* = .75, *SE* = .02) learning strategy. Comprehension mean scores for comparisons at both trials are depicted in Figure 29 below.

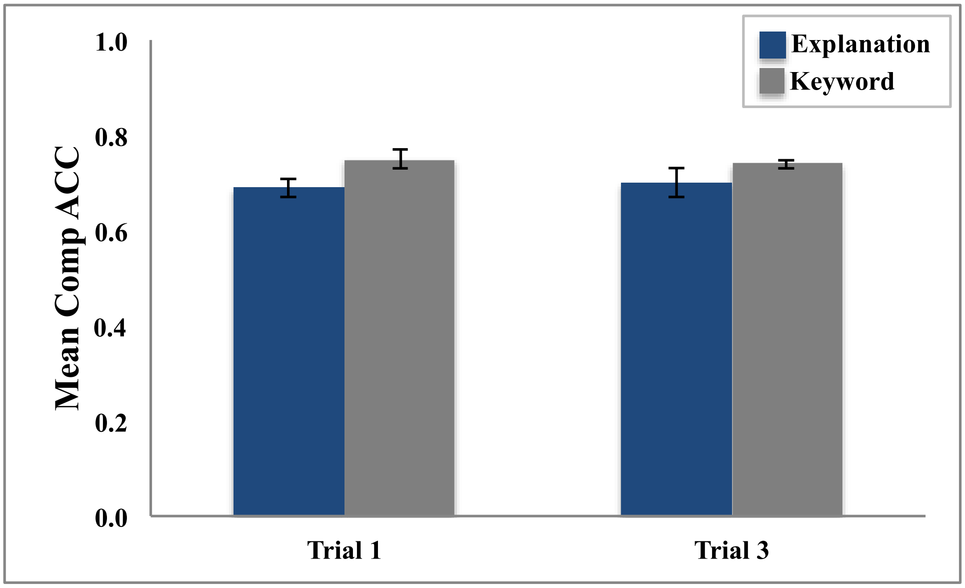

Figure 29. Comprehension accuracy for Learning Strategy at Trial 1 and Trial 3 in Experiment 2

Comprehension accuracy results for Hypothesis 2C. Error bars represent standard errors.

Table 20

GLMM fixed effects results for planned comparisons for Hypothesis 2C in Experiment 2 (DV: comprehension accuracy)

Hypothesis 2C	Predictors	Estimate	*SE*	*z*-Wald	*p*
Main Effect of Learning Strategy at Trial 3					
E = KW	Intercept	1.35	.48	2.79	< .01
	TT[2]xLS[3]	.22	.31	.69	.49
Main Effect of Learning Strategy at Trial 1					
E = KW	Intercept	.92	.20	4.70	< .01
	TT[2]xLS[3]	- .36	.21	1.77	.08
Main Effect of Trial Type at Trial 3					
Multi > Single	Intercept	1.75	.43	4.09	< .01
	TT[2]	- .92	.28	-3.27	$< .01^{**}$
Interaction Effect of Trial Type and Learning Strategy at Trial 3					
Multi-E as comparison group					
Multi-E = Multi-KW	Intercept	1.65	.52	3.19	< .01
	TT[2]xLS[3]	.13	.36	.35	.73
Multi-E > Single-E	TT[2]xLS[3]	-1.50	.71	-2.12	$< .05^{*}$
Multi-E = Single-KW	TT[2]xLS[3]	- .70	.41	-1.72	.09
Multi-KW as comparisons group					
Multi KW > Single E	Intercept	1.77	.43	4.11	< .01
	TT[2]xLS[3]	-1.63	.67	-2.42	$< .05^{*}$
Multi-KW = Single-KW	TT[2]xLS[3]	- .82	.31	-2.63	$< .01^{**}$
Single-KW as comparison group					
Single-KW = Single-E	Intercept	.95	.44	2.17	< .05
	TT[2]xLS[3]	- .81	.68	-1.18	.24
Main Effect of Learning Strategy Order at Trial 3					
Mixed = Same	Intercept	1.74	.53	3.30	< .01
	LSO[6]	.13	.33	.40	.73

[2] Trial Type; [3] Learning Strategy; [6]Learning Strategy Order
$^{*}p < .05$; $^{**}p < .01$ for fixed effects of conditions
Note. Full GLMM output for these analyses can be located in Appendix E.

Additional comprehension accuracy analyses. The main effect of Trial Type on comprehension accuracy at Trial 3, also shown in Table 21, was computed to aid in the interpretation of the metacomprehension accuracy hypothesis presented later. The main effect of Trial Type was significant, consistent with the findings from Experiment 1. Comprehension accuracy for participants in the multi trial conditions (M = .76, SE = .02) was significantly better than for those in the single conditions (M = .64, SE = .03). Additionally, Trial Type interacted significantly with Learning Strategy; comprehension accuracy was significantly lower for participants who selected to the explanation learning strategy in the single condition as compared to all other conditions. All other planned comparisons were not significant. Although not statistically significant, comprehension accuracy was numerically highest for participants in the Multi Keyword condition (M = .74, SE = .01), second highest for participants in the Multi Explanation condition (M = .74, SE = .01), followed by participants in the Single Keyword condition (M = .66, SE = .02). Participants in these three conditions scored significantly higher than participants in the Single Explanation condition (M = .52, SE = .06). Mean comparison scores for all conditions are depicted in Figure 30 below.

In Experiment 2, there were multiple trials for which the learning strategies alternated between Trials 1 and 2 and multiple trials for which the same learning strategy was assigned to both trials. Another analysis was conducted just for multiple trials and collapsed across Learning Strategy using Learning Strategy Order as the factor. The effect of mixed versus same learning strategy on comprehension accuracy at Trial 3 was computed. The main effect of Learning Strategy Order was not significant; performance for participants who experienced both learning strategies (M = .74, SE = .01) was similar to participants who experienced only one of the learning strategies (M = .76, SE = .01).

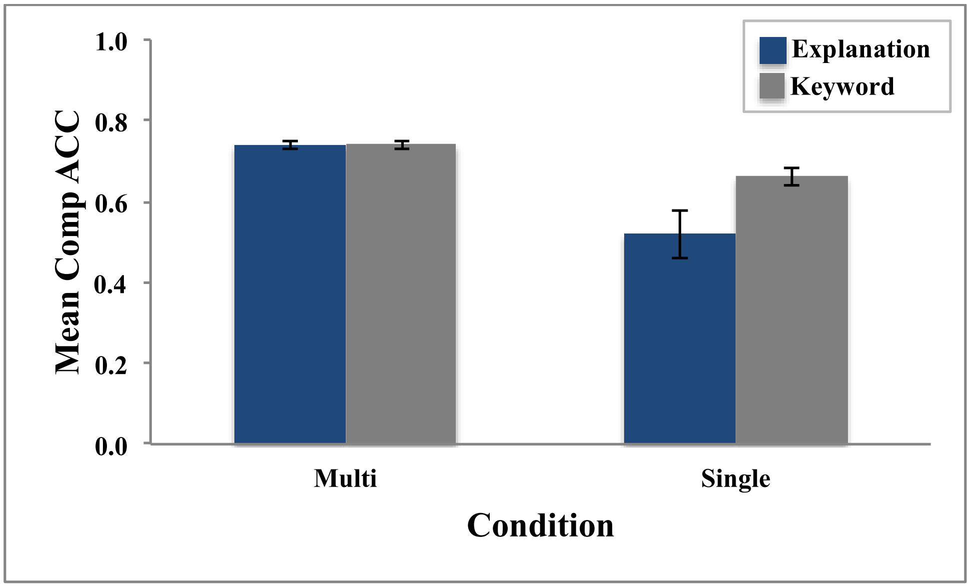

Figure 30. Comprehension accuracy for Trial Type and Learning Strategy at Trial 3 in Experiment 2

Comprehension accuracy results for Hypothesis 2A interpretation. Error bars represent standard errors.

Delayed judgments of learning (DJOL) sensitivity. The magnitude of DJOL predictions was measured using the LMEM approach. The main effect of Learning Strategy on the magnitude of DJOLs was computed for both Trials 3 and 1. Results for the all the LMEM regressions conducted for DJOL sensitivity are presented in Table 21. At Trial 3, there was no main effect of Learning Strategy on DJOL sensitivity. DJOLs tracked the comprehension accuracy findings; DJOLs were similar for the explanation (*M* = 73.64, *SE* = .93) and keyword learning strategies (*M* = 76.58, *SE* = 1.97). There was also not a main effect of Learning Strategy on DJOL sensitivity at Trial 1; DJOLs tracked the comprehension accuracy findings for Trial 1 as well. Mean DJOL sensitivity was similar when the explanation (*M* = 66.93, *SE* = 1.22) and the keyword learning strategy

(*M* = 71.10, *SE* = 1.94) were assigned. Mean DJOL sensitivity for each comparison is depicted in the graph below (Figure 31).

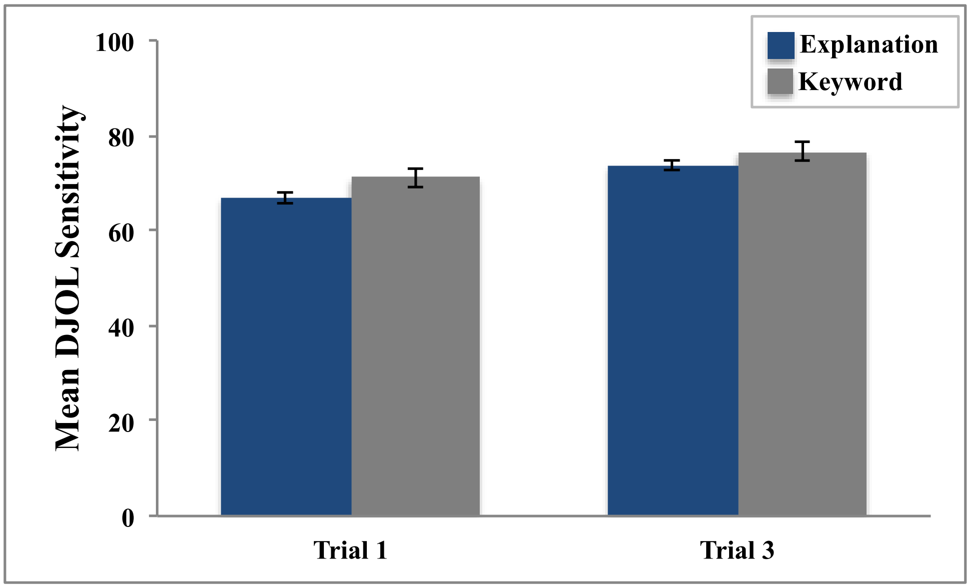

Figure 31. Delayed judgments of learning for Learning Strategy at Trials 3 and 1 in Experiment 2

Mean DJOL sensitivity results for Hypothesis 2C. Error bars represent standard errors.

To supplement the additional comprehension accuracy main effect of Trial Type results presented earlier, the main effect of Trial Type on DJOL sensitivity in Trial 3 was also computed. There main effect of Trial Type was significant, consistent with the findings from Experiment 1 and comprehension accuracy in this experiment. DJOLs were higher for multiple trials (*M* = 79.86, *SE* = .93) than for single trials (*M* = 63.05, *SE* = 1.97). DJOLs sensitivity did not significantly interact between Trial Type and Learning Strategy. Mean comparison scores for all conditions are depicted in Figure 32 below.

The effect of mixed versus same learning strategy on DJOL sensitivity at Trial 3 was also computed. The main effect of Learning Strategy Order was not significant;

DJOLs made by participants who experienced both learning strategies (*M* = 81.01, *SE* = 2.33) were similar to participants who experienced only one of the learning strategy (*M* = 78.76, *SE* = 3.00). This finding suggests that regardless of whether participants experienced both learning strategies or only one across the first two trials, DJOL sensitivity did not vary significantly.

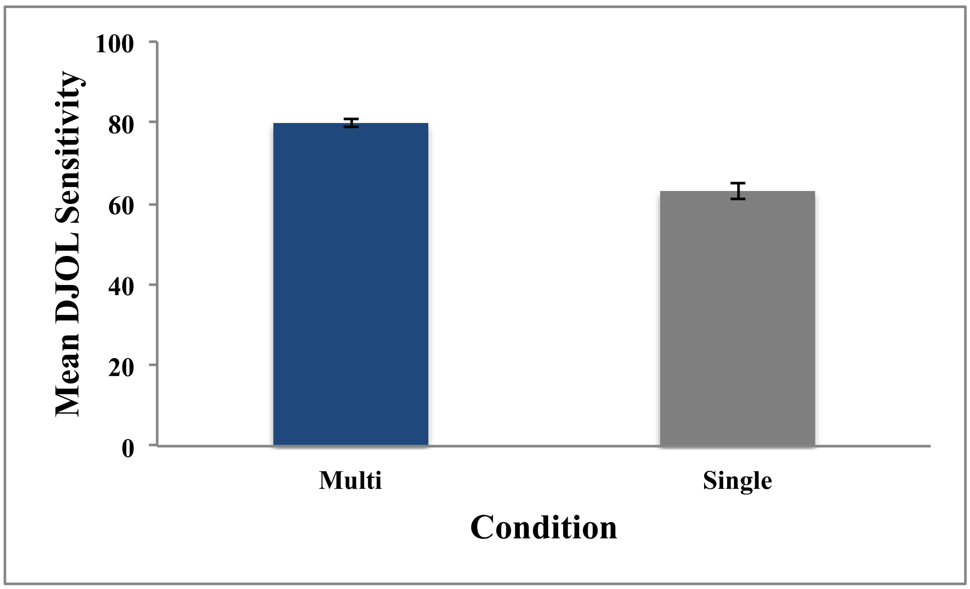

Figure 32. Delayed judgments of learning for Trial Type at Trial 3 in Experiment 2

Mean DJOLs sensitivity results for Hypothesis 2A interpretation. Error bars represent standard errors.

Table 21

LMEM fixed effects results for planned comparisons for DJOL Sensitivity for Trial 3 and Trial 1 in Experiment 2 (DV: DJOL)

	Predictors	Estimate	*SE*	*df*	*t*	*p*
Main Effect of Learning Strategy at Trial 3						
E = KW	Intercept	73.65	3.64	105	20.25	< .01
	LS³	- 1.32	4.49	121	.76	.45
Main Effect of Learning Strategy at Trial 1						
E = KW	Intercept	66.93	3.53	23	18.94	< .01
	LS³	4.17	3.24	129	1.29	.20
Main Effect of Trial Type @ Trial 3						
Multi > Single	Intercept	79.86	2.03	32	39.26	< .01**
	TT²	-16.81	3.17	167	-5.30	< .01**
Interaction Effect of Trial Type and Learning Strategy on Trial 3						
Multi-E as comparison group						
Multi-E = Multi-KW	Intercept	77.06	3.74	105	20.58	< .01
	TT²xLS³	3.42	4.05	123	.85	.40
Multi-E > Single-E	TT²xLS³	-19.82	9.05	96	-2.19	< .05*
Multi-E > Single-KW	TT²xLS³	-13.14	4.85	114	-2.71	< .01**
Multi-KW as comparisons group						
Multi-KW > Single-E	Intercept	80.49	2.45	27	34.29	< .01
	TT²xLS³	-23.24	8.78	72	-2.65	< .01**
Multi-KW > Single-KW	TT²xLS³	-16.56	4.21	44	-3.94	< .01**
Single-KW as comparison group						
Single-KW = Single-E	Intercept	63.93	3.71	43	17.26	< .01
	TT²xLS³	-6.69	9.04	80	- .74	.46
Main Effect of Learning Strategy Order at Trial 3						
Mixed = Same	Intercept	81.01	2.33	65	34.72	< .01
	LSO⁶	-2.25	3.00	105	- .75	.46

[2] Trial Type; [3] Learning Strategy; [6]Learning Strategy Order
* $p < .05$; ** $p < .01$ for fixed effects of conditions
Note. Full LMEM output for these analyses can be located in Appendix E.

Hypotheses Testing of Metacomprehension Accuracy

To measure metacomprehension accuracy, comprehension accuracy was entered as the dependent variable in the GLMM analyses; centralized DJOLs were the predictor variable. Additionally, the effects of Trial Type and Learning Strategy were also included as predictor variables with metacomprehension accuracy as the dependent variable. Change in metacomprehension accuracy will be explained by the degree to which one condition varies as compared to the comparison condition. Therefore, when reporting the mean and standard error values for metacomprehension accuracy for these conditions, the estimate value for the comparison group will also be reported to provide relative comparison. Using the estimates generated from the logistic regression output, line graphs were created to represent the differences in slope change relative to the dependent variable amongst the comparison conditions.

Hypothesis 2A. Metacomprehension accuracy was predicted to be higher for all of the multi trial conditions as compared to the single trial condition, regardless of learning strategy.

To test for Hypothesis 2A, metacomprehension accuracy was compared between multiple and single trials testing the main effect of Trial Type. The multi trial conditions served as the comparison group. Results for the GLMM logistic regression is presented in Table 23 and depicted in Figure 33. There was no significant difference on metacomprehension accuracy between multi and single trials; metacomprehension accuracy slopes were similar for the multi ($M_{\text{Est.}}$ = .44, SE = .10) and single conditions (M_{change} = -.09, SE = .18). Therefore, Hypothesis 2A was not supported. This finding was consistent with Experiment 1, but not with the comprehension accuracy findings from this experiment. Metacomprehension and comprehension accuracy were dissociated in terms of Trial Type.

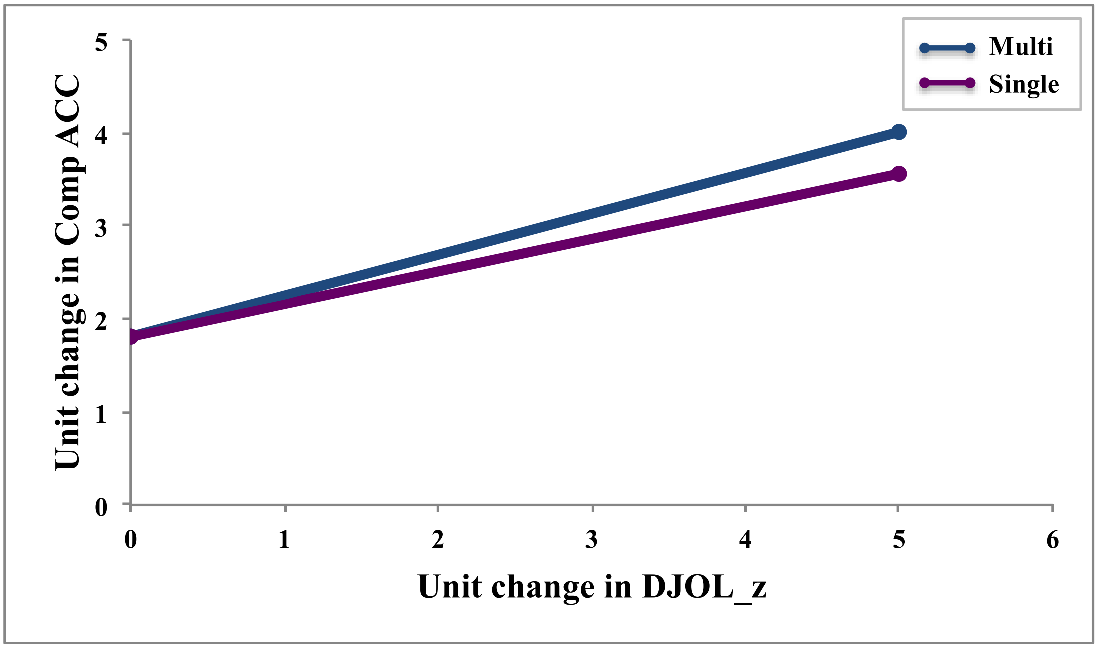

Figure 33. Metacomprehension accuracy for Trial Type at Trial 3 in Experiment 2

Logistic regression comparison for DJOL accuracy between multi and single trial conditions with multi trial condition serving as comparison group for Hypothesis 2A.

Although not included in the predictions, the main effect of Learning Strategy on metacomprehension accuracy was also computed. Results for the additional GLMM logistic regressions are also presented in Table 22 and depicted in Figure 34. There was no significant difference in metacomprehension accuracy between participants who selected the explanation ($M_{Est.}$ = .52, *SE* = .20) and participants who selected the keyword (M_{change} = -.12, *SE* = .24) learning strategy. The interaction between Learning Strategy and Trial Type in metacomprehension accuracy was not significant. This finding did not replicate Experiment 1; in Experiment 1, when explanation was assigned as the learning strategy for Trial 3, metacomprehension accuracy was better than when keyword was assigned as the learning strategy for Trial 3. The metacomprehension accuracy findings paralleled those for comprehension accuracy with regard to Learning Strategy.

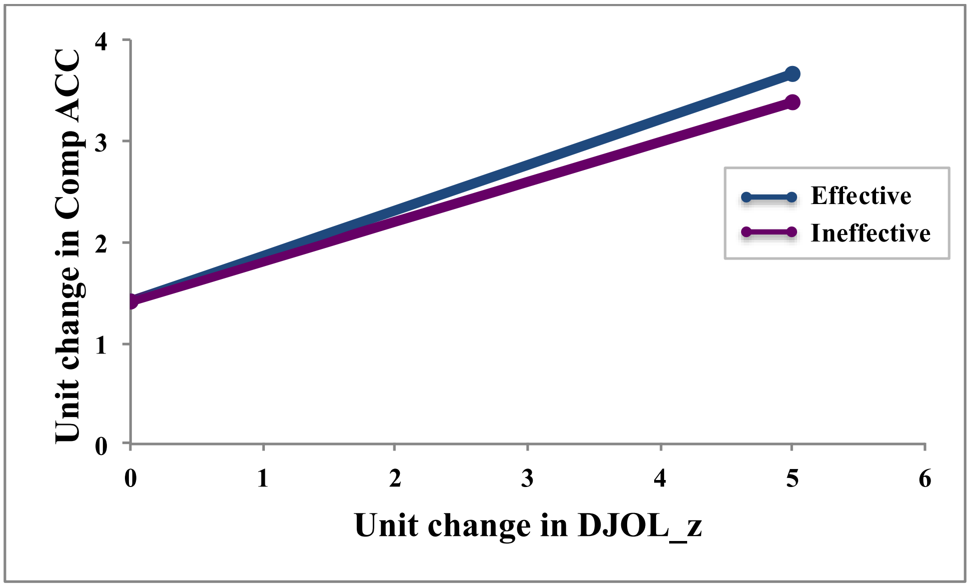

Figure 34. Metacomprehension accuracy for Learning Strategy at Trial 3 in Experiment 2

Logistic regression comparison for DJOL accuracy between effective and ineffective learning strategy conditions with effective learning strategy serving as comparison group.

The main effect of Learning Strategy Order on metacomprehension accuracy was also computed just for the multiple trials. The mixed condition served as the comparison group. Metacomprehension accuracy did not differ between participants in the mixed condition ($M_{Est.}$ = .47, *SE* = .14) and participants in the same condition (M_{change} = -.06, *SE* = .18). This finding shows that, regardless of whether the participant experienced both learning strategies or only one during the first two trials, metacomprehension accuracy did not differ significantly at Trial 3.

Table 22

GLMM fixed effects results for planned comparisons for metacomprehension accuracy for Hypothesis 2A in Experiment 2 (DV: comprehension accuracy)

Hypothesis 2	Predictors	Estimate	*SE*	*z*-Wald	*p*
H2A: Multi = Single	$DJOL_{Est.}$	.44	.10	4.23	< .01
	DJOL[1] xTT[2]	- .09	.18	- .50	.61
Main Effect of Learning Strategy					
E = KW	$DJOL_{Est.}$	.52	.20	2.63	< .01
	DJOL[1]xLS[3]	- .12	.24	- .52	.60
Interaction Effect of Trial Type and Learning Strategy at Trial 3					
Multi-E as comparison group					
Multi-E = Multi-KW	$DJOL_{Est.}$	.53	.27	1.98	< .05
	TT[2]xLS[3]	- .08	.31	- .24	.81
Multi-E = Single-E	TT[2]xLS[3]	.60	.97	.61	.54
Multi-E = Single-KW	TT[2]xLS[3]	- .28	.38	- .74	.46
Multi-KW as comparisons group					
Multi KW = Single E	$DJOL_{Est.}$	.46	.13	3.52	< .01
	TT[2]xLS[3]	.67	.87	.77	.44
Multi-KW = Single-KW	TT[2]xLS[3]	- .20	.20	-1.01	.31
Main Effect of Learning Strategy Order at Trial 3					
Mixed = Same	$DJOL_{Est.}$	.47	.14	3.28	< .01
	LSO[6]	- .06	.18	- .34	.73

[2] Trial Type; [3] Learning Strategy; [6]Learning Strategy Order
Note. Full GLMM output for these analyses can be located in Appendix E.

Summary of Hypotheses Testing Results

There was no difference in the frequency of learning strategy selection between Trial Types. The same proportion of participants selected the explanation and keyword learning strategies in both the multiple and single trial conditions. Furthermore, amongst all the multi trial conditions, most participants chose the keyword over the explanation learning strategy. Out of all the multi trials, participants in the Multi Mix E-KW condition were twice as likely to select the explanation learning strategy over the keyword strategy, *p* = .05.

For comprehension accuracy, there was no effect of Learning Strategy at either Trial 3 or 1. For both trials, comprehension was the same for the explanation and keyword strategies. DJOLs magnitude tracked this pattern of comprehension accuracy; there was no effect of Learning Strategy at either trial. Consistent with Experiment 1, metacomprehension accuracy was not higher for multiple than for single trials. However, contrary to the findings of Experiment 1, metacomprehension accuracy did not vary with learning strategy. Participants who selected the explanation learning strategy did not have better metacomprehension accuracy than those participants who selected the keyword learning strategy.

Confidence Judgments

Although specific hypotheses were not generated for CJs, the sensitivity and accuracy of CJs are reported here.

Confidence judgment sensitivity. The magnitude of CJ predictions was analyzed using the LMEM approach. The effects of Trial Type and Learning Strategy on the magnitude of CJ predictions were computed. Results for the LMEM regression are presented in Table 23. Participants who had multiple trials (*M* = 85.62, *SE* = .84) gave significantly higher CJs than participants who had a single trial (*M* = 71.67, *SE* = 1.78). The main effect of Learning Strategy was not significant. Participants who selected the explanation learning strategy (*M* = 80.62, *SE* = 1.83) gave similar CJs as participants who selected the keyword learning strategy (*M* = 82.88, *SE* = .73). Mean CJ sensitivity for each condition is depicted in the graph below (see Figure 35).

Table 23

LMEM fixed effects results for planned comparisons for CJ sensitivity for Trial 3 in Experiment 2 (DV: CJ)

	Predictors	Estimate	*SE*	*df*	*t*	*p*
Main Effect of Trial Type and Learning Strategy						
Multi > Single	Intercept	85.63	2.01	24	42.70	< .01**
	TT[2]	-13.96	2.85	167	-4.89	< .01**
E = KW	Intercept	80.62	3.74	48	21.55	< .01
	LS[3]	2.26	3.47	140	.66	.52

[2] Trial Type; [3] Learning Strategy
** $p < .01$ for fixed effects of conditions
Note. Full LMEM output from these analyses can be located in Appendix E.

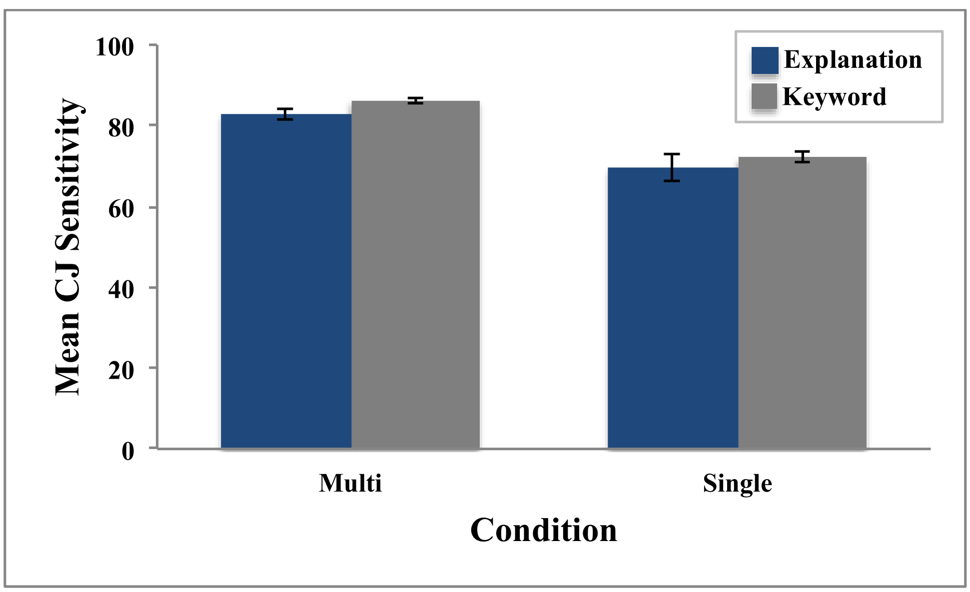

Figure 35. Confidence judgments magnitude for all conditions at Trial 3 in Experiment 2

Mean CJ sensitivity for all conditions at Trial 3. Error bars represent standard errors.

Confidence judgment accuracy. The accuracy of CJs was calculated using the GLMM approach. Centralized CJs served as predictors for comprehension accuracy to determine CJ accuracy. The effects of Trial Type and Learning Strategy on CJ accuracy were analyzed. Results for the GLMM logistic regression are presented in Table 24 and depicted in Figure 36 and 37. There were no significant main or interaction effects.

Table 24

GLMM fixed effects results for planned comparisons for CJ accuracy for Trial 3 in Experiment 2 (DV: comprehension accuracy)

	Predictors	Estimate	*SE*	*z*-Wald	*p*
Main Effect of Trial Type and Learning Strategy					
Multi = Single	$CJ_{Est.}$	.90	.17	5.15	< .01
	CJ[1] xTT[2]	- .22	.17	-1.30	.19
E = KW	CJ_{Est}	.61	.23	2.61	< .01
	CJ[1]xLS[3]	.29	.20	1.46	.14

[4] CJ z-scores; [2] Trial Type; [3] Learning Strategy
[*] $p < .05$ for fixed effects of conditions
Note. Full GLMM output for these analyses can be located in Appendix E.

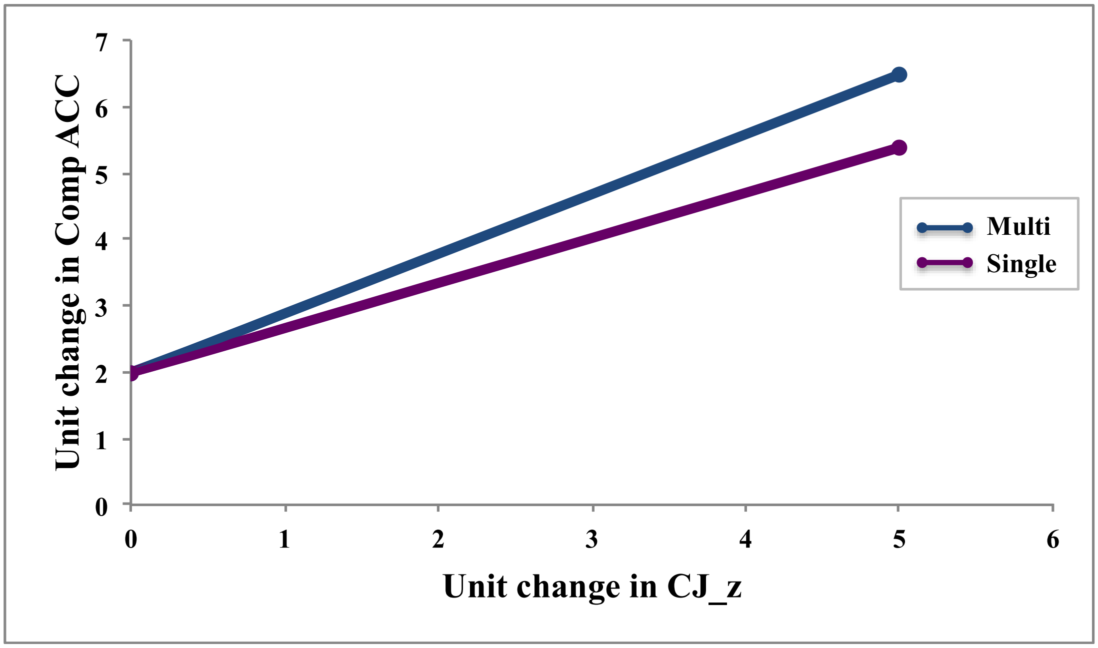

Figure 36. CJ accuracy for Trial Type at Trial 3 in Experiment 2

Logistic regression comparison for CJ accuracy between multi and single trial conditions with multi conditions serving as comparison group.

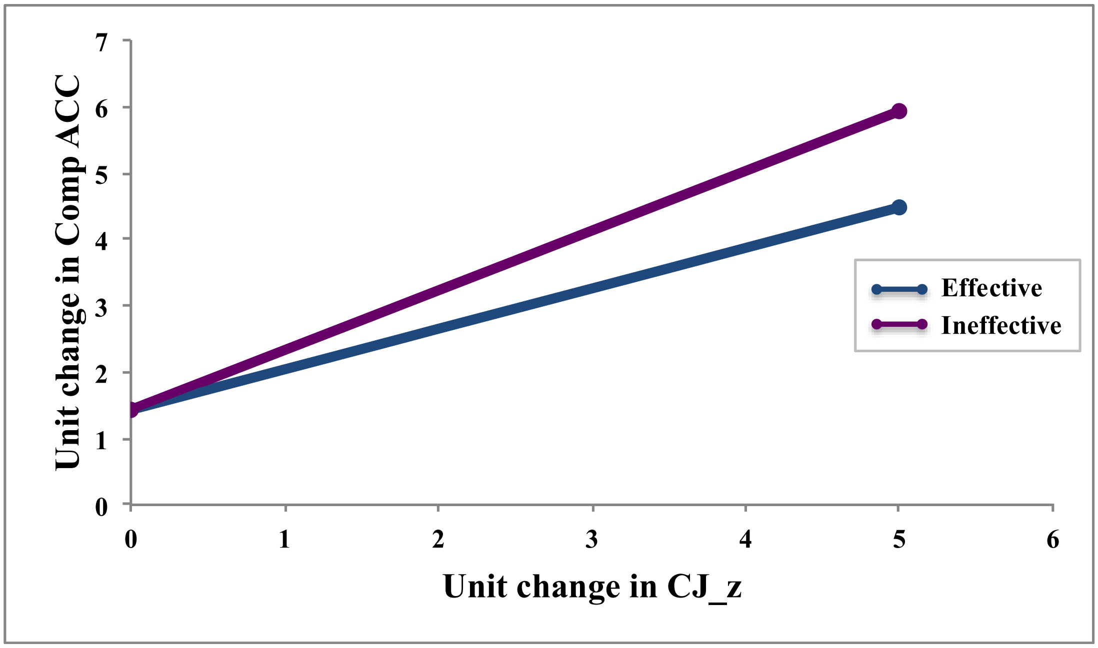

Figure 37. CJ accuracy for Learning Strategy at Trial 3 in Experiment 2

Logistic regression comparison for CJ accuracy between effective and ineffective learning strategy conditions with effective learning strategy serving as comparison group.

Additional Analysis for Multi Trial Conditions Only

The main aim of this experiment was to investigate the impact on the control process on comprehension accuracy as informed by metacomprehension accuracy that presumably improved across multiple trials and with the effective learning strategy. Additional analyses were conducted just on the multiple trials to examine changes in both comprehension accuracy—as a test of the impact of control processes—and metacomprehension accuracy—as a test of the impact of monitoring processes.

Comprehension accuracy. The main effect of Trial on comprehension performance was measured. GLMM logistic regression results for these comparisons are presented in Table 30 and depicted in Figure 38. There was a main effect of Trial on comprehension accuracy; participants performed the best at Trial 3 (M = .76, SE = .02) as compared to Trial 2 (*M* = .73, *SE* = .02) and Trial 1 (*M* = .72, *SE* = .02) regardless of the condition they were assigned at Trials 1 and 2 or the learning strategy they selected at Trial 3.

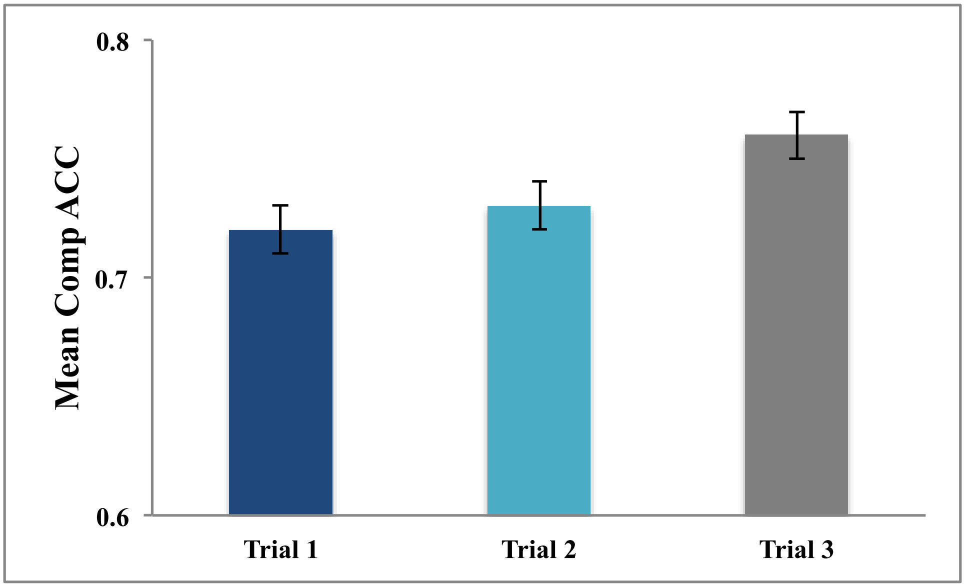

Figure 38. Comprehension accuracy across Trials in Experiment 2

Mean comprehension accuracy across trials within multi trial conditions. Error bars represent standard errors. Note that the y-axis is zoomed in to highlight the differences.

The two mixed conditions—Multi Mix E-KW and Multi Mix KW-E— were combined to form a "Both" condition. The change in comprehension accuracy was examined across trials to determine the effect of experiencing both learning strategies on comprehension accuracy. The benefit of multiple trials on comprehension when both learning strategies were experienced was only observed at Trial 3. The same analysis was done for each of the same learning strategy conditions, independently. When the keyword learning strategy was repeated in Trials 1 and 2, the only significant increase in comprehension was observed at Trial 3. When the explanation learning strategy was repeated in Trials 1 and 2, a significant increase in comprehension was observed at Trials 2 and 3. These analyses were not conditionalized in terms of what learning strategy participants selected at Trial 3 because most participants selected the keyword learning strategy. The GLMM logistic regression results for these comparisons are presented in Table 25.

Table 25

GLMM fixed effects results for planned comparisons for comprehension accuracy for multi trial conditions in Experiment 2 (DV: comprehension accuracy)

	Predictors	Estimate	*SE*	*z*-Wald	*p*
Main Effect of Trial: Trial 1 as comparison group					
Trial 1 = Trial 2	Intercept	1.12	.19	5.76	< .01
	T[5]	- .03	.10	- .30	.76
Trial 1 < Trial 3	T[5]	.32	.11	2.96	< .01**
Main Effect of Trial: Trial 2 as comparison group					
Trial 2 < Trial 3	Intercept	1.09	.18	5.95	< .01
	T[5]	.37	.11	3.12	< .01**
Interaction Effect of Trial and Learning Strategy: Both					
Trial 1 = Trial 2	Intercept	1.23	.24	5.12	< .01
	T[5]	- .23	.15	-1.53	.13
Trial 1 = Trial 3	T[5]	.08	.17	.47	.64
Trial 2 < Trial 3	Intercept	1.00	.22	4.63	< .01
	T[5]	.31	.16	1.98	< .05*
Interaction Effect of Trial and Learning Strategy: KW					
Trial 1 = Trial 2	Intercept	.97	.23	4.28	< .01
	T[5]	.32	.19	1.72	.08
Trial 1 < Trial 3	T[5]	.50	.21	2.36	< .05*
Trial 2 = Trial 3	Intercept	1.29	.22	5.82	< .01
	T[5]	- .32	.19	-1.72	.08
Interaction Effect of Trial and Learning Strategy: E					
Trial 1 = Trial 2	Intercept	1.10	.27	4.10	< .01
	T[5]	- .03	.22	- .14	.89
Trial 1 < Trial 3	T[5]	.66	.24	2.73	< .01**
Trial 2 < Trial 3	Intercept	1.07	.22	4.97	< .05
	T[5]	.69	.29	2.35	< .05*

[5] Trial
* $p < .05$; ** $p < .01$ for fixed effects of conditions
Note. Full GLMM output from these analyses can be located in Appendix E.

Metacomprehension accuracy. To test the impact of multiple trials on improving metacomprehension accuracy, the effect of Trial on metacomprehension accuracy was calculated. If the control process effects of learning strategy selection on comprehension accuracy that produced better comprehension accuracy on Trial 3 was informed by a meta level that was informed by accurate monitoring, metacomprehension accuracy should have improved from Trials 1 to 2. There main effect of Trial, however, was not significant; in fact, metacomprehension accuracy did not change across all three trials. No other comparisons produced significant results. Results from the GLMM logistic regression for the planned comparisons are presented in Table 26 and depicted in Figure 39.

Table 26

GLMM fixed effects results for planned comparisons for metacomprehension accuracy for multi trial conditions in Experiment 2 (DV: comprehension accuracy)

	Predictors	Estimate	*SE*	*z*-Wald	*p*
Main Effect of Trial: Trial 1 as comparison group					
Trial 1 = Trial 2	$DJOL_{Est.}$	.36	.09	4.15	< .01
	$DJOL^1$ xT^5	.01	.12	.12	.90
Trial 1 = Trial 3	$DJOL^1$ xT^5	.02	.12	.16	.87
Main Effect of Trial: Trial 2 as comparison group					
Trial 2 = Trial 3	$DJOL_{Est.}$	.37	.08	4.50	< .01
	$DJOL^1$ xT^5	.00	.11	.04	.97

[1] DJOL z-scores; [5] Trial

Note. Full GLMM output from these analyses can be located in Appendix E.

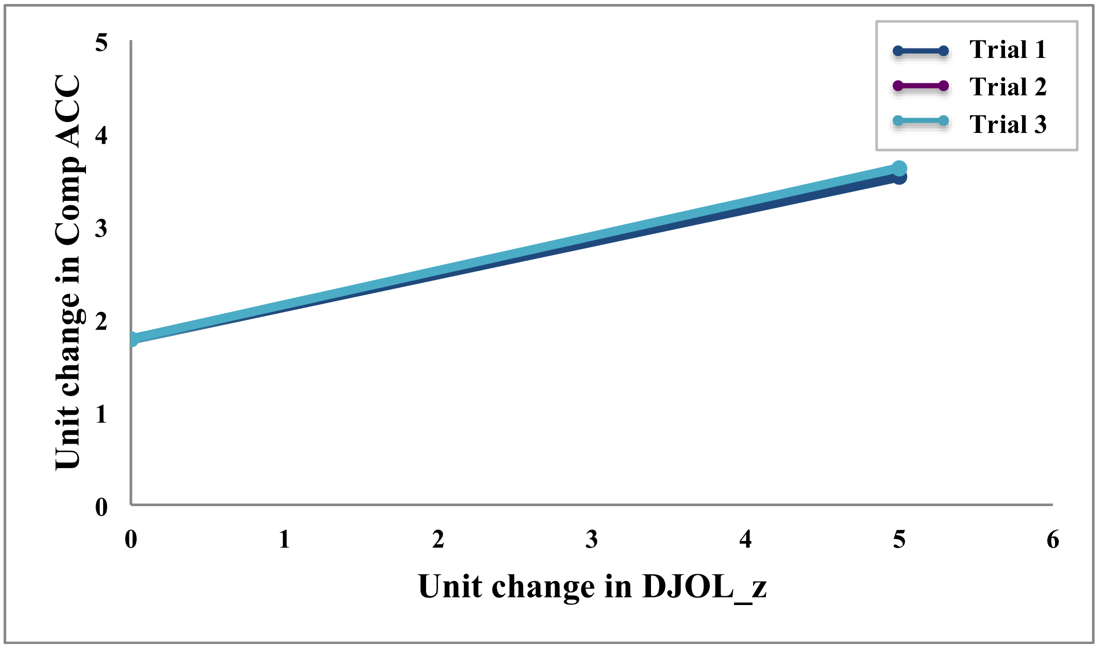

Figure 39. Metacomprehension accuracy across Trials in Experiment 2

Logistic regression comparison for DJOL accuracy across trials with Trial 1 serving as comparison group

Participant-level Metacomprehension Analysis

Although the main effect of Trial on metacomprehension accuracy across trials was not significant, the main effect of Trial on comprehension accuracy across trials was. Comprehension performance was highest at Trial 3, followed by Trial 2 and was the lowest at Trial 1. Comprehension accuracy did not vary significantly between Trials 1 and 2. Theoretically, this improvement in comprehension accuracy across trials could be attributable to the implementation of the appropriate learning strategy by the control process. Based on this theoretical viewpoint, the meta level assessment of the object level representation of the text, updated by improved monitoring, should have informed the control process such that the explanation learning strategy was selected. This selection of the explanation learning strategy should then have resulted in improved comprehension accuracy. However, this does not seem to be the case in Experiment 2. Although there was an aggregate improvement in comprehension accuracy across trials, it was not because participants selected the explanation learning strategy at Trial 3. In fact, the majority of the participants selected the keyword learning strategy at Trial 3.

Because the main aim of this experiment was to investigate the impact on control processes of improved metacomprehension accuracy, a participant-level analysis was conducted investigate whether participants who did show improvement in metacomprehension accuracy across trials, also showed an improvement in comprehension accuracy. Similar to the findings from Experiment 1, on closer examination of the participant-level data, some participants did show improved metacomprehension accuracy across trials; other participants did not. Employing the same methods from Experiment 1, using participant's coefficient values generated from the GLMM analysis that accounted for both random effects of participant and question, the relative metacomprehension accuracy change of slope was used to create change scores between each trial for each participant. Using this change value, a new post-hoc categorical variable was created—Metacomprehension Accuracy Improvement—for each two change phases (T1-T2 and T2-T3). Participants were categorized in either the "improved" or "unimproved" group based on the difference value for each change phase. Table 27 presents the descriptive statistics for the number of participants in the newly coded categories for each change phases.

Table 27

Descriptive statistics for newly formed category based on participant's change in metacomprehension (DJOL) accuracy between trials

N = 131	Trial 1 – Trial 2	Trial 2 – Trial 3
Improved	69	70
Unimproved	62	61

As presented in Table 27, there was almost an equal split of participants who did and did not improve across each change phase. Half of the participants showed improvement in metacomprehension accuracy from Trials 1 to 2. Almost the same number of participants improved—and did not improve—their metacomprehension accuracy from Trials 2 to 3. Aggregating the data across trials for the hypothesis testing analyses obscured these findings. Univariate ANOVAs were conducted to measure whether Metacomprehension Accuracy Improvement—improved versus unimproved—from Trials 1 to 2 impacted comprehension accuracy at Trial 2, and to measure whether Metacomprehension Accuracy Improvement from Trials 2 to 3 impacted comprehension accuracy at Trial 3.

Metacomprehension Accuracy Improvement from Trials 1 to 2 significantly impacted comprehension accuracy at Trial 2, F (1,127) = 8.36, $p < .01$, partial $\eta^2 = .06$. Pairwise comparisons showed that comprehension accuracy was higher at Trial 2 for participants with improved metacomprehension accuracy from Trials 1 to 2 compared to participants with unimproved metacomprehension accuracy from Trials 1 to 2, $p < .05$. Mean comparisons are presented in Table 28 and depicted in Figure 40.

Similar to Experiment 1, a Pearson product-moment correlation was conducted between the change in metacomprehension accuracy from Trials 1 to 2 and comprehension accuracy at Trial 2. There was a significant positive relationship between the two factors, $r(131) = .40$, $p < .01$. Although the correlation was not as strong as in Experiment 1, these findings provide strong support for the initial Univariate ANOVA analysis that comprehension accuracy was influenced by metacomprehension accuracy.

Table 28

Descriptive statistics for comprehension accuracy by metacomprehension accuracy improvement between Trials 1 and 2 and Trials 2 and 3 in Experiment 2

	MetaComp ACC	Mean Comp ACC
T1-T2	Improved	.77 (.01)
	Unimproved	.71 (.02)
T2-T3	E at T3 – Improved	.63 (.04)
	E at T3 – Unimproved	.88 (.04)
T2-T3	KW at T3 – Improved	.72 (.02)
	KW at T3 – Unimproved	.86 (.02)

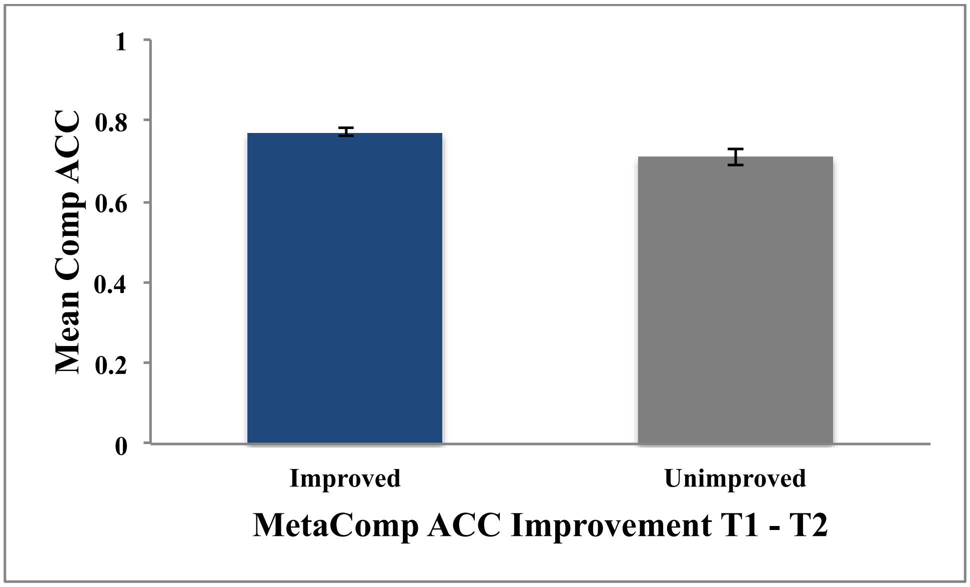

Figure 40. Metacomprehension accuracy improvement from Trials 1 to 2 on comprehension accuracy at Trial 2 in Experiment 2

Mean comprehension accuracy at Trial 2 for improved vs. unimproved metacomprehension accuracy from Trials 1 to 2. Error bars represents standard errors.

There was also a significant main effect of Metacomprehension Accuracy Improvement from Trials 2 to 3 on comprehension accuracy at Trial 3, F (1,127) = 39.63, $p < .01$, partial $\eta^2 = .24$. However, when taking into account the learning strategy selected at Trial 3, there was no significant interaction effect of Metacomprehension Accuracy Improvement and Learning Strategy selection on comprehension accuracy at Trial 3, F (1,127) = 3.20, $p > .05$, partial $\eta^2 = .03$. Although there was no significant interaction effect, the pattern that emerged for the effect of metacomprehension accuracy improvement for Trials 2 to 3 was the opposite from participants who showed metacomprehension accuracy improvement for Trials 1 to 2. The finding showed that participants who did not show improvement in metacomprehension accuracy for Trials 2 and 3 were the ones performing significantly better in comprehension accuracy at Trial 3 that participants who did show improvement in metacomprehension accuracy for Trials 2 and 3, $p < .05$. This finding was obtained regardless of learning strategy selection. Mean comparisons are presented in Table 28 and depicted in Figure 41.

The Pearson's correlation between the change in metacomprehension accuracy from Trials 2 to 3 and comprehension accuracy at Trial 3 was significant, but the correlation was negative, $r(131) = -70$, $p < .01$. This negative correlation mirrors the flipped pattern of comprehension accuracy shown in the ANOVA analysis and can be particularly attributed to the participants who showed improvement in metacomprehension accuracy from trials 2 to 3, but showed declined comprehension accuracy.

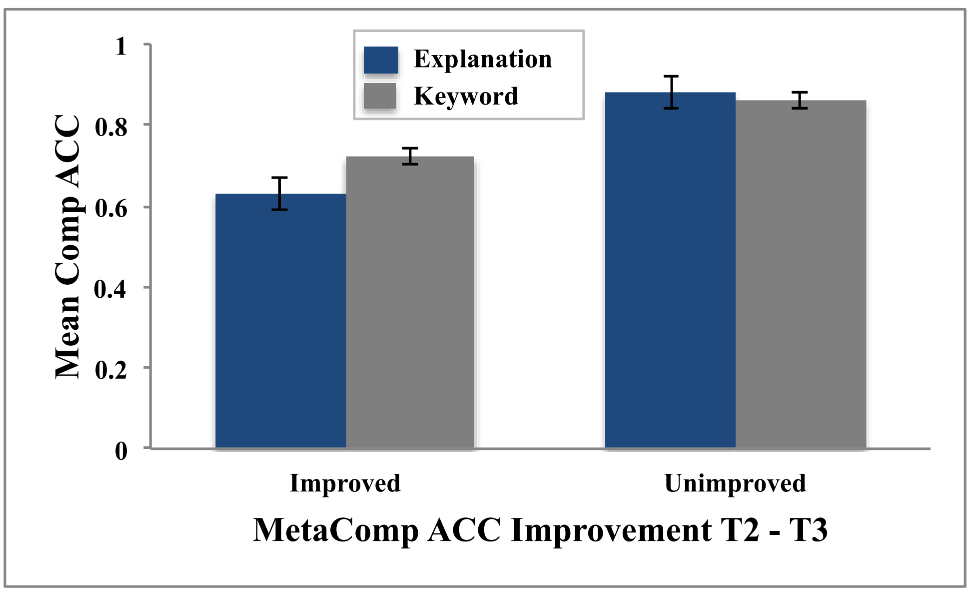

Figure 41. Metacomprehension accuracy improvement from Trials 2 to 3 on comprehension accuracy at Trial 3 in Experiment 2

Mean comprehension accuracy at Trial 3 for improved vs. unimproved metacomprehension accuracy from Trials 2 to 3. Error bars represents standard errors.

The finding of the effect of metacomprehension accuracy improvement from Trials 1 and 2 on comprehension accuracy is consistent with Experiment 1; comprehension accuracy was better when metacomprehension accuracy improved over the previous trial. Similar to Experiment 1, participants were also assigned learning strategies at Trials 1 and 2. However, at Trial 3, when participants were given the opportunity to select a learning strategy, apparently participants were not basing their learning strategy selection purely on improved metacomprehension accuracy. Only 24 participants selected the explanation learning strategy whereas 70 participants were found to have improved metacomprehension accuracy from Trials 2 to 3. Additionally, those 24 participants were a mixture of participants who both showed improvement and no improvement in metacomprehension accuracy from Trials 2 to 3.

This finding although surprising is in fact comparable to the finding from the pilot experiment. In the pilot experiment, participant's metacomprehension accuracy significantly plummeted at the final trial (Trial 4) after a significant improvement and peaked at Trial 3. However, despite the significant drop in metacomprehension accuracy at the final trial, comprehension accuracy at the final trial was the highest. This finding was attributed to the underconfidence with-practice (UWP) effect where participants compromised their ability to accurately assess their comprehension by switching to less diagnostic cues at later trials (Koriat, 1997). For example, in this case, participants could have switched to using mnemonic cues such as the text fluency and familiarity.

To confirm that the findings could in fact be attributed to the UWP effect, a Univariate ANOVA was conducted to measure whether improvement in metacomprehension accuracy at Trials 1 and 2 for comprehension test performance at

Trial 3. There was a significant main effect of metacomprehension accuracy improvement from Trials 1 to 2 on comprehension accuracy at Trial 3, F (1,127) = 8.27, $p < .01$, partial $\eta^2 = .06$. Pairwise comparisons showed that participants with improved metacomprehension from Trials 1 to 2 performed significantly better at Trials 3 compared to participants who did not show improvement in metacomprehension accuracy from Trials 1 to 2, $p < .05$. When taking into account the learning strategy selected at Trial 3, there was no significant interaction effect of Metacomprehension Accuracy Improvement and Learning Strategy selection on comprehension accuracy at Trial 3, F (1,127) = .75, $p > .05$, partial $\eta^2 = .01$. Mean comparisons are presented in Table 29 and depicted in Figure 42. The Pearson's correlation between the change in metacomprehension accuracy from Trials 1 to 2 and comprehension accuracy at Trial 3 found the two to be moderately, but significantly, correlated, $r(131) = .41$, $p < .01$. Comprehension accuracy was associated with the change in metacomprehension accuracy.

Table 29

Descriptive statistics for comprehension accuracy at Trial 3 by metacomprehension accuracy improvement between Trials 1 and 2 in Experiment 2

	MetaComp ACC	Mean Comp ACC
T1-T2	E at T3 – Improved	.82 (.04)
	E at T3 – Unimproved	.69 (.05)
T1-T2	KW at T3 – Improved	.82 (.02)
	KW at T3 – Unimproved	.75 (.02)

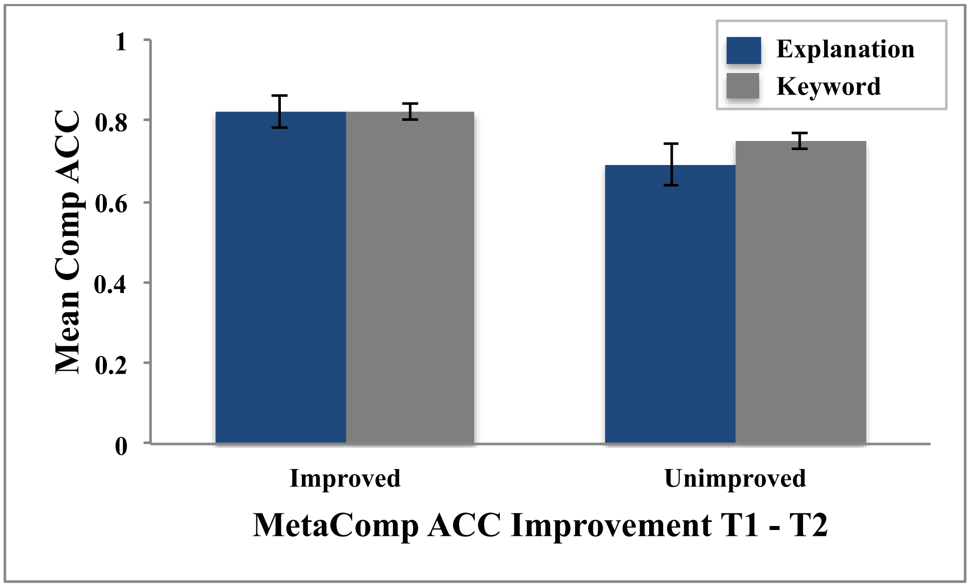

Figure 42. Metacomprehension accuracy improvement from Trials 1 to 2 on comprehension accuracy at Trial 3 in Experiment 2

Mean comprehension accuracy at Trial 3 for improved vs. unimproved metacomprehension accuracy from Trials 1 to 2. Error bars represents standard errors.

Similar to Experiment 1, the categorization of participants change values in metacomprehension categories used thus far in this participant-level analysis is a between subjects comparison, participants categorized as improved from trials 1 to 2 may not the same participants as those categorized as improved from trials 2 to 3. A separate Univariate ANOVA analysis was conducted to compare participants who showed metacomprehension accuracy improvement across trials (*n* = 19) and those who did not (*n* = 112) on comprehension accuracy at Trial 3. There was a significant main effect of Metacomprehension Accuracy Improvement across trials on comprehension accuracy at Trial 3, F (1,129) = 7.03, $p < .01$, partial $\eta^2 = .05$. Pairwise comparison showed comprehension accuracy to be higher at Trial 3 for participants with unimproved metacomprehension accuracy across trials ($M = .80$, $SE = .01$) when compared to participants with improved metacomprehension accuracy across trials (*M* = .70, *SE* = .03). Mean comparison is depicted in Figure 43. Although a flipped pattern than the one predicted was found, it should be noted that only 19 participants showed improvement in metacomprehension across trials in this experiment.

The Pearson's correlation between change in metacomprehension accuracy between Trials 1 and 3 (across trials) and comprehension accuracy at Trial 3 was not significant, $r(131) = .12$, $p > .05$. However, the small number of participants who showed metacomprehension accuracy improvement across trials was small.

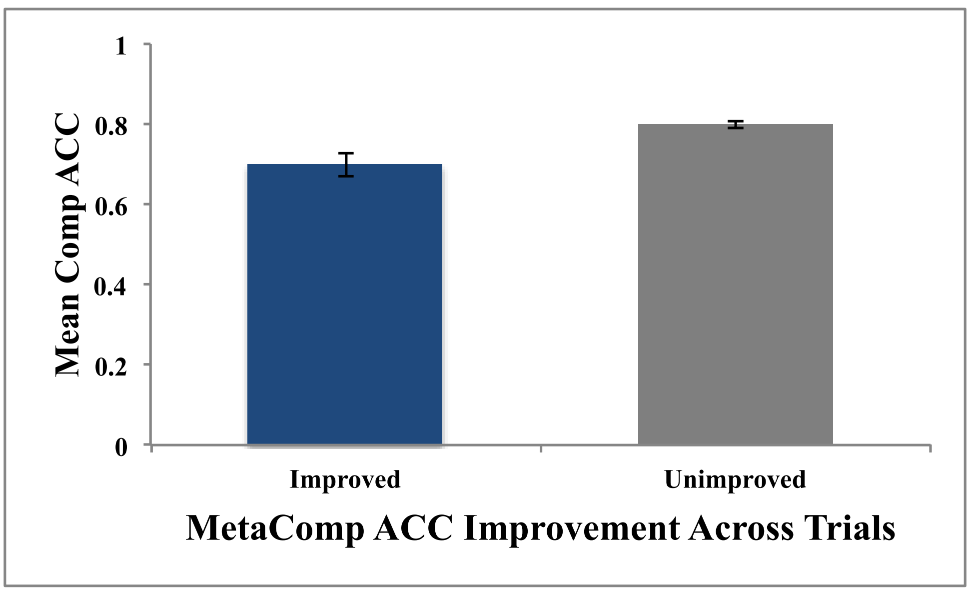

Figure 43. Metacomprehension accuracy improvement across trials on comprehension accuracy at Trial 3 in Experiment 2

Mean comprehension accuracy at Trial 3 for improved vs. unimproved metacomprehension accuracy across all trials. Error bars represents standard errors.

In conclusion, there are enough evidence to attribute this finding of improvement in comprehension accuracy but a decreased in metacomprehension accuracy at Trial 3 to the UPW effect (Koriat, 1997). However, in terms of fulfilling the aim of this experiment, there is still no evidence that participants selected the explanation learning strategy even when metacomprehension accuracy was improved. However, other factors can explain why participants may not select the learning strategy they use purely based on improved metacomprehension accuracy, as will be discussed in the General Discussion. It is also possible that the control process measure of this experiment via the selection of learning strategy is not sensitive enough to detect the impact of other control processes that could also improve comprehension accuracy, such as the time allocated to reading the text during restudy opportunity.

Post-Experiment Interview Questionnaire

After participants completed all sections of the experiment, they answered some questions about the experiment using the Post-Experiment Interview. Participants were instructed to answer as honestly as possible about their experience with the experiment. Each of the post-experiment interview questions were analyzed using the appropriate statistical test in order to determine whether there was a difference in the answers provided by participants across the different conditions. No participants were eliminated based on their responses on the post-experiment interview. There was, however, a rather interesting pattern of responses from the participants from the multi trial conditions who selected the keyword learning strategy at Trial 3. Participants assigned to the Multi Mix conditions were more likely to report that the reason they selected the keyword learning strategy was because it was the more effective learning strategy. In addition to believing that the keyword learning strategy was more effective, participants in the Multi Mix conditions also reported that experiencing both learning strategies helped their ability to predict memory across trials, although this supposition is contrary to the findings for metacomprehension accuracy in this experiment. This finding suggests that participants' failure to select the effective learning strategy should not purely be attributed to the failure in implementation of appropriate learning strategy selection of the control process. Instead, participants may have been monitoring familiarity of the text and repetition of the trials rather than the effect of the learning strategy on comprehension, especially by the time they were exposed to the text a second and third time.

On the other hand, participants assigned to the Multi Same conditions were more likely to report that they were selecting the learning strategy that was easier to use. In addition to thinking that they were choosing to use the easier learning strategy, they did not think their assigned same learning strategy affected their ability to predict their memory performances across trials. Details of the analysis of the post-experiment interview questions are presented in Appendix F.

Discussion

The aim of Experiment 2 was to examine the impact of increased metacomprehension accuracy on the control process of learning strategy selection at Trial 3. The multi trial metacomprehension paradigm was used in order to compare metacomprehension and comprehension accuracy for multiple versus single trials. The effect of learning strategy on metacomprehension and comprehension accuracy was also compared. The multi trial metacomprehension paradigm was appended to include two conditions with alternating effective and ineffective learning strategies at Trials 1 and 2, and these conditions were compared to two conditions for which the same learning strategy was used.

Participants selected the learning strategy used at Trial 3 creating two groups: those who selected the keyword learning strategy and those who selected the explanation learning strategy. Participants in the single trial condition were six times more likely to select the keyword learning strategy over the explanation learning strategy. This finding demonstrated the propensity of students to fail to spontaneously select strategies that will optimize their comprehension (Karpicke et al, 2009; McCabe, 2011). This propensity did not change with multiple trials; the proportion of participants who selected the keyword learning strategy was consistently higher across all multi trial conditions. Between the two conditions with alternative effective and ineffective learning strategy experiences at Trials 1 and 2, those who first experienced the effective learning strategy at Trial 1 were twice as likely to select the effective learning strategy at Trial 3 as compared to those who first experienced the ineffective learning strategy at Trial 1. One possible explanation for this finding is that people were reacting to a recency effect in that they did not want to reuse the strategy that they just experienced. Alternatively, this finding could have been affected by demand characteristics; participants could have thought that they should not repeat the learning strategy they used in Trial 2 because their experience thus far was to use alternating strategies across trials. They could have assumed that the researcher wanted them to pick the alternate strategy to the one they just used, rather than the one they wanted to use.

Comprehension accuracy was compared at Trial 1 to measure the baseline effectiveness of the two learning strategies; comprehension accuracy did not differ between the two learning strategies. This finding was surprising given that the learning strategies were selected based on prior research from the comprehension literature suggesting that there is a benefit to comprehension when a learning strategy promotes the integration of new information with long-term knowledge, as should have been required for the delayed explanation learning strategy. Not only did past research show that delayed explanation promoted comprehension (i.e., Bretzing & Kulhavy, 1979; Chi et al., 1994; Tan & Eakin, 2012), the nature of an elaborative summary requires the understanding of the gist of the text. Additionally, when written at a delay, information is retrieved from long-term memory to do explanation and therefore enhance the opportunity to integrate the new information with existing long-term knowledge.

Comprehension accuracy was also not better on Trial 3 for participants who chose the effective learning strategy as compared to those participants who chose the ineffective learning strategy. This finding replicated Experiment 1 for which the learning strategies on Trial 3 were researcher assigned, but went against the supposition that the reason no difference was obtained in Experiment 1 was because learning strategy was assigned rather than selected. However, concluding that metacomprehension accuracy did not improve the control process of strategy selection would be premature. The control process may not have had a differential effect on learning strategy selection at Trial 3 because there was also no effect of Trial Type or Learning Strategy on metacomprehension accuracy at Trial 3. Metacomprehension accuracy was the same for the multiple and single trials and the same for the two learning strategies. In Experiment 1, metacomprehension was better for the effective than ineffective learning strategy at Trial 3 when the learning strategy was assigned. However, when the learning strategy was self-selected, metacomprehension accuracy was not better for the effective learning strategy. This finding is perhaps not surprising in this analysis of only Trial 3, because most of the participants selected the ineffective learning strategy. When using the keyword learning strategy, the cues generated are not diagnostic of comprehension. Therefore, when participants monitor these cues, their metacomprehension accuracy suffers and it is not surprising that there was no metacomprehension accuracy advantage at Trial 3.

From the analysis of just the multiple trials, a main effect of Trial on comprehension accuracy was found; participants had the highest comprehension at Trial 3 followed by Trials 2 and 1, which did not differ from one another. At Trial 3, there was no effect of Learning Strategy on comprehension accuracy; comprehension was not better for the effective learning strategy. This could be because at this point, participants would have read the text three times and therefore have a diminishing return on the effect of one learning strategy over the other. Although comprehension accuracy changed significantly across the three trials, metacomprehension accuracy did not. However, comprehension accuracy was already high at Trial 1 and the change in comprehension accuracy, while significant, was not large across the three trials. Therefore, rather than concluding that metacomprehension accuracy had no effect on comprehension, it could be more accurate to conclude that metacomprehension accuracy—similar to comprehension accuracy—was actually at its peak at Trial 1 with little room for improvement in either monitoring or control. Although the GLMM fixed effect logistic regression analysis does not report absolute accuracy values, all of the estimates reported were significantly different from chance, indicating that people had accurate metacomprehension, and that improving their level of accuracy might not have been possible.

Other additional analyses were conducted to continue the investigation of the aim for this experiment, beginning with comprehension accuracy. It was suggested earlier that including the single trial conditions in the initial analysis could have obscured comprehension test scores. Therefore, additional analyses of comprehension accuracy were conducted on just the multiple trial conditions.

Although there was no main effect of Trial on metacomprehension accuracy overall, participant-level analyses were conducted to investigate whether participants who did show improvement in metacomprehension accuracy across trials also showed improvement in comprehension accuracy. Participants whose metacomprehension accuracy improved from Trials 1 to 2 showed improvement in comprehension accuracy Trial 2. Comprehension accuracy did not improve for a comparison group of participants whose metacomprehension did not improve.

Participants with improved metacomprehension accuracy from Trials 2 to 3 had lower comprehension accuracy at Trial 3 than participants with unimproved metacomprehension accuracy from Trials 2 and 3. Those participants with unimproved metacomprehension accuracy actually had better comprehension accuracy than those with improved metacomprehension accuracy. This finding is comparable to those of the pilot experiment, in which comprehension accuracy improved but metacomprehension accuracy did not. This finding was explained by the underconfidence with-practice (UWP) effect. According to the cue utilization hypothesis proposed by Koriat (1997), the fact that participants have had multiple exposures to the text serves as a less-diagnostic cue for comprehension. Participants are more likely to be influenced by the familiarity of the text when making DJOLs after reading the text for the third time than by the more-diagnostic cues generated by using an effective learning strategy. Improved metacomprehension accuracy from Trials 1 to 2 led to better comprehension accuracy at Trial 3 as compared to unimproved metacomprehension accuracy, providing further support for the possibility that the UPW effect influenced DJOLs in this experiment. Although comprehension accuracy improved, participants could have been underconfident if their metacomprehension judgments were based on less-diagnostic mnemonic cues.

A different explanation is possible for the group whose metacomprehension accuracy improved between Trials 2 and 3, but whose comprehension accuracy did not improve. It could have been the case that their improvement in metacomprehension accuracy came too late to affect comprehension performance at Trial 3. Given a fourth trial, it might have been possible to see the benefit of this metacomprehension accuracy improvement in comprehension accuracy.

The Pearson's correlations conducted provided more evidence of the relationship between metacomprehension accuracy and comprehension accuracy. Improved metacomprehension accuracy across trials was associated with improved comprehension accuracy and vice versa. However, this finding was obtained only across Trials 1 and 2, when the learning strategy was assigned. This correlation replicated those obtained in Experiment 1. Conversely, the correlation between change in metacomprehension accuracy and comprehension accuracy from Trials 2 to 3 was negative. Similar to the flipped pattern obtained in the ANOVA analysis, positive changes in metacomprehension accuracy was associated with lower comprehension accuracy and vice versa. For the very small number of participants who consistently improved across all three trials, the correlation between their change in metacomprehension accuracy and comprehension accuracy was positive, but not significant.

In conclusion, although there were no straightforward findings of the impact of improved metacomprehension accuracy on the control process of learning strategy selection, there was still evidence that suggested some of the participants benefited from the multi trial metacomprehension paradigm. Although not evident with the aggregate comparisons, participant-level analysis showed that when participant’s metacomprehension accuracy improved from the previous trial, it translated into improved comprehension accuracy on a subsequent trial. Because of the small proportion of participants who selected the effective learning strategy, comprehension accuracy improvement cannot be attributed to the control process of learning strategy at Trial 3. However, the selection of learning strategy is not the only control process that could impact comprehension performances.

CHAPTER VII

GENERAL DISCUSSION

The main aim of this dissertation was to investigate whether improving metacomprehension accuracy via the monitoring process impacts learning strategies implemented by the control process, such as which learning strategy to use while reading texts. A new paradigm—the multi trial metacomprehension paradigm—was introduced to serve this aim. The typical metacomprehension paradigm includes one trial consisting of reading comprehension, metacomprehension judgment, and comprehension test phases. The multi trial metacomprehension paradigm used this paradigm, but added two more trials of these phases. The goal of the new paradigm was to allow improvements in metacomprehension accuracy resulting from making repeated metacomprehension judgments in the first and second trials the opportunity to impact control processes during the third trial. The impact of these control processes was measured by allowing people to decide which learning strategy to use on the third trial. Experiment 1 first tested whether metacomprehension accuracy improved across multiple trials; learning strategies were assigned in all three trials and each trial was assigned the same learning strategy. Experiment 2 added conditions in which the learning strategy was alternated between the first two trials. For the third trial, the learning strategy was under the control of the participant; they could choose the learning strategy they wanted to use.

The learning strategies for the two experiments were selected to include one learning strategy that was previously found in the comprehension literature to be effective toward optimal comprehension, delayed explanation (Chi et al., 1994). As a comparison, the second learning strategy was selected because it was found to be less effective toward optimal comprehension, keyword-identification (Bretzing & Kulhavy, 1979). In addition, the effective learning strategy had previously been found to improve metacomprehension accuracy more than the ineffective learning strategy in a single trial (Thiede & Anderson, 2003; Thiede et al., 2003). Both metacomprehension accuracy and comprehension accuracy was compared across the two learning strategy conditions.

Metacomprehension judgments were absolute delayed Judgments of Learning (DJOLs) and Confidence Judgments (CJs) that were made on a continuous scale of 0 (certain not to remember) – 100 (certain to remember). Metacomprehension was evaluated in terms of sensitivity, or the mean magnitude of the judgments, organized by the experimental conditions. Metacomprehension was also evaluated in terms of accuracy using the generalized linear mixed effects modeling (GLMM) approach. The GLMM approach used comprehension accuracy as the dependent variable and metacomprehension judgments—centralized DJOLs—as the main predictor. Results obtained from the GLMM analysis are based on logistic regression, using one condition as a baseline group. For each comparison analysis, metacomprehension accuracy was represented as the amount of change in the slope relative to the dependent variable in comparison to the baseline group.

DJOL sensitivity and comprehension accuracy results were obtained using the multi-level mixed modeling approach. By using this approach, both participant and question level effects could be accounted for and included as random effects in the model. Additionally, because this approach calculated accuracy across all participants, constant values for any one participant and missing data can be accounted for without losing any participant, producing more complete results.

Highlights of the main findings informed the hypotheses generated from the Aims. In Experiment 1, hypotheses tests were done based on comparisons between conditions at Trial 3. Comprehension accuracy was better for participants who had multiple trials than for those who had a single trial. Multiple trials produced higher comprehension accuracy probably because participants read the same text three times over the course of three trials. There was no benefit of the presumably effective learning strategy over the ineffective learning strategy on comprehension accuracy. This finding is surprising given that comprehension literature has shown that the explanation learning strategy was effective in comprehension performance (i.e., Chi et al., 1994). The explanation learning strategy should have promoted the integration of new information with long-term knowledge, which based on prior research from the comprehension literature, should have benefitted comprehension; this benefit has been demonstrated in prior research (i.e., Bretzing & Kulhavy, 1979; Chi et al., 1994; Tan & Eakin, 2012). In addition, in order to write an elaborative summary, an understanding of the gist of the text is required. Additionally, when written at a delay, information used to write the summary is retrieved from long-term memory and therefore continues to increase the opportunity to integrate new information with existing long-term knowledge and serves as a retrieval-practice opportunity, which research on the testing effect has also shown to improve comprehension (Butler & Roediger, 2007; Karpicke & Roediger, 2007; 2008; Roediger & Karpicke, 2006; Tan & Eakin, 2012).

The findings for metacomprehension accuracy were the reverse of those found for comprehension accuracy. Metacomprehension accuracy was not better for multiple trials than single trials. The finding that metacomprehension accuracy was not better for multiple trials as compared to single trials was also surprising. These findings were unexpected because they were not consistent with findings from past research that has shown that having multiple trials improves metacomprehension accuracy (i.e., Glenberg & Epstein, 1985; Maki, 1998a). Although the multi trial metacomprehension paradigm was partly informed by the findings of Glenberg and Epstein (1985) who found more calibrated metacomprehension judgments after multiple trials, multiple trials in Experiment 1 did not produce this benefit over single trial conditions. Glenberg and Epstein (1985) did not include a single trial condition as comparison, so the analogy is not exact. However, as will be discussed, their finding of improved metacomprehension accuracy with repeated trials was replicated for some participants.

This finding of equal metacomprehension accuracy for multiple and single trials could suggest that the typical paradigm used in prior literature is sufficient in improving metacomprehension accuracy via the monitoring process. The single trial conditions were comparable to the typical paradigm used in metacomprehension literature (i.e., Anderson & Thiede, 2008; Thiede et al, 2003; Maki, 1998a), and metacomprehension accuracy was equal to that for multiple trials. However, because the monitoring process does not act alone during the comprehension process, the multi trial metacomprehension paradigm allows for a measurement of the impact caused by the updated meta level, informed by the monitoring process, on the control process.

It could be that the benefit of the effective learning strategy overrode any potential additional benefit of repeated opportunities to make and calibrate metacomprehension judgments. Subsequent analyses of just the multiple trial conditions showed that there were some participants whose metacomprehension accuracy improved across the three trials. For these participants, the expected improvement in comprehension accuracy also was obtained. These hypothesized findings that were not observed in the model that included all participants—and the model that made comparisons to single trials—actually provide strong support that in circumstances when metacomprehension is improved across trials, comprehension is also improved. This model was the only one that demonstrated the expected improvement in comprehension accuracy across trials.

As predicted, metacomprehension accuracy was better for the effective learning strategy of explanation than for the ineffective learning strategy of keyword identification. This finding, taken together with the comprehension accuracy findings, is consistent with past research (i.e., Linderholm et al., 2012; Thiede & Anderson, 2003) demonstrating that effective learning strategies impact metacomprehension accuracy, but not always comprehension accuracy. The adjustment to the metacomprehension paradigm used in this dissertation—allowing for repeated study trials with the text—did not change this finding. However, one possibility is that the benefit of using the effective learning strategy was obtained during the first trial, with little room for improvement across trials. This explanation also informs the finding of no difference between the single and multiple trials. The maximal benefit of using the effective learning strategy could have occurred at Trial 1 with little room for improvement across trials. This conclusion is supported by the finding of equal metacomprehension accuracy for the delayed explanation condition for Trials 1 and 3. Another reason that improved metacomprehension accuracy did not impact comprehension accuracy is because the learning strategy selection at the third trial was controlled by the researcher. If the control processes was informed by an improved metacomprehension, there was not an opportunity to implement a different learning strategy; the assigned learning strategy had to be used.

Although the reason that comprehension was not also improved with a supposedly effective learning strategy is still unclear, the finding that using an effective learning strategy is also effective at improving metacomprehension is not trivial. The goal is to improve metacomprehension accuracy so that appropriate control processes can be implemented to improve comprehension, given the opportunity. However, it is still beneficial for people to get better at assessing their comprehension, even if what they are getting better at is correctly assessing that they do not comprehend the information. Participants got better at this—metacomprehension accuracy improved—because the explanation learning strategy generated cues that were more diagnostic of future comprehension than the keyword learning strategy. This finding is consistent with the accessibility viewpoint (Koriat, Lichtenstein, & Fischhoff, 1980; Koriat, 1993) and the cue utilization hypothesis proposed by Hertzog et al., (2012).

However, the surprising finding for this analysis was that the learning strategy that produced the improved metacomprehension accuracy across the multiple trials was not the effective one. Rather, it was the keyword learning strategy that showed improvement in metacomprehension accuracy for the second trial, over the first and third trials. Although no significant difference in metacomprehension was found across the three trials for participants assigned to the explanation learning strategy, metacomprehension accuracy at the first and third trials was still significantly higher than improved metacomprehension accuracy at the second trial of the keyword condition. The conclusion that can be drawn from this interaction effect is that the explanation learning strategy was so effective at improving metacomprehension at the first trial that improvement across multiple trials was not possible. This conclusion requires the supposition that participants did not gain significantly more cues that were diagnostic of comprehension when using delayed explanation on the second and third trials than they already gained from using explanation on the first trial. This finding is an interesting one that has not previously been observed in the literature. Using an effective learning strategy not only improves metacomprehension, but does so efficiently from the first trial. Conversely, if an ineffective learning strategy is used, metacomprehension accuracy can be improved, but it will require multiple trials of using that less effective strategy to improve metacomprehension accuracy, and that improvement will still not reach the level of accuracy that using an effective strategy just once achieves. This lower metacomprehension accuracy could have occurred with the keyword learning strategy because when participants were given the opportunity to identify keywords on the second trial, they were able to identify new keywords that they missed during the first trial. In addition, in trying to avoid repeating keywords from the previous trial, they were in effect performing a retrieval practice on those words, which increased the connection between information in long-term memory and the information related to those keywords. They could then associate those words to the gist of the text, resulting in better comprehension. This multiple trial effect on the keyword identification task could be the reason that comprehension did not vary overall between the two learning strategies. Although still not as effective as the explanation learning strategy, multiple trials partially—not completely—made up for the use of the ineffective learning strategy.

The finding that participants whose metacomprehension accuracy improved also showed improved comprehension accuracy shows that, although the researcher controlled the learning strategy to be used for each trial, control processes were still being implemented that improved comprehension across trials. Strategies still under the control of the participant include allocation of study time or rereading portions of the text that they assessed to not be well learned. The focus on new keywords and retrieval practice for already identified keywords as previously discussed are also evidence of the implementation of control processes other than learning strategy selection. This adjustment at the control level due to more accurate metacomprehension could explain the why comprehension accuracy improved for this group of participants who also showed improved metacomprehension, regardless of which learning strategy to which they were assigned.

McCabe (2011) found that participants can be trained to identify the more effective learning strategy over the less effective after some form of training, but that does not automatically translate into actual selection and use of these effective learning strategies when actually studying for an exam (i.e., Karpicke et al., 2009; Tan & Eakin, 2012). Participants who only experienced a single trial overwhelmingly selected the ineffective learning strategy to use while reading the text for that trial. Additionally, in this experiment, using a supposedly effective learning strategy did not automatically result in better comprehension accuracy as compared to a supposedly ineffective one. However, using an effective learning strategy did improve metacomprehension, and that improved metacomprehension was associated with improved comprehension. Theoretically, this improvement happened because the cues generated by using the effective learning strategy were also diagnostic of future comprehension; this almost incidental association led to accurate metacomprehension when judgments were based on these diagnostic cues (Hertzog et al., 2010). The effect of improved metacomprehension was the implementation of appropriate control processes—even when learning strategies were prescribed—which led to improved comprehension accuracy. The findings from Experiment 1 support the theoretically understanding of the dynamic interplay between the meta and object levels, and between the monitoring and control processes of the metacomprehension framework (Nelson & Narens, 1990). To my knowledge, the findings from Experiment 1 are the first in the literature showing that improving metacomprehension accuracy can also improve comprehension accuracy.

Experiment 2 was designed to directly measure the impact of improving metacomprehension accuracy on the control process of learning strategy selection by allowing participants to choose the learning strategy they wanted to use for the third trial. To aid in their selection, some participants used alternating strategies in the first two trials. Other conditions replicated those used in Experiment 1 in which only one learning strategy was used for the first two trials—either explanation or keyword—and a single trial condition as the baseline condition during which participants were allowed to select their learning strategy on a single trial.

In Experiment 2, hypotheses testing were also done based on comparisons between conditions at Trial 3 using the same analysis procedures as used in Experiment 1. The unique aspect of Experiment 2 was that participants selected the learning strategy to use during Trial 3. Strikingly, for the single trial, participants overwhelmingly selected the ineffective keyword learning strategy. This finding supports findings in the literature that students do not know effective versus ineffective learning strategies and frequently fail to choose strategies that will benefit comprehension (Karpicke et al, 2009; McCabe, 2009). The prediction that the effective learning strategy would be selected more frequently by participants who use that learning strategy in Trials 1 and/or 2 was not born out overall. Participants in multiple trials were no more likely to select the effective learning strategy than those in the single trial. In addition, participants who experienced both learning strategies were no more likely to select the effective learning strategy than those who used the same learning strategy in the first two trials; neither were those who experienced the effective learning strategy as compared to the one condition in which participants never experienced the effective learning strategy. The only comparison that produced a difference was the comparison just between the two conditions for which learning strategy alternated; there was an effect of order of the learning strategy in this comparison. Participants who first used explanation and then keyword were twice as likely to select the explanation at Trial 3 than those who first used keyword and then explanation. Possible explanations for this finding are that participants simply chose whatever strategy was different from the one they had just used, producing a kind of recency effect. Participants might have actually presumed that the researcher wanted them to use a different strategy. The pattern of the previous trials alternated and could have inadvertently created a demand characteristic that led participants to choose the learning strategy on the third trial that fit the pattern produced by previous trials. Although it would be interesting to investigate whether metacomprehension accuracy differed across trials within these two groups of participants, this model would not compile due to low power.

In Experiment 2, comprehension accuracy was better for the multiple trials than for the single trial. This is not surprising because participants in the multi trial conditions had three opportunities to read the text. Also consistent with Experiment 1, using a presumably effective learning strategy did not benefit comprehension accuracy over using a presumably ineffective one. This finding is in contrast to findings from the comprehension literature that not only found a benefit to comprehension from using the explanation learning strategy (i.e., Chi et al., 1994), but also the process of explanation should fit with theories about why some learning strategies are more effective than others. The explanation learning strategy, especially when implemented after a delay, should allow for integration of new information with long-term knowledge, allowing for a deeper understanding of the gist of the text, and allowing for repeated retrieval from long-term memory; all of these effects have been demonstrated to improve comprehension (Bretzing & Kulhavy, 1979; Chi et al., 1994; Karpicke & Roediger, 2006; Karpicke & Roediger, 2007; 2008; Tan & Eakin, 2012

Metacomprehension accuracy did not improve, overall, with multiple trials over single trials; this finding was consistent with Experiment 1. Additionally, metacomprehension accuracy was the same for both learning strategies; metacomprehension accuracy was not better for the effective learning strategy. This finding is not consistent with Experiment 1 and past research (i.e., Thiede & Anderson, 2003; Thiede et al., 2003). One major difference is that in Experiment 2, learning strategy was self-selected. Apparently, giving participants the opportunity to self-select which learning strategy to use, any benefit of metacomprehension accuracy with experience of the effective learning strategy, and thereby comprehension accuracy, is lost. The comparison might have been hindered by the low number of participants who selected the explanation learning strategy. However, even for the subset of participants who showed improved metacomprehension accuracy with multiple trials, this benefit of improved metacomprehension for comprehension accuracy was mixed. Consistent with findings from Experiment 1, participants whose metacomprehension accuracy improved also showed better comprehension accuracy than those whose metacomprehension accuracy was not improved. However, this was true only when participants showed the metacomprehension accuracy improvement between the first two trials, not for participants who showed improvement in metacomprehension accuracy only at the third trial. Comprehension accuracy was actually lower for participants whose metacomprehension accuracy improved across the second and third trials and better for participants whose metacomprehension accuracy did not improve.

For participants whose metacomprehension accuracy improved, it might have been too late to observe the impact of that improvement on comprehension accuracy. That benefit might have been observed if there had been a fourth trial on which the effect of improved metacomprehension accuracy on control processes during the next comprehension phase could have been observed. The finding for the second (unimproved) group, which was also obtained in the pilot experiment, could be attributed to the underconfidence with-practice effect (Koriat, 1997). On the third trial, participants could have switched the basis of their metacomprehension judgments from diagnostic to nondiagnostic cues. Repeatedly experience with the text and test questions could have resulted in DJOLs being based more on familiarity than on integrated information from the text, producing inaccurate metacomprehension for this trial as compared to previous trials for which they based their predictions on more diagnostic cues. An additional analysis was conducted to compare participant's performances at the third trial with their metacomprehension improvement from the first two trials. Comprehension accuracy was found to be in the appropriate predicted direction; participants who showed improvement in metacomprehension at the second trial showed better comprehension accuracy at the third trial.

Theoretically, accurate monitoring—metacomprehension accuracy—results in updating of the meta level about the state of the object level such that appropriate learning strategies can be implemented by the control process on the object level. However, the findings from the two experiments in this dissertation demonstrated that participants do not always select that strategy. Other factors could also have influenced their learning strategy selection, other than the degree to which their monitoring is an accurate prediction of their comprehension. For instance, there was a "metacognitively savvy" group of participants in Tan and Eakin (2012) who presumably selected the less effective learning strategy because the benefit for comprehension accuracy was not enough to justify the increased effort required to use the effective learning strategy.

Because the “effective” learning strategy did not actually produce better comprehension than the “ineffective” learning strategy, the participants in Experiment 2 could also have done a similar cost-benefit analysis and determined that using the more effortful learning strategy was not worth the extra effort it required. Alternatively, participants might not have been savvy at all, and because there was no real-life cost to doing poorly on the comprehension test in an experiment, they simply selected the easiest learning strategy to use.

The data from the post experimental interview suggests that participants were neither doing a cost-benefit analysis nor were they choosing the easiest learning strategy. Participants who experienced both learning strategies said that they reason the selected the keyword learning strategy was because they thought it was the more effective one as compared to the explanation learning strategy. Only participants who experienced the same learning strategies across the first two trials reported selecting the learning strategy that was easiest to use; frequency of selecting the keyword learning strategy did not differ between these two groups. Participants who experienced both learning strategies also thought that using the keyword learning strategy led to more accurate metacomprehension predictions; this finding was contrary to the empirical evidence of their metacomprehension accuracy. The analysis of the post-interview questionnaire suggests that participants were selecting the keyword over the explanation learning strategy for other reasons that because it was easier or because they knew it would have no impact on comprehension accuracy.

One factor that could have confounded the results for learning strategy selection was that the same text was used for all three trials. Just as metacomprehension accuracy could have been affected by seeing the text three times due to the UWP effect (Koriat, 1997), the control process of learning strategy selection also could have been affected. Because participants knew they were going to study the same text a third time, they might not have deemed it necessary to use the effective learning strategy on the third trial, relying instead on the effect of repetition on comprehension. One way to counter this in a future experiment would be to present a new text during the third trial, as was done by Tan & Eakin (2012). Participants would need to be informed that they would be studying a new text, which would perhaps influence them to select learning strategies based on their monitoring accuracy instead of other factors.

The dissertation experiments focused on the control process of learning strategy selection, but that is not the only process implemented by control. As discussed, other control processes include allocation of study time and termination of study. The dissertation experiments did not experimentally manipulate and/or hold consistent the other control processes that could have impacted comprehension accuracy, nor would it have been possible to do so. In addition, perhaps the two learning strategies selected did not produce disparate enough results on comprehension accuracy to measure any differential effects of metacomprehension accuracy on selection of one learning strategy over the other. Perhaps, a more effective learning strategy to use would be that of *concept mapping*. Although difficult to implement experimentally, this learning strategy has been empirically shown to improve both metacomprehension and comprehension accuracy. Concept mapping has been shown to produce a high correlation between the quality of the concept map produced and metacomprehension accuracy and comprehension accuracy (i.e., Tan & Eakin, 2015), suggesting that it is effective in applying the characteristics required by the comprehension literature for producing good comprehension as well as in generating cues that, when monitored, are diagnostic of future comprehension.

Another factor that could have impacted the decision on learning strategy selection is the type of test. Unlike most metacomprehension accuracy findings in the literature, and these dissertation experiments, the pilot experiment reported a steady and significant improvement of metacomprehension accuracy across trials, which also led to improved comprehension accuracy. One main difference between the pilot experiment and the experiments conducted in this dissertation is the type of test administered. Participants in the pilot experiment answered short answer test questions whereas participants in the dissertation experiments answered multiple-choice questions. Although attempts were made to ask questions that were inferential rather than surface level, the keyword learning strategy was effective enough to produce at least 70% comprehension in Experiments 1 and 2. Linderholm et al. (2012) concluded that metacomprehension accuracy could also be attributed to the amount of cognitive effort a task requires. Therefore, perhaps the multiple choice test questions used did not elicit the cognitive effort required to promote metacomprehension accuracy. This lack of cognitive effort could explain the lack of evidence in the effect of learning strategy on metacomprehension accuracy, especially in Experiment 2. Furthermore, Pressley et al., (1990) concluded that participants who answered short-answer questions were better at monitoring accuracy, which later led to being better at making decisions about whether they needed to restudy or terminate study. The act of generating an answer for the short answer questions indirectly generated cues that were diagnostic of their later comprehension accuracy. This explanation could be the reason that the benefits of using the multi trial metacomprehension paradigm were observed in the pilot experiment, but not the dissertation experiments.

Finally, participants in both experiments saw the same text multiple times across three trials. Degree of text familiarity and number of repetitions could have served as powerful alternative heuristics on which JOLs were based that overrode any diagnostic cues generated from using the effective over the ineffective learning strategy. In effect, because the paradigm itself created these extremely salient cues due to the repeated trials, using the multi trial metacomprehension paradigm could have masked any potential influence of learning strategy on both metacomprehension and comprehension accuracy.

Conclusions

The main contribution of this dissertation is the finding that using an effective learning strategy improved both metacomprehension accuracy and comprehension accuracy, but the two were not impacted in the same way by multiple trials. Using an effective learning strategy was successful at improving metacomprehension accuracy to its highest potential with just one trial. When using an ineffective learning strategy, however, multiple trials were required for metacomprehension accuracy to improve. Comprehension accuracy also improved with multiple trials, but metacomprehension accuracy was never as good as compared to using the effective learning strategy once. Taken together with the finding from this dissertation and other research, students are inclined to choose an ineffective over an effective learning strategy, this finding could explain why students rely so much on repetition to improve their comprehension of texts.

The only way they can improve comprehension when they use ineffective learning strategies is to use them repeatedly.

The implementation of the multi trial metacomprehension paradigm was successful in producing improved metacomprehension accuracy across trials for a subset of the participants. Those whose metacomprehension accuracy improved also showed improved comprehension accuracy as compared to those whose metacomprehension accuracy did not improve. However, this improved metacomprehension accuracy did not lead to selection of the effective learning strategy when given the opportunity to choose. This finding brings into question the motivation behind most of the metacomprehension literature to determine how to improve metacomprehension accuracy so that students trained to do so will then automatically select effective over ineffective learning strategies. Not only is the control process ignored in most of the literature, as the present experiments show, the relationship between monitoring accuracy and effective control processes is not always clear. The theory that providing the meta level with accurate monitoring will result in the implementation of appropriate control processes, such as learning strategy selection, on the object level is also brought into question. Although the theory is somewhat protected by the fact that other non-manipulated and unmeasured control processes could have influenced the results, the push toward only improving metacomprehension accuracy is not. Left to their own devices, even metacomprehensively accurate students did not choose the effective learning strategy.

Even after experiencing both learning strategies, participants did not choose the effective learning strategy. There is unfortunately no easy solution to get students to select an effective learning strategy when reading texts. However, taking all the main findings from this dissertation together, it can be concluded that, while it is not sufficient to only improve metacomprehension accuracy, it is necessary to do so before comprehension can benefit. Because the ultimate goal in improving metacomprehension accuracy is to improve comprehension, this improvement has to happen before the meta level can appropriately implement control processes on the object level. To promote monitoring accuracy, students should not just be told to select the effective learning strategy because we know that to be an unsuccessful method. Instead, students should be directed to think about cues that are diagnostic of comprehension when making metacomprehension assessments. When students learn to make accurate metacomprehension assessments, monitoring accuracy improves and updates the meta level which will in turn impact the control process so that comprehension accuracy will also improve.

REFERENCES

Anderson, M. C. M., & Thiede, K. W. (2008). Why do delayed summaries improve metacomprehension accuracy? *Acta Psychologica*, *128*(1), 110–118. http://doi.org/10.1016/j.actpsy.2007.10.006

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48. doi:10.18637/jss.v067.i01.

Bretzing, B. H., & Kulhavy, R. W. (1979). Notetaking and depth of processing. *ContemporaryEducational Psychology*, *4*(2), 145–153. http://doi.org/10.1016/0361-476X(79)90069-9

Butler, A. C., & Roediger, H. L. I. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology*, *19*(4-5), 514–527. http://doi.org/10.1080/09541440701326097

Chi, M. T. H., de Leeuw, N., Chiu, M.-H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*(3), 439–477.

Dunlosky, J., Baker, J. M. C., Rawson, K. A., & Hertzog, C. (2006). Does aging influence people's metacomprehension? Effects of processing ease on judgments of text learning. *Psychology and Aging*, *21*(2), 390–400. http://doi.org/10.1037/0882-7974.21.2.390

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. http://doi.org/10.1111/j.1467-8721.2007.00509.x

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*,*14*(1), 4–58. http://doi.org/10.1177/1529100612453266

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*(4), 551–565. http://doi.org/10.1016/j.jml.2005.01.011

Eakin, D. K. (2005). Illusions of knowing: Metamemory and memory under conditions of retroactive interference. *Journal of Memory and Language*, *52*(4), 526-534. http://dx.doi.org/10.1016/j.jml.2005.01.009

Eakin, D. K., & Hertzog, C. (2006). Release from implicit interference in memory and metamemory: Older adults know that they can't let go. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *61*(6), 340-347. http://dx.doi.org/10.1093/geronb/61.6.p340

Eakin, D. K., & Hertzog, C. (2012). Age invariance in feeling of knowing during implicit interference effects. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *67*(5), 555-562. http://doi.org/10.1093/geronb/gbr150

Eakin, D. K., Hertzog, C., & Harris, W. (2014). Age invariance in semantic and episodic metamemory: Both younger and older adults provide accurate feeling-of-knowing for names of faces. *Aging, Neuropsychology, and Cognition*, *21*(1), 27-51. http://dx.doi.org/10.1080/13825585.2013.775217

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191.

Fukaya, T. (2010). Factors affecting the accuracy of metacomprehension: A meta-analysis. *Japanese Journal of Educational Psychology*, *58*(2), 236–251. http://doi.org/10.5926/jjep.58.236

Gardiner, J. M., & Hampton, J. A. (1985). Semantic memory and the generation effect: Some tests of the lexical activation hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(4), 732–741. http://doi.org/10.1037/0278-7393.11.1-4.732

Gernsbacher, M. A. (1997). Two decades of structure building. *Discourse Processes*, *23*(3), 265–304. http://doi.org/10.1080/01638539709544994

Glenberg, A. M., & Epstein, W. (1985). Calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(4), 702–718. http://doi.org/10.1037/0278-7393.11.1-4.702

Glenberg, A. M., & Epstein, W. (1987). Inexpert calibration of comprehension. *Memory & Cognition*, *15*(1), 84–93.

Glenberg, A. M., Sanocki, T., Epstein, W., & Morris, C. (1987). Enhancing calibration of comprehension. *Journal of Experimental Psychology: General*, *116*(2), 119–136. http://doi.org/10.1037/0096-3445.116.2.119

Goodman, L. A., & Kruskal, W. H. (1954). Measures of Association for Cross Classifications. *Journal of the American Statistical Association*, (268), 732.

Griffin, T. D., Jee, B. D., & Wiley, J. (2009). The effects of domain knowledge on metacomprehension accuracy. *Memory & Cognition*, *37*(7), 1001–1013. http://doi.org/10.3758/MC.37.7.1001

Hartlep, K. L., & Forsyth, G. A. (2000). The effect of self-reference on learning and retention. *Teaching of Psychology*, *27*(4), 269–271. http://doi.org/10.1207/S15328023TOP2704_05

Hertzog, C., Dunlosky, J., & Sinclair, S. M. (2010). Episodic feeling-of-knowing resolution derives from the quality of original encoding. *Memory & Cognition*, *38*(6), 771-784. http://doi.org/10.3758/MC.38.6.771

Jee, B., Wiley, J., & Griffin, T. (2006). Expertise and the illusion of comprehension. In *Proceedings of the Annual Conference of the Cognitive Science Society* (pp. 387-392).

Karpicke, J. D., Butler, A. C., & Roediger, H. L. I. (2009). Metacognitive strategies in student learning: Do students practice retrieval when they study on their own? *Memory*, *17*(4), 471–479. http://doi.org/10.1080/09658210802647009

Karpicke, J. D., & Roediger, H. L. I. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language*, *57*(2), 151–162. http://doi.org/10.1016/j.jml.2006.09.004

Karpicke, J. D., & Roediger, H. L. I. (2008). The critical importance of retrieval for learning. *Science*, *319*(5865), 966–968. http://doi.org/10.1126/science.1152408

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*(5), 363–394. http://doi.org/10.1037/0033-295X.85.5.363

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review*, *100*(4), 609–639. http://doi.org/10.1037/0033-295X.100.4.609

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. http://doi.org/10.1037/0096-3445.126.4.349

Koriat, A. (1998). Illusions of knowing: The link between knowledge and metaknowledge. In V.Y. Yzerbyt, G. Lories, B. Dardenne, V. Y. (Ed) Yzerbyt, G. (Ed) Lories, & B. (Ed) Dardenne (Eds.), *Metacognition: Cognitive and social dimensions.* (pp. 16–34). Thousand Oaks, CA, US: Sage Publications, Inc.

Koriat, A. (2000). Control processes in remembering. In E. Tulving, F. I. M. Craik, E. (Ed) Tulving, & F. I. M. (Ed) Craik (Eds.), *The Oxford handbook of memory.* (pp. 333–346). New York, NY, US: Oxford University Press.

Koriat, A. (2012). The relationships between monitoring, regulation and performance. *Learning and Instruction*, *22*(4), 296–298. http://doi.org/10.1016/j.learninstruc.2012.01.002

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 107–118. http://doi.org/10.1037/0278-7393.6.2.107

Koriat, A., Sheffer, L., & Ma'ayan, H. (2002). Comparing objective and subjective learning curves: Judgments of learning exhibit increased underconfidence with practice. *Journal of Experimental Psychology: General*, *131*(2), 147–162. http://doi.org/10.1037/0096-3445.131.2.147

Linderholm, T., Wang, X., Therriault, D., Zhao, Q., & Jakiel, L. (2012). The accuracy of metacomprehension judgements: The biasing effect of text order. *Electronic Journal of Research in Educational Psychology*, *10*(1), 111–128.

Maki, R. H. (1998a). Predicting performance on text: Delayed versus immediate predictions and tests. *Memory & Cognition*, *26*(5), 959–964. http://doi.org/10.3758/BF03201176

Maki, R. H. (1998b). Metacomprehension of text: Influence of absolute confidence level on bias and accuracy. In D. L. Medin & D. L. (Ed) Medin (Eds.), *The psychology of learning and motivation: Advances in research and theory, Vol. 38.* (pp. 223–248). San Diego, CA, US: Academic Press.

Maki, R. H., & Berry, S. L. (1984). Metacomprehension of text material. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(4), 663–679. http://doi.org/10.1037/0278-7393.10.4.663

Maki, R. H., Foley, J. M., Kajer, W. K., Thompson, R. C., & Willert, M. G. (1990). Increased processing enhances calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(4), 609–616. http://doi.org/10.1037/0278-7393.16.4.609

Maki, R., Mikkelsen, B., & Gerlach, T. (1988). Calibration of comprehension depends upon type of test.

Maki, R. H., & Serra, M. (1992). The basis of test predictions for text material. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(1), 116–126. http://doi.org/10.1037/0278-7393.18.1.116

Maki, R. H., & Swett, S. (1987). Metamemory for narrative text. *Memory & Cognition*, *15*(1), 72–83. http://doi.org/10.3758/BF03197713

McCabe, J. (2011). Metacognitive awareness of learning strategies in undergraduates. *Memory & Cognition*, *39*(3), 462–476. http://doi.org/10.3758/s13421-010-0035-2

McNamara, D. S. (2004). SERT: Self-Explanation Reading Training. *Discourse Processes*, *38*(1), 1–30. http://doi.org/10.1207/s15326950dp3801_1

Mengelkamp, C., & Bannert, M. (2010). Accuracy of confidence judgments: Stability and generality in the learning process and predictive validity for learning outcome. *Memory & Cognition*, *38*(4), 441–451. http://doi.org/10.3758/MC.38.4.441

Metcalfe, J. (2009). Metacognitive judgments and control of study. *Current Directions in Psychological Science*, *18*(3), 159–163. http://doi.org/10.1111/j.1467-8721.2009.01628.x

Murayama, K., Sakaki, M., Yan, V. X., & Smith, G. M. (2014). Type I error inflation in the traditional by-participant analysis to metamemory accuracy: A generalized mixed-effects model perspective. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(5), 1287–1306. http://doi.org/10.1037/a0036914

Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The "delayed-JOL effect." *Psychological Science*, *2*(4), 267–270. http://doi.org/10.1111/j.1467-9280.1991.tb00147.x

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *The psychology of learning and motivation*, *26*, 125-141.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(2), 175–220. http://doi.org/10.1037/1089-2680.2.2.175

Parr, J. M., & McNaughton, S. (2014). Making connections: The nature and occurrence of links in literacy teaching and learning. *The Australian Journal of Language and Literacy*,*37*(3), 141=150.

Pierce, B. H., & Smith, S. M. (2001). The postdiction superiority effect in metacomprehension of text. *Memory & Cognition*, *29*(1), 62–67. http://doi.org/10.3758/BF03195741

Pressley, M., Ghatala, E. S., Woloshyn, V. E., & Pirie, J. (1990). Sometimes adults miss the main ideas and do not realize it: Confidence in responses to short-answer and multiple-choice comprehension questions. *Reading Research Quarterly*, *25*(3), 232–249. http://doi.org/10.2307/748004

Psychology Software Tools, Inc. [E-Prime 2.0]. (2012). Retrieved from http://www.pstnet.com

Rawson, K. A., & Dunlosky, J. (2002). Are performance predictions for text based on ease of processing? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(1), 69–80. http://doi.org/10.1037/0278-7393.28.1.69

Rawson, K. A., & Dunlosky, J. (2011). Optimizing schedules of retrieval practice for durable and efficient learning: How much is enough? *Journal of Experimental Psychology: General*, *140*(3), 283–302. http://doi.org/10.1037/a0023956

Rawson, K. A., Dunlosky, J., & Theide, K. W. (2000). The rereading effect: Metacomprehension accuracy improves across reading trials. *Memory & Cognition*, *28*(6), 1004–1010. http://doi.org/10.3758/BF03209348

Roediger, H. L. I., & Karpicke, J. D. (2006). Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention. *Psychological Science*, *17*(3), 249–255. http://doi.org/10.1111/j.1467-9280.2006.01693.x

Schunk, D. H., & Rice, J. M. (1989). Learning goals and children's reading comprehension. *Journal of Reading Behavior*, *21*(3), 279–293.

Schunk, D. H., & Rice, J. M. (1991). Learning goals and progress feedback during reading comprehension instruction. *Journal of Reading Behavior*, *23*(3), 351–364.

Serra, M. J., & Dunlosky, J. (2010). Metacomprehension judgements reflect the belief that diagrams improve learning from text. *Memory*, *18*(7), 698–711. http://doi.org/10.1080/09658211.2010.506441

Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(1), 204–221. http://doi.org/10.1037/0278-7393.26.1.204

Tan, E. W., & Eakin, D. K. (2012). Updating Metacomprehension: Selecting Effective Learning Strategies after Experience. Unpublished manuscript.

Tan, E. W., & Eakin, D. K. (2015). *Can Comprehenders Meta-Map?* Unpublished manuscript.

Thiede, K. W., & Anderson, M. C. M. (2003). Summarizing can improve metacomprehension accuracy. *Contemporary Educational Psychology*, *28*(2), 129–160. http://doi.org/10.1016/S0361-476X(02)00011-5

Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology*, *95*(1), 66–73. http://doi.org/10.1037/0022-0663.95.1.66

Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(4), 1024–1037. http://doi.org/10.1037/0278-7393.25.4.1024

Thiede, K. W., Dunlosky, J., Griffin, T. D., & Wiley, J. (2005). Understanding the Delayed-Keyword Effect on Metacomprehension Accuracy. *Journal of Experimental Psychology: Learning, Memory,* and *Cognition*, *31*(6), 1267–1280. http://doi.org/10.1037/0278-7393.31.6.1267

Thiede, K. W., Wiley, J., & Griffin, T. D. (2011). Test expectancy affects metacomprehension accuracy. *British Journal of Educational Psychology*, *81*(2), 264–273. http://doi.org/10.1348/135910710X510494

Trabasso, T., Secco, T., & van den Broek, P. (1984). Causal cohesion and story coherence. In H. Mandl, N. L. Stein, & T. Trabasso (Eds.), *Learning and comprehension of text* (pp. 83- 111). Hillsdale, NJ: Erlbaum. van den Broek, P., White, M. J., Kendeou, P., & Carlson, S. (2009). Reading between the lines: Developmental and individual differences in cognitive processes in reading comprehension. In R. K. Wagner, C. Schatschneider, & C. Phythian-Sence (Eds.), *Beyond decoding: The behavioral and biological foundations of reading comprehension* (pp. 107-123). New York, NY: Guilford Press.

Weaver, C. A. (1990). Constraining factors in calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(2), 214–222. http://doi.org/10.1037/0278-7393.16.2.214

Weaver, C. A., & Bryant, D. S. (1995). Monitoring of comprehension: The role of text difficulty in metamemory for narrative and expository text. *Memory & Cognition*, *23*(1), 12–22. http://doi.org/10.3758/BF03210553

Winne, P. H., & Jamieson-Noel, D. (2002). Exploring students' calibration of self reports about study tactics and achievement. *Contemporary Educational Psychology*, *27*(4), 551–572. http://doi.org/10.1016/S0361-476X(02)00006-1

Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science*, *6*(5), 292–297. http://doi.org/10.1111/j.1467-9280.1995.tb00513.x

APPENDIX A

IRB DOCUMENTS

Mississippi State University
Informed Consent Form for Participation in Research

Title of Research Study: Learning to Comprehend

Researchers: Elaine W. Tan, Mississippi State University
Dr. Deborah K. Eakin, Mississippi State University

You are being asked to be a volunteer in a research study. The purpose of this form is to tell you about the study you will be participating in today and to inform you about your rights as a research volunteer. Before you participate, you should read this consent form carefully and completely. You will be given a copy of this consent form to keep and you do not waive any of your legal rights by signing this consent form.

Thank you for volunteering to participate in this study. Our work could not be done without your help and willingness to give of your time and yourself.

Purpose

The purpose of this research is to investigate people's ability to monitor text comprehension and to accurately predict comprehension test performance.

Procedures

If you decide to participate in this study, you will read and study texts using a learning strategy assigned to you. After studying the texts, you will be tested on your comprehension of the texts. When presented with each question, you will first give a prediction of how well you will be able to answer that question before proceeding to answering the question.

You will repeat the same procedure in three more phases using either the same or a different learning strategy assigned to you at the beginning of the next phase.

This study should take approximately 2 hours.

Risks or Discomforts

There are no major physical discomforts involved in this study. Risks are minimal and do not exceed those of normal office work. Please tell us if you are having trouble with any task or if you need additional rest and the investigators will be happy to accommodate you in any way possible. If you feel any discomfort, please tell the person assisting you immediately.

Benefits

This study will provide insight about whether people can monitor their text comprehension and accurately predict their performance on comprehension tests. You will not likely to benefit personally in any way from joining this study, but thanks to the willingness of people like you, we will continue to learn about how people monitor text comprehension.

MSU HRPP Approved: 10.09.2015 Expires: 12.31.2016
IRB # 15-347

Incentive to participate

The study will take approximately 2 hours, and you will be reimbursed for your participation. You will receive ½ research credit for every half-hour of participation. In addition, all participants who score a 70% or greater on the final comprehension test will receive a .5 research credit bonus. This bonus is to encourage you to continue to work hard throughout the study.

We want you to know, however, that you are free to change your mind and withdraw from this research at any time. There will be no penalty for doing so. You will receive compensation equal to the time involved in the study. You will receive no less than .5 research credit, and will receive .5 research credit for each half-hour of participation.

Confidentiality

All of your responses will be kept strictly confidential. To protect the confidentiality of this information, we will assign your data a code number that will only be known to the members of this research project. All of the information which you provide us today will be marked with the code number, not your name. All information will be stored in a computer for analysis using only your code number for identification. No identification of your individual answers to questions will be given to anyone. We want you to be completely confident that you may feel free to answer all questions without concern that it may affect you in any way.

Your name and identifying information will not be connected in any way to your responses in this study. The online system will automatically grant you credit when you complete the study by submitting your PRP Identity Code back to the SONA system while your responses will be stored in a different database accessible to the researcher. If you are participating in a lab study (in-person research), be sure to bring your "Identity Code" (available under the "My Profile" tab on the PRP website) with you to the study, so that you may be granted credit for your participation.

Please note that these records will be held by a state entity and therefore are subject to disclosure if required by law. Research information may be shared with the MSU Institutional Review Board (IRB) and the Office for Human Research Protection (OHRP).

Questions

If you have any questions about this research project, please feel free to contact Elaine Tan at ewt38@msstate.edu or Dr. Deborah Eakin at de115@msstate.edu.

For questions regarding your rights as a research participant, or to discuss problems, express concerns or complaints, request information, or offer input, please feel free to contact the MSU Research Compliance Office by phone at (662) 325-3994, or by email at irb@research.msstate.edu, or on the web at http://orc.msstate.edu/humansubjects/participant/

Voluntary Participation

Please understand that your **participation is voluntary.** Your **refusal to participate will involve no penalty or loss** of benefits to which you are otherwise entitled. You **may discontinue your participation** at any time without penalty or loss of benefits.

MSU HRPP Approved: 10.09.2015 Expires: 12.31.2016 IRB # 15-347

Please take all the time you need to read through this document and decide whether you would like to participate in this research study.

If you agree to participate in this research study, please sign below. You will be given a copy of this form for your records.

________________________________ __________
Participant Signature Date

________________________________ __________
Investigator Signature Date

Debriefing Statement

Thank you again for participating in our study. Please read this statement carefully, and ask the person assisting you if you have any questions.

The study in which you just participated was design to investigate the degree to which people are able to accurately predict their own comprehension. We used both effective (explaining what you read in a summary) and ineffective (writing down key words on a notecard) strategies, and measured how your comprehension and your assessment of your own comprehension changed. In the future, when you study you should try to use an effective strategy, such as trying to explain what you read in a written summary. Other strategies, such as making a study guide of key words or highlighting key words in the text, are not effective toward learning.

We also told you that only people who scored greater than 70% on the final comprehension test would receive a bonus research credit of .5. In fact, all participants received this bonus. Thank you for your hard work on the study!

Again, thank you for participating in the study. Our research could not be conducted without your help. If you should have any questions, please do not hesitate to contact:

Elaine Tan
(662) 325-5804
ewt38@msstate.edu

Deborah K. Eakin, Ph.D.
Department of Psychology
Mississippi State University
(662) 325-7949
deakin@psychology.msstate.edu

SONA Recruitment Header

Recruitment Title: Learning to Comprehend

You must be a native English speaker to participate in this experiment. You must not be under the age of 18. For this study, you will study texts using a learning strategy assigned to you. You will predict your performance on and take comprehension tests based on the texts you studied. This study will take approximately 2 hours for which you will receive 2 research credits for participating. You can earn a .5 bonus research credit during the experiment.

MSU HRPP Full Accreditation	Approved: 10.09.2015	Expires: 12.31.2016
	IRB # 15-347	

From: jroberts@orc.msstate.edu
Subject: Study 15-347: Investigating the Effects of Metacomprehension Accuracy Improvement on Learning Strategy Selection [SONA Header - Learning to Comprehend]
Date: October 9, 2015 at 1:11 PM
To: ewt38@msstate.edu
Cc: jroberts@orc.msstate.edu, jroberts@orc.msstate.edu, de115@msstate.edu, glb2@psychology.msstate.edu

J

Protocol Title: Investigating the Effects of Metacomprehension Accuracy Improvement on Learning Strategy Selection [SONA Header - Learning to Comprehend]

Protocol Number: 15-347

Principal Investigator: Ms. Elaine Tan Wei-Ling

Date of Determination: 10/9/2015

Qualifying Exempt Category: 45 CFR 46.101(b)(2)

Dear Ms. Tan Wei-Ling:

The Human Research Protection Program has determined the above referenced project exempt from IRB review.

Please note the following:

- Retain a copy of this correspondence for your records.
- An approval stamp is required on all informed consents. You must use the stamped consent form for obtaining consent from participants.
- Only the MSU staff and students named on the application are approved as MSU investigators and/or key personnel for this study.
- The approved study will expire on 12/31/2016, which was the completion date! indicated on your application. If additional time is needed, submit a continuation request. (SOP 01-07 Continuing Review of Approved Applications)
- Any modifications to the project must be reviewed and approved by the HRPP prior to implementation. Any failure to adhere to the approved protocol could result in suspension or termination of your project.
- Per university requirement, all research-related records (e.g. application materials, letters of support, signed consent forms, etc.) must be retained and available for audit for a period of at least 3 years after the research has ended.
- It is the responsibility of the investigator to promptly report events that may represent unanticipated problems involving risks to subjects or others.

This determination is issued under the Mississippi State University's OHRP Federalwide Assurance #FWA00000203. All forms and procedures can be found on the HRPP website: www! .orc.msstate.edu.

Thank you for your cooperation and good luck to you in conducting this research project. If you have questions or concerns, please contact me at jroberts@orc.msstate.edu or call 662-325-2238.

Finally, we would greatly appreciate your feedback on the HRPP approval process. Please take a few minutes to complete our survey at https://www.surveymonkey.com/s/PPM2FBP.

Sincerely,

Jodi Roberts, Ph.D.
HRPP Officer

cc: Deborah Eakin, Advisor
SONA

APPENDIX B

EXPERIMENTAL MATERIALS

Comprehension Text Passage

Digitized Signals Are the Future of the Black Box

Signals of any kind are a way to deliver a message to a destination. When digital signals transmit information, they do so by turning signals into code. This is binary code, which is very specific and easily quantified. When that code is sent via wave pulses, the transmission of the signal is very reliable.

What makes this so reliable is the fact that digital signals are actually quite resistant to outside noise disturbances. While other kinds of communication will almost always be transmitted along with some kind of undesirable noise (making a recording much harder to hear), digital signals can be encoded and sent without too much outside interference. One of today's commonly used devices made the switch from analog to digital signaling within the last 20 years. You might know it as the black box.

Many have heard of "the black box," a device used for recording what happens during an airplane's flight. What most people don't know is that the black box is really a common term for two pieces of recording equipment that are onboard every commercial and corporate airplane.

The first is called a cockpit voice recorder, or CVR. The CVR is attached to multiple microphones located in the cockpit and it records any communication and all the sounds in the cockpit. In the case of an accident, the investigators who listen to a CVR recording can actually hear two things: first, what was said by the pilots and/or crew right before the incident; and second, the sounds in the background. Well-trained investigators can detect unusual engine noise, strange pops and other signals that help alert them to figure out what went wrong with the flight.

The second part of the so-called black box is the flight data recorder, or FDR. This piece of equipment does not record the people onboard, but all technical aspects of a flight. Sensors all over the plane detect and send information to a flight data acquisition unit, which, in turn, is hooked up to the FDR. The FDR is usually attached to the plane's tail, where it's least likely to be damaged in case of an accident. In the U.S., the Federal Aviation Administration requires FDRs to record at least 88 parameters, or aspects, of a commercial flight. As a few examples, these parameters can include the time, altitude, airspeed, direction, movement of the flaps on the wings, the flow of fuel, and use of autopilot. Then, in case something happens, investigators can use this information to recreate a simulation of the entire flight, from takeoff to the incident. In conjunction with the information from the cockpit voice recorder, they can get a picture of what happened.

Making a recording of some aspect of a flight began with the beginning of flight itself. The Wright brothers, who created the first airplane, actually used a device to record their propeller rotations. (Think of it as the very first FDR, except that it only recorded a single kind of data!)

Some basic recording devices were invented and used during the 1930s and during World War II, but they weren't commonplace. It was two decades later that aviation recorders began to become more widespread. The modern day black box is credited as an invention by an Australian scientist, Dr. David Warren.

Warren came up with the idea that multiple aspects of all flights should be recorded while he was working at the Aeronautical Research Laboratory in Melbourne. He was helping investigate an accident by the world's first jet-powered commercial aircraft, the

Comet. Without any kind of recording, the crash was a total mystery to him and his co-investigators. He demonstrated the first basic flight data recorder in 1957. It was called a "red egg" for its shape and color. The red egg was fireproof and shockproof. It could reliably record both a plane's instrument readers and the pilots' voices, using only one wire. It also included a device to then decode all this information back on the ground.

The red egg wasn't put into widespread use immediately. In 1960, however, there was another unexplained plane crash in Australia; this time in Queensland. After that, Australia became the first country in the world to mandate that the device be used on all commercial aircraft.

The black box is now used on all commercial aircraft and corporate jets. It's unclear exactly where the term came from, but it's possible it came from something a journalist told Dr. Warren about his red egg. Supposedly, he said, "this is a wonderful black box." At any rate, the phrase doesn't refer to the black box's color—the equipment is actually painted bright orange, in order to make it easier to find.

The modern device is used around the world and is highly regulated. International standards mandate that it be able to withstand high acceleration and deceleration, high and low temperature fires, deep sea pressure, submersion in seawater or other liquids, and high impact and being crushed.

Beginning in the 1990s, the technology employed by the black box was greatly improved. Newer black boxes were being built with solid-state memory boards, which use memory chips to record and store information. This digital system is an improvement over the original system, magnetic tape technology, for several reasons. First off, magnetic tape needs to be pulled across an electromagnetic head. Solid-state technology, however, has no moving parts making it both more reliable as an encoder of information and less likely to break. Second, the original cockpit voice recorder could only hold about a half-hour of information. It would record in a loop, recording over every half-hour, so the last half-hour of a flight was all investigators could hear. With solid-state technology, the CVR can record up to two hours, which provides much more information. Furthermore, the flight data recorder can hold up to 25 hours using solid-state technology.

Solid-state memory boards are also better than magnetic tape technology concerning what the flight data recorder can record. While the old technology was able to record up to 100 different aspects or parameters of a flight, solid-state technology records up to 700.

What has remained the same, from one technology to the next, is the way the black box is powered. Both types draw energy from two generators, which are powered by the plane's engines.

The black box records and provides a huge amount of information. However, its technology helps determine how quickly investigators can analyze and use that information. In the case of an investigation, it can take weeks, even months, for investigators to download all the information from black boxes still using magnetic tape technology. And that's before they can even start studying and processing what happened! Using digitally equipped black boxes, however, they're able to download all the information from a flight in a matter of minutes. What a vast improvement! Black box manufacturers have made a complete switch to digital signaling from the old analog ways, and no longer make the magnetic tape recorders.

Comprehension Test Questions

Trial 1

1. **How do digital signals transmit information?**
    a. The signals are turned into binary code and sent via wave pulses
    b. The signals are converted into Morse code and sent via wave pulses
    c. The signals are turned into white noise and sent via binary pulses
    d. The signals are converted into binary waves and sent via binary pulses
2. **How is the CVR component of the black box important to a plane crash investigation?**
    a. It records unusual engine noises, strange pops, and other signals
    b. It is resistant to outside noise disturbances
    c. It records technical aspects of the flights such as time, altitude, airspeed, etc.
    d. It collects information from all over the airplane
3. **Why is the FDR typically located on the tail of the airplane?**
    a. The tail of the airplane is least likely to be damaged in a plane crash
    b. The tail of the airplane is conducive for the FDR component
    c. The tail is where the FDR collects its data during a flight
    d. The FDR is not located on the tail of the airplane
4. **Why are both the CVR and FDR components equally crucial to the black box during a plane crash investigation?**
    a. To accurately simulate what happened during the flight
    b. To analyze what the communication between the pilot and co-pilot minutes before the plane crash
    c. To calculate the precise speed of the airplane shortly before the plane crashed
    d. To detect any unusual engine noises, and strange pops during the flight
5. **Why did Dr. Warren come up with the idea of recording multiple aspects of the flight?**
    a. He was in the midst of investigating a mysterious plane crash
    b. He was in the midst of designing an upgrade of the black box
    c. He and his colleague was part of the U.S. Federal Aviation Administration team appointed to devise a better airplane recording device
    d. The recording device in the airplane that crashed only had one flight data recorded
6. **Why did Australia become the first country to mandate recording devices on all commercial aircrafts?**
    a. An unexplained plane crash had occurred in Queensland, Australia
    b. The U.S. Federal Aviation Administration selected Australia to be the first country
    c. Dr. Warren was an Australian scientist and he advocated mandating recording devices on all commercial aircrafts
    d. Australia was the only country producing airplane-recording devices

7. **What is the significance of the color of the black box?**
    a. To make it easy to locate after a plane crash
    b. It is named after the color of the device
    c. It represents the mystery it holds after a plane crash
    d. It was named after the journalist who interviewed Dr. Warren
8. **What was one problem with the older recording devices?**
    a. The magnetic tapes could get damaged when being stretched across an electromagnetic head
    b. The color of the device made it difficult to find after a plane crash
    c. The transmission could get interrupted if the black box was too close in proximity to a magnetic field
    d. The magnetic tapes did not provide clear video images
9. **What is the one aspect of the modern black box that has not changed with the advancement of its technology?**
    a. The power source remained the same
    b. The method of information transmission remained the same
    c. The method of data storage remained the same
    d. The types of flight data recorded during the flight remained the same
10. **How often are magnetic tape recorders used on modern aircrafts?**
    a. They are currently obsolete technology
    b. They are used as often as any new airplanes are built
    c. They are remodeled every four years according to the FAA standards
    d. They are only used on private jets

Manipulation check questions for Trial 1

11. **What does your heart pump?**
    a. Blood
    b. Bone marrow
    c. Leukocyte
    d. Oxygen
12. **What does “CVR” stand for?**
    a. Cockpit voice recorder
    b. Content validity receiver
    c. Current viewing recorder
    d. Computer voice response

Trial 2

1. **What is one of the benefits of using binary codes to transmit information?**
    a. It is easily quantifiable
    b. It is not as vague as analog codes
    c. It is reliable most of the time
    d. It is difficult to intercept the codes
2. **How is the FDR component of the black box important to a plane crash investigation?**
    a. It records technical aspects of the flight such as time, altitude, airspeed, etc.
    b. It records the voices of all that was onboard the airplane at the time of the crash
    c. It is the component that turns the sound waves into pulses for transmission
    d. It provides a video recording of the cockpit and in the fuselage
3. **Why does the U.S Federal Aviation Administration require FDRs to record at least 88 flight parameters?**
    a. To collect enough flight data to simulate the entire flight
    b. The smallest black box can record only 88 flight parameters
    c. These 88 flight parameters are all is needed to assist an investigation after a plane crashes
    d. It was decided based on data collected from previous plane crashes from older black boxes
4. **What decade can be accredited to the beginning of the widespread of the black box?**
    a. The 1950s
    b. The 1990s
    c. The 1930s
    d. The 1920s
5. **Why was the red egg invented?**
    a. To demonstrate the importance of recording multiple aspects of a flight
    b. To prove that a recording device should be both fire and waterproof
    c. To demonstrate the reliability of the technology used
    d. To convince the FAA to mandate recording devices on all aircrafts
6. **In this era, what determines whether an aircraft is required to have a recording device?**
    a. All aircrafts are required to have a recording device
    b. Only if it is a commercial aircraft
    c. The size and regular flight route of the aircraft
    d. Only if the amount of passengers could exceed 19
7. **How is the modern black box an improvement from those used prior to the 1950s?**
    a. The modern black box now uses the more stable solid-state memory chips
    b. The modern black box now favors the magnetic tape technology
    c. The modern black box is now powered by the plane's engines
    d. The modern black box no longer needs an external power source

8. **How sturdy are black boxes using the solid-state technology as compared to those using magnetic tape technology?**
    a. <u>There are no moving and unstable parts in solid-state technology black boxes</u>
    b. The FDR is always unstable because it is located on the tail of the airplane
    c. The solid-state technology black boxes are comparatively more stable than the magnetic tape technology black boxes but is still prone to breakage occasionally
    d. Both black box technology are equally stable
9. **What is the time frame it takes to download information from black boxes using the magnetic tape technology?**
    a. <u>It can take weeks to months</u>
    b. It can take minutes to hours
    c. It can take up to six months
    d. It can take up to a year
10. **What is the main function of a black box?**
    a. <u>To record important flight data during a flight</u>
    b. To record all audio data during a flight
    c. To record the decisions and actions of the flight crew during a flight
    d. To record all data in the flight deck of the aircraft

Manipulation check questions for Trial 2

11. **How many sides does a triangle have?**
    a. <u>Three</u>
    b. Five
    c. Seven
    d. Two

12. **What does "FDR" stand for?**
    a. <u>Flight data recorder</u>
    b. Future digital radio
    c. Frequency data radio
    d. Flight data review

Trial 3

1. **What is the main improvement of the black box in the last 20 years?**
    a. <u>The reliability of the technology used</u>
    b. The number of components in the black box
    c. The device is no longer resistant to background sound
    d. The shape of the device and its name
2. **How do the functions of the CVR differ from those of the FDR?**
    a. <u>The CVR records sounds in the cockpit whereas the FDR records technical information</u>
    b. Both the CVR and FDR records identical information as a backup in case one component malfunctions
    c. The CVR complements the FDR by recording a video in the cockpit
    d. The FDR records only 88 parameters whereas the CVR records the communications transmitted via the intercom during the flight
3. **How does the black box on the airplane record information?**
    a. <u>It uses microphones in the cockpit and sensors all over the airplane</u>
    b. It videotapes the fuselage, and the cockpit all throughout the flight
    c. It uses a close circuit television transmissions
    d. Pilots and flight attendants manually input information into a computer program linked to the black box
4. **How far back can the FDR device be traced back to?**
    a. <u>The inventors of the first airplanes</u>
    b. There is no mention of the origins of the FDR component
    c. The pilots of fighter planes in World War I
    d. Dr. Warren was the first to suggest the importance of the FDR
5. **How was the red egg an upgrade from the first airplane-recording device?**
    a. <u>The red egg recorded multiple aspects of the flight data using only a single wire</u>
    b. The red egg was both fire and waterproof
    c. The red egg was not an upgrade, it was the first recording device ever invented
    d. The shape of the red egg made it easier to attach to the fuselage
6. **Where did the name "black box" originate?**
    a. <u>Potentially from a journalist interviewing Dr. Warren</u>
    b. The color and shape of the device
    c. Dr. Warren gave the device its new name
    d. Because of the nature of what the device contains
7. **Why is the modern day black box mandated to withstand extreme environment conditions?**
    a. <u>To safeguard the data recorded during the flight after a plane crash</u>
    b. To maintain the structure integrity of the black box for the next twenty years even after a plane crash
    c. To assure that the black box is easily located after a plane crash
    d. To assist well-trained investigators after a plane crash

8. **Why is the solid-state memory technology more advantages than magnetic tape technology?**
    a. Solid-state memory technology can record 600 more technical aspects than magnetic tape technology
    b. Solid-state memory technology receives a more steady flow of power than the magnetic tape technology
    c. Solid-state memory technology can record up to 25 hours on the CVR whereas the magnetic tape technology can only record for half an hour on the CVR
    d. Solid-state memory technology can record data for as long as there is a power source whereas the magnetic tape technology does not have sufficient storage space
9. **What is the unit of measurement used to describe the time it will take to download information from black boxes using the digital technology?**
    a. Minutes
    b. Seconds
    c. Days
    d. Hours
10. **Why do digital signals work better than other kinds of communications?**
    a. Digital signals are resistant to undesirable background noises
    b. Digital signals are the future of black boxes
    c. Digital signals are capable of translating magnetic tape technology
    d. Digital signals are encrypted signals

Manipulation check questions for Trial 3

11. **What is H_20?**
    a. Water
    b. Iron
    c. Gold
    d. Ammonia
12. **What is a "Black Box"?**
    a. An aircraft recording device
    b. A mystery box with dark secrets
    c. The engine of an airplane
    d. The airplane controller box

APPENDIX C

MIXED-EFFECTS MODELING OUTPUTS FOR EXPERIMENT 1

Table C1

Full GLMM results for Hypothesis 1B on Trial Type in Experiment 1 (Table 7)

Formula: CompACC ~ TrialType + (1 | Participant) + (1 + TrialType | Question)					
Family: binomial (logit)					
Control: glmerControl(optimizer = “bobyqa”, optCtrl = list(maxfun = 2e+05))					
Number of observations = 1400; groups: Participant, 140; Question, 10					
Random Effects		Variance	*SD*	Correlation	
Participant	(Intercept)	0.78	0.88		
Question	(Intercept)	1.25	1.12		
	TT	0.14	0.37	-0.76	
Fixed Effects		Estimate	*SE*	z-Wald	*p*
(Intercept)		1.41	0.39	3.66	< .05
TrialType (TT)		-0.85	0.24	-3.57	< .01

Note. Multi as comparison group

Table C2

Full GLMM results for Hypothesis 1B on Learning Strategy in Experiment 1 (Table 7)

Formula: CompACC ~ LearningStrategy + (1 | Participant) + (1+ LearningStrategy | Question)					
Family: binomial (logit)					
Control: glmerControl(optimizer = “bobyqa”, optCtrl = list(maxfun = 2e+05))					
Number of observations = 1400; groups: Participant, 140; Question, 10					
Random Effects		Variance	*SD*	Correlation	
Participant	(Intercept)	0.87	0.93		
Question	(Intercept)	1.21	1.12		
	LS	0.07	0.37	-0.76	
Fixed Effects		Estimate	*SE*	z-Wald	*p*
(Intercept)		1.02	0.38	2.66	< .05
LearningStrategy (LS)		-0.19	0.23	-0.85	0.40

Note. Explanation as comparison group

Table C3

Full LMEM results for Trial Type on DJOL sensitivity for Trial 3 in Experiment 1 (Table 8)

Formula: DJOL ~ TrialType + (1 | Participant) + (1 + TrialType | Question) Control: lmerControl(optCtrl = list(maxfun = 2e+05)) Number of observations = 1400; groups: Participant, 140; Question, 10						
Random Effects		Variance	*SD*	Correlation		
Participant	(Intercept)	271.42	16.47			
Question	(Intercept)	34.45	5.87			
	TT	25.57	5.06	-0.36		
Residual		570.60	23.89			
Fixed Effects		Estimate	*SE*	*df*	*t*-value	*p*
(Intercept)		80.26	2.96	32	27.14	< .01
TrialType (TT)		-21.49	3.47	55	-6.19	< .01

Note. Multi as comparison group

Table C4

Full LMEM results for Learning Strategy on DJOL sensitivity for Trial 3 in Experiment 1 (Table 8)

Formula: DJOL ~ LearningStrategy + (1 | Participant) + (1 + LearningStrategy | Question) Control: lmerControl(optCtrl = list(maxfun = 2e+05)) Number of observations = 1400; groups: Participant, 140; Question, 10						
Random Effects		Variance	*SD*	Correlation		
Participant	(Intercept)	386.90	19.67			
Question	(Intercept)	59.58	7.72			
	LS	19.44	4.41	-1.00		
Residual		571.71	23.91			
Fixed Effects		Estimate	*SE*	*df*	*t*-value	*p*
(Intercept)		59.43	3.92	31	15.16	< .01
LearningStrategy (LS)		-1.32	4.49	69	-0.30	0.80

Note. Explanation as comparison group

Table C5

Full LMEM results for interaction effect of Trial Type and Learning Strategy on DJOL sensitivity for Trial 3 in Experiment 1, Multi E as comparison group (Table 8)

Formula: DJOL ~ AssignedCondition + (1 | Participant) + (1 + AssignedCondition | Question)						
Control: lmerControl(optCtrl = list(maxfun = 2e+05))						
Number of observations = 1400; groups: Participant, 140; Question, 10						
Random Effects		Variance	*SD*	Correlation		
Participant	(Intercept)	275.90	16.61			
Question	(Intercept)	60.28	7.76			
	MultiKW	32.55	5.71	-0.75		
	SingleE	9.20	3.03	-0.13	-0.31	
	SingleKW	42.40	6.51	-0.76	0.18	0.69
Residual		562.91	23.77			
Fixed Effects		Estimate	*SE*	*df*	*t*-value	*p*
(Intercept)		78.92	4.05	35	19.48	< .01
Multi-KW		2.78	4.97	71	0.56	0.60
Single-E		-19.49	4.45	95	-4.37	< .01
Single-KW		-20.81	4.81	65	-4.33	<. 01

Note. Multi Explanation as comparison group

Table C6

Full LMEM results for interaction effect of Trial Type and Learning Strategy on DJOL sensitivity for Trial 3 in Experiment 1, Multi KW as comparison group (Table 8)

Formula: DJOL ~ AssignedCondition + (1 | Participant) + (1 + AssignedCondition | Question) Control: lmerControl(optCtrl = list(maxfun = 2e+05)) Number of observations = 1400; groups: Participant, 140; Question, 10						
Random Effects		Variance	*SD*	Correlation		
Participant	(Intercept)	275.90	16.61			
Question	(Intercept)	26.50.	5.15			
	MultiE	32.55	5.71	0.02		
	SingleE	52.32	7.23	-0.21	0.92	
	SingleKW	61.52	7.84	-0.77	0.58	0.79
Residual		562.91	23.73			
Fixed Effects		Estimate	*SE*	*df*	*t*-value	*p*
(Intercept)		81.69	3.37	61	22.05	< .01
Single-E		-19.49	4.45	95	-4.37	< .01
Single-KW		-20.81	4.81	65	-4.33	<. 01

Note. Multi Keyword as comparison group

Table C7

Full LMEM results for interaction effect of Trial Type and Learning Strategy on DJOL sensitivity for Trial 3 in Experiment 1, Single E as comparison group (Table 8)

Formula: DJOL ~ AssignedCondition + (1 | Participant) + (1 + AssignedCondition | Question) Control: lmerControl(optCtrl = list(maxfun = 2e+05)) Number of observations = 1400; groups: Participant, 140; Question, 10						
Random Effects		Variance	*SD*	Correlation		
Participant	(Intercept)	304.52	17.45			
Question	(Intercept)	57.18	7.56			
	SingleKW	8.95	3.00	-1.00		
Residual		718.80	26.81			
Fixed Effects		Estimate	*SE*	*df*	*t*-value	*p*
(Intercept)		58.11	3.34	55	17.41	< .01
Single-KW		1.32	4.41	85	0.30	0.76

Note. Single Explanation as comparison group

Table C8

Full GLMM results for Hypothesis 1A on Trial Type in Experiment 1 (Table 9)

Formula: CompACC ~ DJOL_z * TrialType + (1 + DJOL_z | Participant) +
(1 + DJOL_z * TrialType | Question)
Family: binomial (logit)
Control: glmerControl(optimizer = “bobyqa”, optCtrl = list(maxfun = 2e+05))
Number of observations = 1400; groups: Participant, 140; Question, 10

Random Effects		Variance	*SD*	Correlation		
Participant	(Intercept)	0.79	0.88			
	DJOL_z	0.00	0.00	-1.00		
Question	(Intercept)	1.21	1.10			
	DJOL_z	0.04	0.20	0.89		
	TT	0.12	0.35	-0.75	-0.84	
	DJOLz:TT	0.02	0.13	-0.96	-0.73	0.58

Fixed Effects	Estimate	*SE*	z-Wald	*p*
(Intercept)	1.41	0.39	3.66	< .01
DJOL_z	0.32	0.15	2.17	0.03
TrialType (TT)	-0.85	0.23	-3.57	< .01
DJOLz:TrialType	-0.08	0.16	-0.48	0.63

Note. Multi as comparison group

Table C9

Full GLMM results for Hypothesis 1C on Learning Strategy in Experiment 1 (Table 9)

Formula: CompACC ~ DJOL_z * LearningStrategy+ (1 + DJOL_z | Participant) + (1 + DJOL_z * LearningStrategy | Question)
Family: binomial (logit)
Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 2e+05))
Number of observations = 1400; groups: Participant, 140; Question, 10

Random Effects		Variance	*SD*	Correlation		
Participant	(Intercept)	0.91	0.95			
	DJOL_z	0.00	0.00	-1.00		
Question	(Intercept)	1.17	1.08			
	DJOL_z	0.07	0.26	0.24		
	LS	0.08	0.28	-0.89	-0.65	
	DJOLz:LS	0.02	0.15	0.48	-0.74	-0.03
Fixed Effects		**Estimate**	***SE***	**z-Wald**	***p***	
(Intercept)		1.02	0.38	2.71	< .01	
DJOL_z		0.43	0.14	3.00	< .01	
LearningStrategy (LS)		-0.19	0.23	-0.83	0.41	
DJOLz:LearningStrategy		-0.33	0.16	-2.10	0.04	

Note. Explanation as comparison group

Table C10

Full GLMM results for Hypothesis 1D on interaction effect of Trial Type and Learning Strategy in Experiment 1, Multi E as comparison group (Table 9)

Formula: CompACC ~ DJOL_z * AssignedCondition + (1 + DJOL_z | Participant) + (1 + DJOL_z * AssignedCondition | Question)
Family: binomial (logit)
Control: glmerControl(optimizer = “bobyqa”, optCtrl = list(maxfun = 2e+05))
Number of observations = 1400; groups: Participant, 140; Question, 10

Random Eff.	Var.	*SD*	Correlation						
Participant(Int)	0.81	0.91							
DJOL_z	0.00	0.02	1.00						
Question(Int)	1.91	1.38							
DJOL_z	0.14	0.37	0.35						
M-KW	0.31	0.55	-0.99	-0.43					
S-E	0.37	0.61	-0.86	-0.35	0.90				
S-KW	0.47	0.68	-0.86	-0.59	0.91	0.96			
DJOLz:M-KW	0.13	0.36	0.43	-0.70	-0.34	-0.29	-0.07		
DJOLz:S-E	0.04	0.20	-0.38	-0.76	0.38	0.03	0.27	0.41	
DJOLz:S-KW	0.05	0.23	-0.38	-0.79	0.50	0.67	0.79	0.50	0.23
Fixed Effects		Estimate		*Std.Error*		*z*-Wald		*p*	
(intercept)		1.49		0.50		2.98		< .01	
DJOL_z		0.52		0.25		2.14		0.03	
Multi-KW		-0.13		0.37		-0.35		0.72	
Single-E		-0.78		0.36		-2.17		0.03	
Single-KW		-1.04		0.37		-2.78		0.01	
DJOLz:M-KW		-0.38		0.29		-1.29		0.20	
DJOLz:S-E		-0.16		0.26		-0.62		0.54	
DJOLz:S-KW		-0.44		0.26		-1.70		0.09	

Note. Multi Explanation as comparison group

Table C11

Full GLMM results for Hypothesis 1D on interaction effect of Trial Type and Learning Strategy in Experiment 1, Single E as comparison group (Table 9)

Formula: CompACC ~ DJOL_z * AssignedCondition + (1 + DJOL_z | Participant) + (1 + DJOL_z * AssignedCondition | Question)
Family: binomial (logit)
Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 2e+05))
Number of observations = 1400; groups: Participant, 140; Question, 10

Random Eff.	Var.	*SD*	Correlation						
Participant(Int)	0.81	0.91							
DJOL_z	0.00	0.02	1.00						
Question(Int)	0.84	0.92							
DJOL_z	0.06	0.25	-0.02						
M-KW	0.07	0.26	-0.45	0.44					
M-E	0.37	0.61	0.63	0.48	0.40				
S-KW	0.04	0.19	-0.33	-0.77	0.24	-0.27			
DJOLz:M-KW	0.11	0.34	0.82	-0.56	-0.48	0.34	0.24		
DJOLz:M-E	0.04	0.20	0.54	-0.30	-0.72	0.03	-0.84	0.5	
DJOLz:S-KW	0.07	0.27	0.30	-0.61	-0.97	-0.55	-0.03	0.46	0.55

Fixed Effects	Estimate	*Std.Error*	*z*-Wald	*p*
(intercept)	0.71	0.35	2.03	0.04
DJOL_z	0.36	0.16	2.29	0.02
Multi-KW	0.65	0.31	2.09	0.04
Multi-E	0.78	0.36	2.17	0.03
Single-KW	-0.25	0.28	-0.92	0.36
DJOLz:M-KW	-0.22	0.24	-0.90	0.37
DJOLz:M-E	0.16	0.26	0.62	0.54
DJOLz:S-KW	-0.28	0.21	-1.33	0.18

Note. Single Explanation as comparison group

Table C12

Full LMEM results for Trial Type on CJ sensitivity for Trial 3 in Experiment 1 (Table 10)

Formula: CJ ~ TrialType + (1 | Participant) + (1 + TrialType | Question) Control: lmerControl(optCtrl = list(maxfun = 2e+05)) Number of observations = 1400; groups: Participant, 140; Question, 10						
Random Effects		Variance	*SD*	Correlation		
Participant	(Intercept)	172.83.42	13.15			
Question	(Intercept)	23.23	4.82			
	TT	7.55	2.75	-0.34		
Residual		461.51	21.48			
Fixed Effects		Estimate	*SE*	*df*	*t*-value	*p*
(Intercept)		86.74	2.42	30	35.85	< .01
TrialType (TT)		-17.05	2.66	64	-6.40	< .01

Note. Multi as comparison group

Table C13

Full LMEM results for Learning Strategy on CJ sensitivity for Trial 3 in Experiment 1 (Table 10)

Formula: CJ ~ LearningStrategy + (1 | Participant) + (1 + LearningStrategy | Question) Control: lmerControl(optCtrl = list(maxfun = 2e+05)) Number of observations = 1400; groups: Participant, 140; Question, 10						
Random Effects		Variance	*SD*	Correlation		
Participant	(Intercept)	245.82	15.68			
Question	(Intercept)	41.03	6.41			
	LS	21.83	4.67	-0.87		
Residual		457.89	21.30			
Fixed Effects		Estimate	*SE*	*df*	*t*-value	*p*
(Intercept)		77.64	2.87	25	27.10	< .01
LearningStrategy (LS)		-0.81	3.24	59	-0.25	0.80

Note. Explanation as comparison group

Table C14

Full LMEM results for interaction effect of Trial Type and Learning Strategy on CJ sensitivity for Trial 3 in Experiment 1, Multi E as comparison group (Table 10)

Formula: CJ ~ AssignedCondition + (1 | Participant) + (1 + AssignedCondition | Question)						
Control: lmerControl(optCtrl = list(maxfun = 2e+05))						
Number of observations = 1400; groups: Participant, 140; Question, 10						
Random Effects		Variance	*SD*	Correlation		
Participant	(Intercept)	174.70	13.22			
Question	(Intercept)	35.37	5.95			
	SingleE	6.37	2.52	0.23		
	MultiKW	6.50	2.55	-0.87	0.27	
	SingleKW	48.28	6.95	-0.82	0.37	0.99
Residual		454.50	21.32			
Fixed Effects		Estimate	*SE*	*df*	*t*-value	*p*
(Intercept)		85.50	3.23	36	26.49	< .01
Single-E		-14.31	3.63	103	-3.94	< .01
Multi-KW		2.57	3.86	105	0.67	0.51
Single-KW		-17.32	4.17	41	-4.16	<. 01

Note. Multi Explanation as comparison group

Table C15

Full LMEM results for interaction effect of Trial Type and Learning Strategy on CJ sensitivity for Trial 3 in Experiment 1, Multi KW as comparison group (Table 10)

Formula: CJ ~ AssignedCondition + (1 | Participant) + (1 + AssignedCondition | Question)						
Control: lmerControl(optCtrl = list(maxfun = 2e+05))						
Number of observations = 1400; groups: Participant, 140; Question, 10						
Random Effects		Variance	*SD*	Correlation		
Participant	(Intercept)	134.06	11.58			
Question	(Intercept)	15.83	3.98			
	MultiE	4.55	2.13	0.80		
	SingleE	9.50	3.08	0.94	0.96	
Residual		421.70	20.54			
Fixed Effects		Estimate	*SE*	*df*	*t*-value	*p*
(Intercept)		88.07	2.73	42	32.25	< .01
Multi-E		-2.57	3.44	57	-0.75	0.46
Single-E		-16.88	3.37	64.04	-5.01	<. 01

Note. Multi Keyword as comparison group

Table C16

Full LMEM results for interaction effect of Trial Type and Learning Strategy on CJ sensitivity for Trial 3 in Experiment 1, Single E as comparison group (Table 10)

Formula: CJ ~ AssignedCondition + (1 | Participant) + (1 + AssignedCondition | Question)						
Control: lmerControl(optCtrl = list(maxfun = 2e+05))						
Number of observations = 1400; groups: Participant, 140; Question, 10						
Random Effects		Variance	*SD*	Correlation		
Participant	(Intercept)	174.70	13.22			
Question	(Intercept)	48.73	6.98			
	MultiE	6.37	2.52	-0.56		
	SingleKW	41.70	6.46	-0.83	-0.01	
	MultiKW	9.35	3.06	-1.00	0.60	0.80
Residual		454.50	21.32			
Fixed Effects		Estimate	*SE*	*df*	*t*-value	*p*
(Intercept)		71.20	3.24	26	21.95	< .01
Multi-E		14.31	3.63	103	3.94	< .01
Single-KW		-3.01	3.93	44	-0.77	0.45
Multi-KW		16.88	3.73	98	4.52	< .01

Note. Single Explanation as comparison group

Table C17

Full GLMM results for Trial Type on CJ accuracy in Experiment 1 (Table 11)

Formula: CompACC ~ CJ_z * TrialType + (1 + CJ_z | Participant) + (1 + CJ_z * TrialType | Question)						
Family: binomial (logit)						
Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 2e+05))						
Number of observations = 1400; groups: Participant, 140; Question, 10						
Random Effects		Variance	*SD*	Correlation		
Participant	(Intercept)	0.86	0.93			
	CJ_z	0.00	0.00	1.00		
Question	(Intercept)	1.26	1.12			
	CJ_z	0.11	0.33	0.73		
	TT	0.18	0.42	-0.72	-1.00	
	CJz:TT	0.03	0.17	0.74	0.08	-0.07
Fixed Effects		Estimate	*SE*	z-Wald	*p*	
(Intercept)		1.49	0.40	3.76	< .01	
CJ_z		0.55	0.17	3.36	< .01	
TrialType (TT)		-0.88	0.26	-3.45	< .01	
CJz:TrialType		0.14	0.16	0.83	0.41	

Note. Multi as comparison group

Table C18

Full GLMM results for Learning Strategy on CJ accuracy in Experiment 1 (Table 11)

Formula: CompACC ~ CJ_z * LearningStrategy+ (1 + CJ_z | Participant) + (1 + CJ_z * LearningStrategy | Question)						
Family: binomial (logit)						
Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 2e+05))						
Number of observations = 1400; groups: Participant, 140; Question, 10						
Random Effects		Variance	*SD*	Correlation		
Participant	(Intercept)	0.95	0.97			
	CJ_z	0.00	0.00	-1.00		
Question	(Intercept)	1.22	1.10			
	CJ_z	0.22	0.47	0.74		
	LS	0.01	0.32	-0.93	-0.94	
	CJz:LS	0.01	0.30	-0.07	-0.72	0.44
Fixed Effects		Estimate	*SE*	z-Wald	*p*	
(Intercept)		1.05	0.38	2.73	< .01	
CJ_z		0.66	0.19	3.54	< .01	
LearningStrategy (LS)		-0.18	0.24	-0.76	0.45	
CJz:LearningStrategy		-0.03	0.18	-0.16	0.87	

Note. Explanation as comparison group

Table C19

Full GLMM results for interaction effect of Trial Type and Learning Strategy on CJ accuracy in Experiment 1, Multi E as comparison group (Table 11)

Formula: CompACC ~ CJ_z * AssignedCondition + (1 + CJ_z | Participant) + (1 + CJ_z * AssignedCondition | Question)
Family: binomial (logit)
Control: glmerControl(optimizer = “bobyqa”, optCtrl = list(maxfun = 2e+05))
Number of observations = 1400; groups: Participant, 140; Question, 10

Random Eff.	Var.	*SD*	Correlation						
Participant(Int)	0.85	0.92							
CJ_z	0.00	0.01	-1.00						
Question(Int)	2.04	1.43							
CJ_z	0.34	0.58	0.58						
M-KW	0.33	0.58	-0.99	-0.64					
S-E	0.43	0.66	-0.89	-0.75	0.95				
S-KW	0.58	0.76	-0.89	-0.83	0.94	0.99			
CJz:M-KW	0.20	0.45	-0.31	-0.95	0.40	0.58	0.66		
CJz:S-E	0.10	0.32	-0.04	-0.83	0.11	0.26	0.37	0.93	
CJz:S-KW	0.19	0.44	0.16	-0.58	-0.03	0.28	0.30	0.78	0.71

Fixed Effects	Estimate	*SE*	*z*-Wald	*p*
(intercept)	1.56	0.51	3.02	< .01
CJ_z	0.56	0.27	2.07	0.04
Multi-KW	-0.15	0.38	-0.39	0.70
Single-E	-0.85	0.37	-2.27	0.02
Single-KW	-1.08	0.39	-2.75	< .01
CJz:M-KW	0.03	0.29	0.10	0.92
CJz:S-E	0.17	0.26	0.65	0.51
CJz:S-KW	0.10	0.28	0.36	0.72

Note. Multi Explanation as comparison group

Table C20

Full GLMM results for interaction effect of Trial Type and Learning Strategy on CJ accuracy in Experiment 1, Multi KW as comparison group (Table 11)

Formula: CompACC ~ CJ_z * AssignedCondition + (1 + CJ_z | Participant) +
(1 + CJ_z * AssignedCondition | Question)
Family: binomial (logit)
Control: glmerControl(optimizer = “bobyqa”, optCtrl = list(maxfun = 2e+05))
Number of observations = 1400; groups: Participant, 140; Question, 10

Random Eff.	Var.	*SD*	Correlation						
Participant(Int)	0.85	0.92							
CJ_z	0.00	0.01	-1.00						
Question(Int)	0.74	0.86							
CJ_z	0.04	0.21	0.94						
M-E	0.33	0.58	0.97	0.94					
S-E	0.05	0.22	0.03	-0.10	-0.20				
S-KW	0.08	0.29	-0.29	-0.49	-0.48	0.88			
CJz:M-E	0.20	0.45	0.25	0.52	0.40	-0.71	-0.94		
CJz:S-E	0.04	0.19	0.58	0.68	0.75	-0.79	-0.92	0.79	
CJz:S-KW	0.09	0.30	0.74	0.85	0.65	0.32	-0.16	0.36	0.25
Fixed Effects		Estimate		*SE*		*z*-Wald		*p*	
(intercept)		1.41		0.36		3.89		< .01	
CJ_z		0.59		0.18		3.31		< .01	
Multi-E		0.15		0.38		0.39		0.70	
Single-E		-0.70		0.31		-2.24		0.03	
Single-KW		-0.93		0.32		-2.93		< .01	
CJz:M-E		-0.03		0.29		-0.10		0.92	
CJz:S-E		0.14		0.22		0.62		0.53	
CJz:S-KW		0.07		0.23		0.30		0.77	

Note. Multi Keyword as comparison group

Table C21

Full GLMM results for interaction effect of Trial Type and Learning Strategy on CJ accuracy in Experiment 1, Single E as comparison group (Table 11)

Formula: CompACC ~ CJ_z * AssignedCondition + (1 + CJ_z | Participant) + (1 + CJ_z * AssignedCondition | Question)
Family: binomial (logit)
Control: glmerControl(optimizer = “bobyqa”, optCtrl = list(maxfun = 2e+05))
Number of observations = 1400; groups: Participant, 140; Question, 10

Random Eff.	Var.	*SD*	Correlation						
Participant(Int)	0.85	0.92							
CJ_z	0.00	0.01	-1.00						
Question(Int)	0.80	0.89							
CJ_z	0.13	0.37	0.70						
S-KW	0.02	0.14	-0.54	-0.82					
M-KW	0.05	0.22	-0.27	0.47	-0.24				
M-E	0.43	0.66	0.69	0.97	-0.66	0.51			
CJz:S-KW	0.10	0.31	0.53	0.02	-0.39	-0.79	-0.13		
CJz:M-KW	0.04	0.19	-0.37	-0.91	0.64	-0.79	-0.92	0.38	
CJz:M-E	0.10	0.32	-0.13	0.44	-0.76	0.52	0.26	0.03	-0.50

Fixed Effects	Estimate	*SE*	*z*-Wald	*p*
(intercept)	0.71	0.34	2.05	0.04
CJ_z	0.72	0.18	4.00	< .01
Single-KW	-0.23	0.28	-0.83	0.40
Multi-KW	0.70	0.31	2.24	0.03
Multi-E	0.85	0.37	2.27	0.02
CJz:S-KW	-0.07	0.22	-0.31	0.76
CJz:M-KW	-0.14	0.22	-0.62	0.53
CJz:M-E	-0.17	0.26	-0.66	0.51

Note. Single Explanation as comparison group

Table C22

Full GLMM results for Trial on metacomprehension accuracy in Experiment 1, Trial 2 as comparison group (Table 12)

Formula: CompACC ~ DJOL_z * Trial + (1 + Trial * DJOL_z | Participant) + (1 + DJOL_z | Question) Family: binomial (logit) Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 2e+05)) Number of observations = 1400; groups: Participant, 140; Question, 10							
Random Eff.	Var.	*SD*	Correlation				
Participant(in)	0.64	0.80					
DJOL_z	0.07	0.26	0.72				
Trial3	0.07	0.26	0.48	0.72			
Trial1	0.04	0.20	0.81	0.98	0.78		
DJOLz:T3	0.11	0.34	-0.65	-0.98	-0.61	-0.93	
DJOLz:T1	0.04	0.21	-0.51	-0.66	-0.99	-0.75	0.52
Question(Int)	0.25	0.50					
DJOL_z	0.04	0.19	1.00				

Fixed Effects	Estimate	*SE*	*z*-Wald	*p*
(intercept)	1.11	0.21	5.19	< .01
DJOL_z	0.32	0.12	2.57	0.01
Trial3	0.13	0.15	0.84	0.40
Trial1	0.24	0.16	1.58	0.12
DJOLz:T3	-0.06	0.16	-0.39	0.70
DJOLz:T1	-0.06	0.16	-0.36	0.72

Note. Trial 2 as comparison group

Table C23

Full GLMM results of Trial on metacomprehension accuracy for explanation in Experiment 1, Multi E at Trial 1 as comparison group (Table 12)

Formula: CompACC ~ DJOL_z * Trial + (1 + Trial*DJOL_z | Participant) + (1 + DJOL_z | Question) Family: binomial (logit) Control: glmerControl(optimizer = “bobyqa”, optCtrl = list(maxfun = 2e+05)) Number of observations = 1400; groups: Participant, 140; Question, 10						
Random Eff.	Var.	*SD*	Correlation			
Participant(Int)	1.03	1.02				
DJOL_z	0.16	0.40	0.06			
Trial2	0.17	0.41	-0.87	-0.51		
Trial3	0.02	0.13	-0.27	0.34	0.23	
DJOLz:T2	0.01	0.12	0.33	-0.83	0.03	-0.76
DJOLz:T3	0.26	0.51	-0.14	-0.99	0.54	-0.44
Question(Int)	0.32	0.56				
DJOL_z	0.11	0.33	1.00			

Fixed Effects	Estimate	*SE*	*z*-Wald	*p*
(intercept)	1.36	0.30	4.48	< .01
DJOL_z	0.48	0.22	2.22	< .01
Trial2	-0.32	0.23	-1.41	0.16
Trial3	-0.11	0.22	-0.50	0.62
DJOLz:T2	-0.41	0.24	-1.74	0.08
DJOLz:T3	0.01	0.25	0.03	0.98

Note. Explanation at Trial 1 as comparison group

Table C24

Full GLMM results on interaction effect of Trial and Learning Strategy on metacomprehension accuracy in Experiment 1, Multi KW at Trial 2 as comparison group (Table 12)

Formula: CompACC ~ DJOL_z * Trial *LearningStrategy +
(1 + Trial * DJOL_z | Participant) + (1 + DJOL_z * LearningStrategy | Question)
Family: binomial (logit)
Control: glmerControl(optimizer = “bobyqa”, optCtrl = list(maxfun = 2e+05))
Number of observations = 1400; groups: Participant, 140; Question, 10

Random Eff.	Var.	*SD*	Correlation				
Participant(Int)	0.65	0.81					
DJOL_z	0.06	0.25	0.64				
Trial1	0.05	0.21	0.72	0.99			
Trial3	0.08	0.28	0.44	0.70	0.78		
DJOLz:T1	0.03	0.17	-0.33	-0.42	-0.54	-0.94	
DJOLz:T3	0.09	0.17	-0.83	-0.98	-0.94	-0.55	
Question(Int)	0.19	0.43					
DJOL_z	0.00	0.07	0.36				
LS_E	0.03	0.17	0.81	0.84			
DJOLz:LS_E	0.11	0.33	1.00	0.28	0.75	0.24	
Fixed Effects		Estimate		*SE*		*z*-Wald	*p*
(intercept)		1.19		0.25		4.70	< .01
DJOL_z		0.52		0.16		3.23	< .01
Trial1		0.21		0.22		0.95	0.34
Trial3		0.07		0.22		0.32	0.75
LS_E		-0.13		0.29		-0.44	0.66
DJOLz:T1		-0.51		0.22		-2.27	0.02
DJOLz:T3		-0.50		0.23		-2.17	0.03
DJOLz:LS_E		-0.41		0.24		-1.71	0.09
T1:LS_E		0.06		0.29		0.22	0.83
DJOLz:T1:E		0.89		0.30		2.93	< .01
DJOLz:T3:E		0.88		0.32		2.81	< .01

Note. Keyword at Trial 2 as comparison group

APPENDIX D

EXPERIMENT 1: POST-EXPERIMENT INTERVIEW ANALYSIS

The following analyses were conducted for each question in the post-experiment interview and will be discussed in the order of presentation to the participants. All analyses conducted were subjected to a two-factor analysis with two levels of Trial Type (multi, single) and two levels of Learning Strategy (explanation, keyword).

Question 1. Were you aware of the experimental manipulation in the experiment? A count of “Yes” and “No” responses was tabulated; frequency of each answer option is presented in Table D1. A Chi-square test of independence was performed to examine the relation between the assigned conditions and participant’s awareness of the manipulation of the experiment. The difference between these variables was significant, X^2 (3, N = 140) = 7.91, $p < .05$. A higher proportion of participants, regardless of the condition they were assigned to especially participants in the Single Effective condition, reported not noticing any experimental manipulation during the experiment. Amongst the participants who reported that they were aware of the experimental manipulation (n = 40), 90% of participants of those participants reported being aware of the distractor tasks (i.e., Tower of Hanoi) prior to taking the comprehension test, or the fact that they read the same texts and repeated tasks multiple times across trials. Only one participant reported being aware that they were learning to be more aware of their studying methods via multiple trials. That one response was the closest answer to the actual manipulation of this experiment. However, upon further inspection, that participant did not show any improvement in metacomprehension nor comprehension accuracy across trials.

Table D1

Frequency of answer options for Question 1 by Trial Type and Learning Strategy in Experiment 1

	Explanation		Keyword	
	Yes	No	Yes	No
Multi	12	18	12	20
Single	5	34	11	28

Question 2. How did you think the learning strategy assigned affected your ability to predict memory performance across trials? A count for "Helped", "Hindered", and "Did not affect" ability responses was tabulated; frequency for each answer option is presented in Table D2. A Chi-square test of independence was performed to examine the relation between the assigned conditions and participant's self-report of whether the learning strategy assigned affected their ability to predict memory performances. The differences in count amongst these responses were insignificant, X^2 (6, N = 140) = 9.41, $p > .05$. The proportion of participants who reported that the learning strategy assigned helped, hindered, or did not affect their ability to predict memory performance was similar between the multi and single conditions. Participants were not aware of the effect of the learning strategy on their ability to make predictions on their comprehension test performance.

Table D2

Frequency of answer options for Question 2 by Trial Type and Learning Strategy in Experiment 1

	Explanation			Keyword		
	Helped	Hindered	Did not affect	Helped	Hindered	Did not affect
Multi	17	3	12	22	1	7
Single	21	9	9	19	7	13

Question 3. How did you think the learning strategy assigned affected your test performance across trials? A count for "Increased", "Decreased", and "Did not affect" performance responses was tabulated; frequency for each answer option is presented in Table D3. A Chi-square test of independence was performed to examine the relation between the assigned conditions and participant's self-report of whether the learning strategy assigned affected their comprehension test performance. The differences in count amongst these responses were insignificant, X^2 (6, N = 140) = 11.14, $p > .05$. There was no difference between the proportion of participants who reported whether the learning strategy increased, decreased, or did not affect their comprehension test performance regardless of the conditions they were assigned to. Participants were not aware of the effect of the learning strategy on their comprehension performance at a future test.

Table D3

Frequency of answer options for Question 3 by Trial Type and Learning Strategy in Experiment 1

	Explanation			Keyword		
	Increased	Decreased	Not affected	Increased	Decreased	Not affected
Multi	22	1	9	22	1	8
Single	20	10	9	23	6	10

Question 4. Rate the difficulty of the test questions using a 0 (not difficult at all) to 10 (extremely difficult). Because this question used a rating scale of 0 – 10 instead of nominal categorical responses, a two-way Analysis of Variance (ANOVA) was conducted on the difficulty ratings made. A two-way ANOVA yielded a significant main effect of Trial Type for the difficulty ratings, $F(1, 136) = 7.32$, $p < .05$, partial $\eta^2 = .05$. The average difficulty ratings were significantly higher for participants who had a single trial ($M = 5.21$, $SE = .27$) than those who had multiple trials ($M = 4.11$, $SE = .30$). The main effect of Learning Strategy was not significant, $F(1, 136) = 1.46$, $p > .05$, partial $\eta^2 = .01$. Participants who were assigned to the explanation learning strategy ($M = 4.42$, $SE = .28$) gave similar difficulty ratings than those assigned to the keyword learning strategy ($M = 4.91$, $SE = .29$). Additionally, there was no significant interaction effect, $F(1, 136) = 1.46$, $p > .05$, partial $\eta^2 = .01$. Participants who experienced only a single trial rated the test questions more difficult than participants who experienced multiple trials. Additionally, learning strategy assigned did not have an impact on difficulty ratings on test questions.

Question 5. Rate the effectiveness of the learning strategy assigned using a 0 (not effective at all) to 10 (extremely effective). This question used a 10-point rating scale; participants rated the effectiveness of the learning strategy assigned using a scale of 0 (not effective at all) – 10 (extremely effective). A two-way ANOVA was conducted on the effectiveness ratings made. The main effect of Trial Type was not significant, $F(1, 136) = 2.28$, $p > .05$, partial $\eta^2 = .02$. There was also no main effect of Learning Strategy, $F(1, 136) = 2.52$, $p > .05$, partial $\eta^2 = .02$. Participants rated the effectiveness of the learning strategy they were assigned to; those assigned to explanation rated the learning strategy ($M = 6.80$, $SE = .25$) equally as high as those assigned to keyword learning strategy ($M = 6.24$, $SE = .25$). There was also not a significant interaction, $F(1, 136) = 2.13$, $p > .05$, partial $\eta^2 = .02$. There was no difference in learning strategy effectiveness ratings when participants experienced explanation multiple times ($M = 6.81$, $SE = .37$) and when participants experienced keyword multiple times ($M = 6.77$, $SE = .38$). The learning strategy assigned or the number of trials participants experienced did not impact their ratings on the effectiveness of the learning strategies.

Question 6. Would you recommend the learning strategy assigned to a friend who wants to do well? A count for "Yes" and "No" responses was tabulated; frequency for each answer option is presented in Table D4. A Chi-square test of independence was performed to examine the difference in a participant's inclination to recommend the learning strategy assigned to a friend who would like to do well in a test between the two learning strategies. The difference between these variables was not significant, X^2 (3, N = 129) = 1.50, $p > .05$. There was no difference in the proportion of participants who would recommend and not recommend the learning strategy they were assigned to between the multi and single trial conditions.

Table D4

Frequency of answer options for Question 6 by Trial Type and Learning Strategy in Experiment 1

	Explanation		Keyword	
	Yes	No	Yes	No
Multi	24	8	23	7
Single	26	13	22	6

Question 7. Was the learning strategy assigned worth the effort to improve test performance? A count for “Yes”, “No”, and “Maybe” responses was tabulated; frequency for each answer option is presented in Table D5. A Chi-square test of independence was performed to examine the differences in a participant’s self-report of whether the learning strategy assigned was worth the effort to improve their test performance between the two learning strategies. The difference in count between the learning strategies was not significant, X^2 (6, N = 140) = 5.54, $p > .05$. There was no difference in the proportion of participants who reported whether the learning strategy was worth the effort to improve comprehension test performance between the multi and single trial conditions.

Table D5

Frequency of answer options for Question 7 by Trial Type and Learning Strategy in Experiment 1

	Explanation			Keyword		
	Yes	No	Maybe	Yes	No	Maybe
Multi	18	2	12	20	3	7
Single	21	4	14	15	6	18

As a conclusion on the findings of participants' responses to all the post-experiment interview questions, it can be inferred that participants were not aware of the manipulation of the experiment in any way. Additionally, the proportion of responses for each answer option is similar between both multi and single conditions. It can be concluded that participants' responses were not in any way a reaction to the awareness of the experimental manipulation.

APPENDIX E

MIXED-EFFECTS MODELING OUTPUTS FOR EXPERIMENT 2

Table E1

Full GLMM results for Hypothesis 2C on Learning Strategy at Trial 3 in Experiment 2 (Table 20)

Formula: CompACC ~ LearningStrategy + (1 | Participant) + (1 + LearningStrategy | Question)					
Family: binomial (logit)					
Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 2e+05))					
Number of observations = 1690; groups: Participant, 169; Question, 10					
Random Effects		Variance	*SD*	Correlation	
Participant	(Intercept)	1.50	1.22		
Question	(Intercept)	1.52	1.23		
	LS	0.00	0.00	-1.00	
Fixed Effects		Estimate	*SE*	z-Wald	*p*
(Intercept)		1.35	0.48	2.79	< .01
LearningStrategy (LS)		0.22	0.31	0.69	0.49

Note. Explanation as comparison group

Table E2

Full GLMM results for Hypothesis 2C on Learning Strategy at Trial 1 in Experiment 2 (Table 20)

Formula: CompACC ~ LearningStrategy + (1 | Participant) + (1 + LearningStrategy | Question)					
Family: binomial (logit)					
Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 2e+05))					
Number of observations = 1310; groups: Participant, 131; Question, 10					
Random Effects		Variance	*SD*	Correlation	
Participant	(Intercept)	0.67	0.82		
Question	(Intercept)	0.19	0.44		
	LS	0.03	0.18	0.14	
Fixed Effects		Estimate	*SE*	z-Wald	*p*
(Intercept)		0.92	0.19	4.70	< .01
LearningStrategy (LS)		0.36	0.21	1.77	0.08

Note. Explanation as comparison group

Table E3

Full GLMM results for Hypothesis 2C on Trial Type at Trial 3 in Experiment 2 (Table 20)

Formula: CompACC ~ TrialType + (1 | Participant) + (1 + TrialType | Question)					
Family: binomial (logit)					
Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 2e+05))					
Number of observations = 1690; groups: Participant, 169; Question, 10					
Random Effects		Variance	*SD*	Correlation	
Participant	(Intercept)	1.41	1.19		
Question	(Intercept)	1.63	1.28		
	TT	0.05	0.23	-0.70	
Fixed Effects		Estimate	*SE*	z-Wald	*p*
(Intercept)		1.75	0.43	4.09	< .01
TrialType (TT)		-0.92	0.28	-3.27	< .01

Note. Multi as comparison group

Table E4

Full GLMM results for Hypothesis 2C on interaction effect of Trial Type and Learning Strategy at Trial 3, Multi E as comparison group (Table 20)

Formula: CompACC ~ LearningStrategy + (1 | Participant) + (1 + LearningStrategy | Question)						
Family: binomial (logit)						
Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 2e+05))						
Number of observations = 1690; groups: Participant, 169; Question, 10						
Random Effects		Variance	*SD*	Correlation		
Participant	(Intercept)	1.39	1.18			
Question	(Intercept)	1.64	1.28			
	MultiKW	0.08	0.28	-0.14		
	SingleE	0.07	0.26	0.12	-1.00	
	SingleKW	0.02	0.16	-0.94	-0.20	0.22
Fixed Effects		Estimate	*SE*	z-Wald	*p*	
(Intercept)		1.65	0.52	3.19	< .01	
Multi-KW		0.13	0.36	0.35	0.73	
Single-E		-1.50	0.72	-2.12	0.03	
Single-KW		-0.70	0.41	-1.72	0.08	

Note. Multi Explanation as comparison group

Table E5

Full GLMM results for Hypothesis 2C on interaction effect of Trial Type and Learning Strategy at Trial 3, Multi KW as comparison group (Table 20)

Formula: CompACC ~ LearningStrategy + (1 | Participant) + (1 + LearningStrategy | Question)						
Family: binomial (logit)						
Control: glmerControl(optimizer = “bobyqa”, optCtrl = list(maxfun = 2e+05))						
Number of observations = 1690; groups: Participant, 169; Question, 10						
Random Effects		Variance	*SD*	Correlation		
Participant	(Intercept)	1.39	1.18			
Question	(Intercept)	1.62	1.27			
	SingleE	0.30	0.54	-0.09		
	SingleKW	0.12	0.35	-0.51	0.90	
	MultiE	0.08	0.28	-0.08	1.00	0.09
Fixed Effects		Estimate	*SE*	z-Wald	*p*	
(Intercept)		1.77	0.43	4.11	< .01	
Single-E		-1.63	0.67	-2.42	0.02	
Single-KW		-0.82	0.31	-2.63	0.01	
Multi-E		-0.13	0.36	-0.35	0.73	

Note. Multi Keyword as comparison group

Table E6

Full GLMM results for Hypothesis 2C on interaction effect of Trial Type and Learning Strategy at Trial 3, Single KW as comparison group (Table 20)

Formula: CompACC ~ LearningStrategy + (1 | Participant) +
(1 + LearningStrategy | Question)
Family: binomial (logit)
Control: glmerControl(optimizer = “bobyqa”, optCtrl = list(maxfun = 2e+05))
Number of observations = 1690; groups: Participant, 169; Question, 10

Random Effects		Variance	*SD*	Correlation		
Participant	(Intercept)	1.39	1.18			
Question	(Intercept)	1.29	1.14			
	SingleE	0.08	0.28	0.69		
	MultiE	0.02	0.16	0.93	0.36	
	MultiKW	0.12	0.35	0.27	-0.52	0.61
Fixed Effects		Estimate	*SE*	z-Wald	*p*	
(Intercept)		0.95	0.44	2.17	0.03	
Single-E		-0.81	0.68	-1.18	0.24	
Multi-E		0.70	0.41	1.72	0.09	
Multi-KW		0.82	0.31	2.63	0.01	

Note. Single Keyword as comparison group

Table E7

Full GLMM results for Hypothesis 2C on Learning Strategy Order at Trial 3 (Table 20)

Formula: CompACC ~ LearningStrategyOrder + (1 | Participant) +
(1 + LearningStrategyOrder | Question)
Family: binomial (logit)
Control: glmerControl(optimizer = “bobyqa”, optCtrl = list(maxfun = 2e+05))
Number of observations = 1310; groups: Participant, 131; Question, 10

Random Effects		Variance	*SD*	Correlation	
Participant	(Intercept)	1.88	1.37		
Question	(Intercept)	2.30	1.52		
	LSO	0.20	0.44	-1.00	
Fixed Effects		Estimate	*SE*	z-Wald	*p*
(Intercept)		1.74	0.53	3.30	< .01
LearningStrategyOrder		0.13	0.33	0.40	0.69

Note. Mixed as comparison group

Table E8

Full LMEM results for Learning Strategy on DJOL sensitivity for Trial 3 in Experiment 2 (Table 21)

Formula: DJOL ~ LearningStrategy + (1 | Participant) + (1 + LearningStrategy | Question)						
Control: lmerControl(optCtrl = list(maxfun = 2e+05))						
Number of observations = 1690; groups: Participant, 169; Question, 10						
Random Effects		Variance	*SD*	Correlation		
Participant	(Intercept)	304.99	17.46			
Question	(Intercept)	13.15	3.63			
	LS	6.94	2.64	0.05		
Residual		402.87	20.07			
Fixed Effects		Estimate	*SE*	*df*	*t*-value	*p*
(Intercept)		73.65	3.64	105	20.25	< .01
LearningStrategy (LS)		2.94	2.88	121	0.76	0.45

Note. Explanation as comparison group

Table E9

Full LMEM results for Learning Strategy on DJOL sensitivity for Trial 1 in Experiment 2 (Table 21)

Formula: DJOL ~ LearningStrategy + (1 | Participant) + (1 + LearningStrategy | Question)						
Control: lmerControl(optCtrl = list(maxfun = 2e+05))						
Number of observations = 1310; groups: Participant, 131; Question, 10						
Random Effects		Variance	*SD*	Correlation		
Participant	(Intercept)	282.36	16.80			
Question	(Intercept)	68.64	8.29			
	LS	0.09	0.30	1.00		
Residual		611.92	24.74			
Fixed Effects		Estimate	*SE*	*df*	*t*-value	*p*
(Intercept)		66.93	3.50	21	19.14	< .01
LearningStrategy (LS)		4.17	3.24	128	1.29	0.20

Note. Explanation as comparison group

Table E10

Full LMEM results for Trial Type on DJOL sensitivity for Trial 3 in Experiment 2 (Table 21)

Formula: DJOL ~ TrialType + (1 | Participant) + (1 + TrialType | Question) Control: lmerControl(optCtrl = list(maxfun = 2e+05)) Number of observations = 1690; groups: Participant, 169; Question, 10						
Random Effects		Variance	*SD*	Correlation		
Participant	(Intercept)	257.09	16.03			
Question	(Intercept)	23.33	4.83			
	TT	45.79	6.77	-0.47		
Residual		395.83	19.90			
Fixed Effects		Estimate	*SE*	*df*	*t*-value	*p*
(Intercept)		79.86	2.14	26	37.24	< .01
TrialType (TT)		-16.81	3.83	47	-4.39	< .01

Note. Multi as comparison group

Table E11

Full LMEM results for interaction effect of Trial Type and Learning Strategy on DJOL sensitivity for Trial 3 in Experiment 2, Multi E as comparison group (Table 21)

Formula: DJOL ~ LearningStrategy + (1 | Participant) + (1 + LearningStrategy | Question) Control: lmerControl(optCtrl = list(maxfun = 2e+05)) Number of observations = 1690; groups: Participant, 169; Question, 10						
Random Effects		Variance	*SD*	Correlation		
Participant	(Intercept)	258.62	16.08			
Question	(Intercept)	16.15	4.02			
	MultiKW	12.18	3.49	-0.04		
	SingleKW	21.38	4.62	0.26	-0.79	
	SingleE	99.84	10.00	-0.18	-0.52	-0.11
Residual		390.85	19.77			
Fixed Effects		Estimate	*SE*	*df*	*t*-value	*p*
(Intercept)		77.06	3.74	105	20.58	< .01
Multi-KW		3.42	4.05	123	0.85	0.40
Single-KW		-13.14	4.86	114	-2.71	< .01
Single-E		-19.82	9.05	96	-2.19	0.03

Note. Multi Explanation as comparison group

Table E12

Full LMEM results for interaction effect of Trial Type and Learning Strategy on DJOL sensitivity for Trial 3 in Experiment 2, Multi KW as comparison group (Table 21)

Formula: DJOL ~ LearningStrategy + (1 | Participant) + (1 + LearningStrategy | Question)						
Control: lmerControl(optCtrl = list(maxfun = 2e+05))						
Number of observations = 1690; groups: Participant, 169; Question, 10						
Random Effects		Variance	*SD*	Correlation		
Participant	(Intercept)	258.62	16.08			
Question	(Intercept)	27.28	5.22			
	SingleKW	59.03	7.68	-0.49		
	SingleE	148.09	12.17	-0.58	0.40	
	MultiE	12.18	3.49	-0.64	0.93	0.71
Residual		390.85	19.77			
Fixed Effects		Estimate	*SE*	*df*	*t*-value	*p*
(Intercept)		80.49	2.35	27	34.29	< .01
Single-KW		-16.56	4.21	44	-394	< .01
Single-E		-23.24	8.78	72	-2.65	< .01
Multi-E		-3.42	4.05	123	-0.85	0.40

Note. Multi Keyword as comparison group

Table E13

Full LMEM results for interaction effect of Trial Type and Learning Strategy on DJOL sensitivity for Trial 3 in Experiment 2, Single KW as comparison group (Table 21)

Formula: DJOL ~ LearningStrategy + (1 | Participant) + (1 + LearningStrategy | Question)						
Control: lmerControl(optCtrl = list(maxfun = 2e+05))						
Number of observations = 1690; groups: Participant, 169; Question, 10						
Random Effects		Variance	*SD*	Correlation		
Participant	(Intercept)	258.62	16.08			
Question	(Intercept)	47.02	6.86			
	SingleE	131.54	11.47	-0.49		
	MultiE	21.38	4.62	-0.82	0.50	
	MultiKW	59.03	7.68	-0.75	0.24	0.96
Residual		390.85	19.77			
Fixed Effects		Estimate	*SE*	*df*	*t*-value	*p*
(Intercept)		63.93	3.71	43	17.26	< .01
Single-E		-6.69	9.04	80	-0.74	0.46
Multi-E		13.14	4.85	114	2.71	< .01
Multi-KW		16.56	4.21	44	3.94	< .01

Note. Single Keyword as comparison group

Table E14

Full LMEM results for Learning Strategy Order on DJOL sensitivity for Trial 3 in Experiment 2 (Table 21)

Formula: DJOL ~ LearningStrategyOrder + (1 | Participant) + (1 + LearningStrategyOrder | Question)						
Control: lmerControl(optCtrl = list(maxfun = 2e+05))						
Number of observations = 1690; groups: Participant, 169; Question, 10						
Random Effects		Variance	*SD*	Correlation		
Participant	(Intercept)	237.37	15.41			
Question	(Intercept)	12.21	3.49			
	LSO	7.41	2.72	1.00		
Residual		328.25	18.12			
Fixed Effects		Estimate	*SE*	*df*	*t*-value	*p*
(Intercept)		81.01	2.33	65	34.72	< .01
LearningStrategyOrder		-2.25	3.00	105	-0.75	0.45

Note. Mixed as comparison group

Table E15

Full GLMM results for Hypothesis 2A on Trial Type at Trial 3 in Experiment 2 (Table 22)

Formula: CompACC ~ DJOL_z * TrialType + (1 + DJOL_z | Participant) + (1 + DJOL_z * TrialType | Question) Family: binomial (logit) Control: glmerControl(optimizer = “bobyqa”, optCtrl = list(maxfun = 2e+05)) Number of observations = 1690; groups: Participant, 169; Question, 10						
Random Effects		Variance	*SD*	Correlation		
Participant	(Intercept)	1.51	1.27			
	DJOL_z	0.01	0.01	1.00		
Question	(Intercept)	1.62	1.27			
	DJOL_z	0.01	0.10	0.90		
	TT	0.06	0.24	-0.63	-0.25	
	DJOLz:TT	0.05	0.22	-0,38	-0.59	0.07
Fixed Effects		Estimate	*SE*	z-Wald	*p*	
(Intercept)		1.81	0.43	4.21	< .01	
DJOL_z		0.44	0.10	4.23	< .01	
TrialType (TT)		-0.97	0.29	-3.32	< .01	
DJOLz:TT		-0.09	0.18	-0.50	0.61	

Note. Multi as comparison group

Table E16

Full GLMM results for Hypothesis 2A on Learning Strategy at Trial 3 in Experiment 2, (Table 22)

Formula: CompACC ~ DJOL_z * LearningStrategy + (1 + DJOL_z | Participant) + (1 + DJOL_z * LearningStrategy | Question) Family: binomial (logit) Control: glmerControl(optimizer = “bobyqa”, optCtrl = list(maxfun = 2e+05)) Number of observations = 1690; groups: Participant, 169; Question, 10						
Random Effects		Variance	*SD*	Correlation		
Participant	(Intercept)	1.60	1.27			
	DJOL_z	0.01	0.10	1.00		
Question	(Intercept)	1.61	1.27			
	DJOL_z	0.08	0.28	0.83		
	LS	0.06	0.08	-.050	-0.90	
	DJOLz:LS	0.20	0.44	-0.51	-0.91	1.00
Fixed Effects		Estimate	*SE*	z-Wald	*p*	
(Intercept)		1.44	0.50	2.87	< .01	
DJOL_z		0.52	1,29	2.63	< .01	
LearningStrategy (LS)		0.17	0.33	.053	0.60	
DJOLz:LS		-0.12	0.24	-0.52	0.60	

Note. Explanation as comparison group

Table E17

Full GLMM results for Hypothesis 2A on interaction of Trial Type and Learning Strategy at Trial 3 in Experiment 2, Multi E as comparison group (Table 22)

Formula: CompACC ~ DJOL_z * TrialType * LearningStrategy +
(1 + DJOL_z | Participant) + (1 + DJOL_z * TrialType * LearningStrategy | Question)
Family: binomial (logit)
Control: glmerControl(optimizer = “bobyqa”, optCtrl = list(maxfun = 2e+05))
Number of observations = 1690; groups: Participant, 169; Question, 10

Random Eff.	Var.	*SD*	Correlation						
Participant(Int)	1.53	1.24							
DJOL_z	0.00	0.08	1.00						
Question(Int)	2.00	1.42							
DJOL_z	0.30	0.54	0.61						
Multi-KW	0.22	0.46	-0.49	-0.77					
Single-KW	0.05	0.23	-0.80	-0.69	0.21				
Single-E	0.78	0.88	0.18	-0.21	-0.43	0.43			
DJOLz:M-KW	0.43	0.66	-0.43	-0.96	0.67	0.60	0.34		
DJOLz:S-KW	0.75	0.87	-0.52	-0.97	0.67	0.67	0.33	0.99	
DJOLz:S-E	5.42	2.33	-0.30	-0.65	0.93	0.00	-0.43	0.65	0.63

Fixed Effects	Estimate	*SE*	*z*-Wald	*p*
(intercept)	1.85	0.57	3.27	< .01
DJOL_z	0.53	0.27	1.98	< .05
Multi-KW	-0.04	0.40	-0.10	0.92
Single-KW	-0.84	0.44	-1.90	0.06
Single-E	-1.72	0.83	-2.07	0.04
DJOLz:M-KW	-0.08	0.31	-0.24	0.81
DJOLz:S-KW	-0.28	0.38	-0.74	0.46
DJOLz:S-E	0.59	0.97	0.61	0.54

Note. Multi Explanation as comparison group

Table E18

Full GLMM results for Hypothesis 2A on interaction of Trial Type and Learning Strategy at Trial 3 in Experiment 2, Multi KW as comparison group (Table 22)

Formula: CompACC ~ DJOL_z * TrialType * LearningStrategy +
(1 + DJOL_z | Participant) + (1 + DJOL_z * TrialType * LearningStrategy | Question)
Family: binomial (logit)
Control: glmerControl(optimizer = “bobyqa”, optCtrl = list(maxfun = 2e+05))
Number of observations = 1690; groups: Participant, 169; Question, 10

Random Eff.	Var.	*SD*	Correlation						
Participant(Int)	1.53	1.24							
DJOL_z	0.00	0.08	1.00						
Question(Int)	1.58	1.26							
DJOL_z	0.04	0.21	0.29						
Single-KW	0.23	0.48	-0.23	-0.08					
Single-E	1.35	1.16	0.11	0.37	0.83				
Multi-E	0.22	0.47	0.18	-0.12	0.88	0.73			
DJOLz:S-KW	0.05	0.23	-0.60	0.41	-0.20	-0.04	-0.62		
DJOLz:S-E	3.87	1.97	0.09	0.20	-0.96	-0.83	-0.88	0.31	
DJOLz:M-E	0.43	0.66	0.24	-0.65	0.37	0.01	0.67	-.92	-0.44

Fixed Effects	Estimate	*SE*	*z*-Wald	*p*
(intercept)	1.81	0.43	4.22	< .01
DJOL_z	0.46	0.13	3.52	< .01
Single-KW	-0.80	0.34	-2.33	0.02
Single-E	-1.69	0.81	-2.08	0.04
Multi-E	0.04	0.40	0.10	0.92
DJOLz:S-KW	-0.20	0.20	-1.01	0.31
DJOLz:S-E	0.67	0.87	0.77	0.44
DJOLz:M-E	0.08	0.31	0.24	0.81

Note. Multi Keyword as comparison group

Table E19

Full GLMM results for Hypothesis 2A on Learning Strategy Order at Trial 3 in Experiment 2 (Table 22)

Formula: CompACC ~ DJOL_z * LearningStrategyOrder + (1 + DJOL_z | Participant) + (1 + DJOL_z * LearningStrategyOrder | Question)						
Family: binomial (logit)						
Control: glmerControl(optimizer = “bobyqa”, optCtrl = list(maxfun = 2e+05))						
Number of observations = 1310 groups: Participant, 131; Question, 10						
Random Effects		Variance	*SD*	Correlation		
Participant	(Intercept)	2.03	1.42			
	DJOL_z	0.00	0.01	-1.00		
Question	(Intercept)	2.38	1.54			
	DJOL_z	0.02	0.14	0.37		
	LSO	0.26	0.51	-0.98	-0.55	
	DJOLz:LSO	0.01	0.08	0.22	-0.82	-0.02
Fixed Effects		Estimate	*SE*	z-Wald	*p*	
(Intercept)		1.81	0.54	3.36	< .01	
DJOL_z		0.47	0.14	.28	< .01	
LearningStrategyOrder		0.10	0.35	0.29	0.77	
DJOLz:LSO		-0.06	0.18	-0.34	0.73	

Note. Mixed as comparison group

Table E20

Full LMEM results for Trial Type on CJ sensitivity for Trial 3 in Experiment 2 (Table 23)

Formula: CJ ~ TrialType + (1 | Participant) + (1 + TrialType | Question)						
Control: lmerControl(optCtrl = list(maxfun = 2e+05))						
Number of observations = 1690; groups: Participant, 169; Question, 10						
Random Effects		Variance	*SD*	Correlation		
Participant	(Intercept)	207.17	14.39			
Question	(Intercept)	23.23	4.72			
	TT	0.02	0.15	-1.00		
Residual		327.10	18.09			
Fixed Effects		Estimate	*SE*	*df*	*t*-value	*p*
(Intercept)		85.63	2.01	23	42	< .01
TrialType (TT)		-13.96	2.86	166	-4.89	< .01

Note. Multi as comparison group

Table E21

Full LMEM results for Learning Strategy on CJ sensitivity for Trial 3 in Experiment 2 (Table 10)

Formula: CJ ~ LearningStrategy + (1 | Participant) + (1 + LearningStrategy | Question) Control: lmerControl(optCtrl = list(maxfun = 2e+05)) Number of observations = 1690; groups: Participant, 169; Question, 10						
Random Effects		Variance	*SD*	Correlation		
Participant	(Intercept)	240.87	15.52			
Question	(Intercept)	45.65	6.77			
	LS	6.27	2.50	-1.00		
Residual		326.20	18.06			
Fixed Effects		Estimate	*SE*	*df*	*t*-value	*p*
(Intercept)		80.62	3.74	48	21.55	< .01
LearningStrategy (LS)		2.26	3.47	140	0.65	0.52

Note. Explanation as comparison group

Table E22

Full GLMM results for Trial Type on CJ accuracy in Experiment 2 (Table 24)

Formula: CompACC ~ CJ_z * TrialType + (1 + CJ_z | Participant) + (1 + CJ_z | Question) Family: binomial (logit) Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 2e+05)) Number of observations = 1690; groups: Participant, 169; Question, 10					
Random Effects		Variance	*SD*	Correlation	
Participant	(Intercept)	1.97	1.40		
	CJ_z	0.10	0.31	1.00	
Question	(Intercept)	1.63	1.28		
	CJ_z	0.21	0.46	0.75	
Fixed Effects		Estimate	*SE*	z-Wald	*p*
(Intercept)		1.98	0.44	4.54	< .01
CJ_z		0.90	0.17	5.15	< .01
TrialType (TT)		-1.03	0.31	-3.35	< .01
CJz:TrialType		-0.22	0.17	1.30	0.20

Note. Multi as comparison group

Table E23

Full GLMM results for Learning Strategy on CJ accuracy in Experiment 2 (Table 24)

Formula: CompACC ~ CJ_z * LearningStrategy + (1 + CJ_z | Participant) + (1 + CJ_z | Question)					
Family: binomial (logit)					
Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 2e+05))					
Number of observations = 1690; groups: Participant, 169; Question, 10					
Random Effects		Variance	*SD*	Correlation	
Participant	(Intercept)	2.10	1.45		
	CJ_z	0.10	0.32	1.00	
Question	(Intercept)	1.63	1.10		
	CJ_z	0.21	0.46	0.75	
Fixed Effects		Estimate	*SE*	z-Wald	*p*
(Intercept)		1.43	0.52	2.76	< .01
CJ_z		0.61	0.23	2.61	< .01
LearningStrategy (LS)		0.38	0.35	1.06	0.29
CJz:LearningStrategy		0.28	0.20	1.46	0.14

Note. Explanation as comparison group

Table E24

Full GLMM results for Trial on comprehension accuracy in Experiment 2, Trial 1 as comparison group (Table 25)

Formula: CompACC ~ Trial + (1 + Trial | Participant) + (1 | Question)					
Family: binomial (logit)					
Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 2e+05))					
Number of observations = 3930; groups: Participant, 131; Question, 10					
Random Effects		Variance	*SD*	Correlation	
Participant	(Intercept)	0.73	0.86		
	Trial2	0.12	0.35	-1.00	
	Trial3	0.13	0.35	0.33	-0.25
Question	(Intercept)	0.27	0.62		
Fixed Effects		Estimate	*SE*	z-Wald	*p*
(Intercept)		1.12	0.19	5.76	< .01
Trial2		-0.03	0.10	-0.30	0.76
Trial3		0.32	0.11	2.96	< .01

Note. Trial 1 as comparison group

Table E25

Full GLMM results for Trial on comprehension accuracy in Experiment 2, Trial 2 as comparison group (Table 25)

Formula: CompACC ~ Trial + (1 + Trial | Participant) + (1 | Question)					
Family: binomial (logit)					
Control: glmerControl(optimizer = “bobyqa”, optCtrl = list(maxfun = 2e+05))					
Number of observations = 3930; groups: Participant, 131; Question, 10					
Random Effects		Variance	*SD*	Correlation	
Participant	(Intercept)	0.26	0.51		
	Trial3	0.31	0.56	0.87	
	Trial1	0.12	0.35	0.99	0.79
Question	(Intercept)	0.27	0.52		
Fixed Effects		Estimate	*SE*	z-Wald	*p*
(Intercept)		1.09	0.18	5.95	< .01
Trial3		0.36	0.11	3.12	< .01
Trial1		0.03	0.10	0.30	0.76

Note. Trial 2 as comparison group

Table E26

Full GLMM results for interaction effect of Trial and Learning Strategy for both conditions on comprehension accuracy in Experiment 2, Both at Trial 1 as comparison group (Table 25)

Formula: CompACC ~ Trial + (1 + Trial | Participant) + (1 | Question)					
Family: binomial (logit)					
Control: glmerControl(optimizer = “bobyqa”, optCtrl = list(maxfun = 2e+05))					
Number of observations = 1920; groups: Participant, 64; Question, 10					
Random Effects		Variance	*SD*	Correlation	
Participant	(Intercept)	1.00	1.00		
	Trial2	0.11	0.38	-0.98	
	Trial3	0.32	0.57	-0.37	0.56
Question	(Intercept)	0.31	0.55		
Fixed Effects		Estimate	*SE*	z-Wald	*p*
(Intercept)		1.23	0.24	5.12	< .01
Trial2		-0.22	0.15	-1.53	0.13
Trial3		0.08	0.17	0.47	0.64

Note. Both at Trial 1 as comparison group

Table E27

Full GLMM results for interaction effect of Trial and Learning Strategy for both conditions on comprehension accuracy in Experiment 2, Both at Trial 2 as comparison group (Table 25)

Formula: CompACC ~ Trial + (1 + Trial | Participant) + (1 | Question)					
Family: binomial (logit)					
Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 2e+05))					
Number of observations = 1920; groups: Participant, 64; Question, 10					
Random Effects		Variance	*SD*	Correlation	
Participant	(Intercept)	0.26	0.51		
	Trial3	0.31	0.56	0.87	
	Trial1	0.12	0.35	0.99	0.79
Question	(Intercept)	0.27	0.52		
Fixed Effects		Estimate	*SE*	z-Wald	*p*
(Intercept)		1.09	0.18	5.95	< .01
Trial3		0.36	0.11	3.12	< .01
Trial1		0.03	0.10	0.30	0.76

Note. Both at Trial 2 as comparison group

Table E28

Full GLMM results for interaction effect of Trial and Learning Strategy for keyword conditions on comprehension accuracy in Experiment 2, KW at Trial 1 as comparison group (Table 25)

Formula: CompACC ~ Trial + (1 + Trial | Participant) + (1 | Question)					
Family: binomial (logit)					
Control: glmerControl(optimizer = “bobyqa”, optCtrl = list(maxfun = 2e+05))					
Number of observations = 1020; groups: Participant, 64; Question, 10					
Random Effects		Variance	*SD*	Correlation	
Participant	(Intercept)	0.38	0.62		
	Trial2	0.02	0.13	-1.00	
	Trial3	0.13	0.37	1.00	-1.00
Question	(Intercept)	0.24	0.49		
Fixed Effects		Estimate	*SE*	z-Wald	*p*
(Intercept)		0.97	0.23	4.28	< .01
Trial2		0.32	0.19	1.72	0.09
Trial3		0.50	0.21	2.36	0.02

Note. KW at Trial 1 as comparison group

Table E29

Full GLMM results for interaction effect of Trial and Learning Strategy for keyword conditions on comprehension accuracy in Experiment 2, KW at Trial 2 as comparison group (Table 25)

Formula: CompACC ~ Trial + (1 + Trial | Participant) + (1 | Question)					
Family: binomial (logit)					
Control: glmerControl(optimizer = “bobyqa”, optCtrl = list(maxfun = 2e+05))					
Number of observations = 1020; groups: Participant, 34; Question, 10					
Random Effects		Variance	*SD*	Correlation	
Participant	(Intercept)	0.24	0.48		
	Trial3	0.24	0.49	1.00	
	Trial1	0.02	0.13	1.00	1.00
Question	(Intercept)	0.24	0.49		
Fixed Effects		Estimate	*SE*	z-Wald	*p*
(Intercept)		1.29	0.22	5.82	< .01
Trial3		0.17	0.22	0.78	0.44
Trial1		-0.32	0.19	-1.72	0.09

Note. KW at Trial 2 as comparison group

Table E30

Full GLMM results for interaction effect of Trial and Learning Strategy for explanation conditions on comprehension accuracy in Experiment 2, E at Trial 1 as comparison group (Table 25)

Formula: CompACC ~ Trial + (1 + Trial | Participant) + (1 | Question)					
Family: binomial (logit)					
Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 2e+05))					
Number of observations = 990; groups: Participant, 33; Question, 10					
Random Effects		Variance	*SD*	Correlation	
Participant	(Intercept)	0.84	0.92		
	Trial2	0.44	0.66	-1.00	
	Trial3	0.17	0.42	1.00	-1.00
Question	(Intercept)	0.28	0.53		
Fixed Effects		Estimate	*SE*	z-Wald	*p*
(Intercept)		1.10	0.27	4.10	< .01
Trial2		-0.03	0.22	-0.14	0.89
Trial3		0.66	0.24	2.74	< .01

Note. E at Trial 1 as comparison group

Table E31

Full GLMM results for interaction effect of Trial and Learning Strategy for explanation conditions on comprehension accuracy in Experiment 2, E at Trial 2 as comparison group (Table 25)

Formula: CompACC ~ Trial + (1 + Trial | Participant) + (1 | Question) Family: binomial (logit) Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 2e+05)) Number of observations = 990; groups: Participant, 33; Question, 10					
Random Effects		Variance	*SD*	Correlation	
Participant	(Intercept)	0.06	0.25		
	Trial3	1.16	1.09	1.00	
	Trial1	0.44	0.66	1.00	1.00
Question	(Intercept)	0.28	0.53		
Fixed Effects		Estimate	*SE*	z-Wald	*p*
(Intercept)		1.07	0.22	4.97	< .01
Trial3		0.69	0.29	2.36	0.02
Trial1		0.03	0.22	0.14	0.89

Note. E at Trial 2 as comparison group

Table E32

Full GLMM results for Trial on metacomprehension accuracy in Experiment 2, Trial 1 as comparison group (Table 26)

Formula: CompACC ~ DJOL_z * Trial + (1 + DJOL_z * Trial | Participant) + (1 + DJOL_z | Question)							
Family: binomial (logit)							
Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 2e+05))							
Number of observations = 3930; groups: Participant, 131; Question, 10							
Random Eff.	Var.	*SD*	Correlation				
Participant(Int)	0.84	0.92					
DJOL_z	0.12	0.34	0.31				
Trial2	0.18	0.43	-0.92	-0.62			
Trial3	0.22	0.47	0.10	-0.88	0.21		
DJOLz:T2	0.33	0.58	0.00	-0.83	0.37	0.70	
DJOLz:T3	0.16	0.40	-0.23	-0.98	0.58	0.84	0.92
Question(Int)	0.27	0.52					
DJOL_z	0.01	0.83	1.00				
Fixed Effects		Estimate		*SE*	*z*-Wald		*p*
(intercept)		1.18		0.20	5.97		< .01
DJOL_z		0.36		0.09	4.15		< .01
Trial2		-0.03		0.11	-0.31		0.76
Trial3		0.31		0.12	2.60		< .01
DJOLz:T2		0.01		0.12	0.12		0.90
DJOLz:T3		0.02		0.12	0.16		0.87

Note. Trial 1 as comparison group

Table E33

Full GLMM results for Trial on metacomprehension accuracy in Experiment 2, Trial 2 as comparison group (Table 26)

Formula: CompACC ~ DJOL_z * Trial + (1 + DJOL_z * Trial | Participant) + (1 + DJOL_z | Question)								
Family: binomial (logit)								
Control: glmerControl(optimizer = “bobyqa”, optCtrl = list(maxfun = 2e+05))								
Number of observations = 3930; groups: Participant, 131; Question, 10								
Random Eff.	Var.	*SD*	Correlation					
Participant(Int)	0.30	0.55						
DJOL_z	0.12	0.35	0.53					
Trial3	0.32	0.56	0.85	0.24				
Trial1	0.18	0.43	0.77	-0.01	0.59			
DJOLz:T3	0.07	0.26	-0.57	-1.00	-0.25	-0.06		
DJOLz:T1	0.33	0.58	-0.30	-0.84	-0.29	0.37	0.80	
Question(Int)	0.27	0.52						
DJOL_z	0.01	0.08	1.00					
Fixed Effects		Estimate		*SE*		*z*-Wald		*p*
(intercept)		1.14		0.18		6.21		< .01
DJOL_z		0.37		0.08		4.50		< .01
Trial3		0.34		0.12		2.89		< .01
Trial1		0.03		0.11		0.31		0.76
DJOLz:T3		0.00		0.11		0.04		0.97
DJOLz:T1		-0.01		0.12		-0.12		0.90

Note. Trial 2 as comparison group

APPENDIX F

EXPERIMENT 2: POST-EXPERIMENT INTERVIEW ANALYSIS

The following analyses were conducted for each question in the post-experiment interview and will be discussed in the order of presentation to the participant. Based on the condition participants were assigned to and they learning strategy they selected at Trial 3, the post-experiment interview questions were adjusted accordingly. For example, if participants were assigned to the Multi Same KW-KW condition and selected the keyword learning strategy at Trial 3, they will not be asked to rate the effectiveness of the explanation learning strategy because they did not experienced it. All analyses conducted were subjected to the two-factor analysis with two levels of Learning Strategy (explanation, keyword) and two levels of Learning Strategy Order (mixed, same) and the question itself. The single trial was included as a baseline comparison.

Question 1. Were you aware of the experimental manipulation in the experiment? A count of "Yes" and "No" responses was tabulated; frequency of each answer option is presented in Table F1. A Chi-square test of independence was performed to examine the relation between the assigned conditions and participant's awareness of the manipulation of the experiment. The differences in count amongst these responses were insignificant, X^2 (4, N = 169) = 4.50, $p > .05$. Participants are equally likely to either report that they were aware of the experimental manipulation or not. Out of the participants (n = 49) who reported that they were aware of the manipulation, all of them reported that being aware of either the distractor tasks (i.e., Tower of Hanoi) prior to taking the comprehension tests, or the fact that they repeated the same tasks multiple times across trials.

Table F1

Frequency of answer options for Question 1 by Learning Strategy Order in Experiment 2

	Yes	No
Multi-Mix KW-E	11	22
Multi-Mix E-KW	10	21
Multi-Same KW-KW	12	21
Multi-Same E-E	10	23
Single Trial	6	32

Question 2. How did you think the learning strategy assigned at Trials 1 and 2 affected your ability to predict memory performances across trials? A count for "Helped", "Hindered", and "Did not affect" ability responses was tabulated; frequency for each answer option is presented in Table F2. A Chi-square test of independence was performed to examine the relation between the assigned conditions and participant's self-report of whether the learning strategy assigned at Trials 1 and 2 affected their ability to predict memory performances. The differences in count amongst these responses were insignificant, X^2 (6, N = 130) = 7.66, $p > .05$. There was no difference between the proportion of participants who reported that the learning strategy helped, hindered, or did not affect their ability to predict memory performance across trials amongst the different multi trial conditions. Across all the multi conditions, majority of the participants reported that the learning strategy assigned to them at Trials 1 and 2 helped their ability to make predictions on their comprehension test performances across trials. Participants were not aware of the effect of their experience with the learning strategy assigned on their ability to make predictions across trials.

Table F2

Frequency of answer options for Question 2 by Learning Strategy Order in Experiment 2

	Helped	Hindered	Did not affect
Multi Mix KW-E	22	2	9
Multi Mix E-KW	27	1	3
Multi Same KW-KW	27	2	4
Multi Same E-E	21	4	8

Question 3. How did you think the learning strategy assigned at Trials 1 & 2 affected your test performances across trials? A count of "Increased", "Decreased", and "Did not affect" performance responses were tabulated; frequency for each answer option is presented in Table F3. A Chi-square test of independence was performed to examine the relation between the assigned conditions and participant's self-report of whether the learning strategy assigned at Trials 1 and 2 affected their comprehension test performances across trials. The differences in count amongst these responses were insignificant, X^2 (6, N = 130) = 4.31, $p > .05$. Amongst all the multiple conditions, there were no differences between the proportion of participants who reported whether the learning strategy increased, decreased, or did not affect their comprehension test performances across trials. Participants were not aware of the effect of learning strategy on their comprehension performance at a future test.

Table F3

Frequency of answer options for Question 3 by Learning Strategy Order in Experiment 2

	Increased	Decreased	Did not affect
Multi Mix KW-E	26	2	5
Multi Mix E-KW	28	1	2
Multi Same KW-KW	25	2	6
Multi Same E-E	23	3	7

Question 4. Why did you select the **learning strategy over the other at Trial 3?** Because this question has a direct relationship to the learning strategy selected at Trial 3, the count for all the coded responses was tabulated separately based on participant's learning strategy selection; frequency for each answer option for participants who selected explanation at Trial 3 is presented in Table F4 whereas frequency for each answer option for participants who selected keyword at Trial 3 is presented in Table F5. Separate Chi-square test of independence was performed based on participant's learning strategy selection at Trial 3. For participants who selected explanation, the differences in the count amongst the responses were insignificant, X^2 (20, N = 31) = 17.93, $p > .05$. The proportion of participants spread across all the reasons why they selected explanation over keyword was not different across the different conditions, most probably due to the low number of participants in this category.

Table F4

Frequency of answer options for Question 4 by Learning Strategy Order and explanation at Trial 3 in Experiment 2

	KW-E	E-KW	KW-KW	E-E	Single
Easier to use	0	1	1	0	1
Less effort	0	2	2	1	0
More effective	0	4	0	2	2
Helped ability to predict	2	1	0	1	3
Helped test performance	1	3	1	1	1
Out of habit	0	0	0	0	0
Wanted to try	0	0	1	0	0

However, as for participants who selected keyword, the differences in count amongst the responses were significant, X^2 (24, N = 197) = 9.41, $p < .05$. Participants in the Mix conditions were more likely to report that the reason they selected keyword over explanation was because it was more effective, helped with their ability to predict their memory performances and to improve comprehension test performances. On the contrary, participants who experienced only one of the two learning strategy (either explanation or keyword) in the earlier trials were more likely to report that they were selecting the learning strategy that was easier to use and takes less effort. Interestingly, only participants who experienced both the learning strategies at Trials 1 and 2 reported that effectiveness of the learning strategy was the reason they selected keyword over explanation.

Table F5

Frequency of answer options for Question 4 by Learning Strategy Order and keyword at Trial 3 in Experiment 2

	KW-E	E-KW	KW-KW	E-E	Single
Easier to use	8	9	15	8	8
Less effort	3	1	9	6	3
More effective	7	8	7	2	5
Helped ability to predict	7	5	5	5	9
Helped test performance	10	5	2	3	11
Out of habit	0	1	0	0	1
Wanted to try	0	0	0	3	0

Question 5. How did you think the learning strategy you selected at Trial 3 affected your ability to predict memory performances across trials? Because this question has a direct relationship to the learning strategy selected at Trial 3, a count for "Helped", "Hindered", and "Did not affect" ability responses was tabulated separately based on participant's learning strategy selection; frequency for each answer option for participants who selected explanation at Trial 3 is presented in Table F6 whereas frequency for each answer option for participants who selected keyword at Trial 3 is presented in Table F7. Separate Chi-square test of independence was performed based on participant's learning strategy selection at Trial 3.

For participants who selected explanation, the differences in the count amongst the responses were insignificant, X^2 (8, N = 28) = 12.73, $p > .05$. The proportion of participants who reported that explanation helped, hindered, or did not affect their ability to predict memory performances at Trial 3 did not differ across all the conditions, most probably due to the low number of participants in this category.

Table F6

Frequency of answer options for Question 5 by Learning Strategy Order and explanation at Trial 3 in Experiment 2

	Helped	Hindered	Did not affect
Multi Mix KW-E	3	0	1
Multi Mix E-KW	8	0	2
Multi Same KW-KW	0	1	3
Multi Same E-E	4	0	1
Single Trial	4	0	1

However, as for participants who selected keyword, the differences in count amongst the responses were significant, X^2 (8, N = 140) = 16.82, $p < .05$. Participants in the Multi Mix conditions were more likely to report that the reason they selected keyword over explanation was because it helped with their ability to predict their memory performances. On the contrary, participants who experienced only one of the two learning strategy (either explanation or keyword; Multi Same conditions) in the earlier trials were more likely to report that the keyword learning strategy they selected did not affect their ability to predict their memory performances at Trial 3.

Table F7

Frequency of answer options for Question 5 by Learning Strategy Order and keyword at Trial 3 in Experiment 2

	Helped	Hindered	Did not affect
Multi Mix KW-E	24	2	3
Multi Mix E-KW	20	0	1
Multi Same KW-KW	18	0	11
Multi Same E-E	18	1	9
Single Trial	26	0	7

Question 6. How did you think the learning strategy you selected at Trial 3 affected your test performances across trials? Because this question has a direct relationship to the learning strategy selected at Trial 3, a count for “Improved”, “Decreased”, and “Did not affect” ability responses was tabulated separately based on participant’s learning strategy selection; frequency for each answer option for participants who selected explanation at Trial 3 is presented in Table F8 whereas frequency for each answer option for participants who selected keyword at Trial 3 is presented in Table F9. Separate Chi-square test of independence was performed based on participant’s learning strategy selection at Trial 3.

For participants who selected explanation, the differences in the count amongst the responses were insignificant, X^2 (4, N = 29) = 4.51, $p > .05$. The proportion of participants who reported that explanation improved, decreased, or did not affect their comprehension performances at Trial 3 did not differ across all the conditions, most probably due to the low number of participants in this category.

Table F8

Frequency of answer options for Question 6 by Learning Strategy Order and explanation at Trial 3 in Experiment 2

	Improved	Decreased	Did not affect
Multi Mix KW-E	2	0	2
Multi Mix E-KW	8	0	3
Multi Same KW-KW	1	0	3
Multi Same E-E	4	0	1
Single Trial	4	0	1

For participants who selected keyword, the differences in the count amongst the responses were insignificant, X^2 (8, N = 140) = 4.51, $p > .05$. The proportion of participants who reported that keyword improved, decreased, or did not affect their comprehension performances at Trial 3 did not differ across all the conditions.

Table F9

Frequency of answer options for Question 6 by Learning Strategy Order and keyword at Trial 3 in Experiment 2

	Improved	Decreased	Did not affect
Multi Mix KW-E	24	1	4
Multi Mix E-KW	18	0	3
Multi Same KW-KW	20	1	8
Multi Same E-E	21	1	6
Single Trial	21	1	11

Question 7. Rate the difficulty of the test questions using a 0 (not difficult at all) to 10 (extremely difficult). Because this question used a rating scale of 0 – 10 instead of normal categorical responses, an ANOVA was conducted on the difficulty ratings made. There was no significant main effect of Trial Type for the difficulty ratings, $F(1, 166) = .22$, $p > .05$, partial $\eta^2 = .00$. The average difficulty ratings did not differ for participants who had a single trial ($M = 4.92$, $SE = .34$) than those who had multiple trials ($M = 4.74$, $SE = .19$). Amongst the multi trial conditions, the main effect of Learning Strategy Order was not significant for the difficulty ratings, $F(3, 120) = 1.67$, $p > .05$, partial $\eta^2 = .04$. The average difficulty ratings did not differ amongst participants in the multi trial conditions.

Question 8. Rate the effectiveness of the keyword learning strategy using a 0 (not effective at all) to 10 (extremely effective). This question used a 10-point rating scale; participants rated the effectiveness of the keyword learning strategy using a 0 (not effective at all) – 10 (extremely effective). An ANOVA was conducted on the effectiveness ratings made. The main effect of Trial Type was not significant, $F(1, 156) = .00$, $p > .05$, partial $\eta^2 = .00$. Participants gave almost identical ratings of effectiveness in the single ($M = 6.79$, $SE = .34$) and in the multi trial conditions ($M = 6.81$, $SE = .18$). Amongst the multi trial conditions, the main effect of Learning Strategy Order was not significant for the effectiveness ratings, $F(3, 121) = .57$, $p > .05$, partial $\eta^2 = .01$. The average difficulty ratings did not differ amongst participants in the multi trial conditions.

Question 9. Rate the effectiveness of the explanation learning strategy using a 0 (not effective at all) to 10 (extremely effective). This question used a 10-point rating scale; participants rated the effectiveness of the explanation learning strategy using a 0 (not effective at all) – 10 (extremely effective). An ANOVA was conducted on the effectiveness ratings made. The main effect of Trial Type was not significant, $F(1, 104) = 1.28$, $p > .05$, partial $\eta^2 = .01$. Participant's effectiveness ratings did not differ between the single ($M = 7.00$, $SE = .99$) and in the multi trial conditions ($M = 5.85$, $SE = .22$). Amongst the multi trial conditions, the main effect of Learning Strategy Order was not significant for the effectiveness ratings, $F(3, 97) = .85$, $p > .05$, partial $\eta^2 = .03$. The average difficulty ratings did not differ amongst participants in the multi trial conditions.

Question 10. Would you recommend the keyword learning strategy to a friend who wants to do well? A count for "Yes" and "No" responses was tabulated; frequency for each answer option is presented in Table F10. A Chi-square test of independence was performed to examine the difference in a participant's inclination to recommend the keyword learning strategy to a friend who would like to do well in a test. The differences between these variables were not significant, X^2 (4, N = 139) = 8.87, $p >$ .05. There was no difference in the proportion of participants who would recommend and not recommend the learning strategy they were assigned to amongst the conditions. Although not significant, it should be noted that participants in the Multi Mix conditions—participants experienced both learning strategies in earlier trials—were less likely to not recommend the keyword learning strategy to a friend as compared to those in the Multi Same conditions—participants only experienced one of the two learning strategies in earlier trials—and in the Single trial conditions.

Table F10

Frequency of answer options for Question 10 by Learning Strategy Order and Learning Strategy in Experiment 2

	Yes	No
Multi Mix KW-E	25	1
Multi Mix E-KW	19	0
Multi Same KW-KW	27	6
Multi Same E-E	21	7
Single Trial	27	6

Question 11. Would you recommend the explanation learning strategy to a friend who wants to do well? A count for "Yes" and "No" responses was tabulated; frequency for each answer option is presented in Table F11. A Chi-square test of independence was performed to examine the difference in a participant's inclination to recommend the explanation learning strategy to a friend who would like to do well in a

test. The differences between these variables was significant, X^2 (4, N = 61) = 9.95, $p < .05$. Participants in the Multi Same conditions—participants only experienced one of the two learning strategies in earlier trials—were more likely to not recommend the explanation learning strategy than participants in the Multi Mix—participants experienced both learning strategies in earlier trials—and Single Trial conditions.

Table F11

Frequency of answer options for Question 11 by Learning Strategy Order in Experiment 2

	Yes	No
Multi Mix KW-E	7	0
Multi Mix E-KW	12	0
Multi Same KW-KW	2	2
Multi Same E-E	24	9
Single Trial	5	0

Question 12. Was the keyword learning strategy worth the effort to improve test performance? A count for "Yes", "No", and "Maybe" responses was tabulated; frequency for each answer option is presented in Table F12. A Chi-square test of independence was performed to examine the differences in participant's self-report of whether the keyword learning strategy was worth the effort to improve their test performances. The differences in count across the conditions was not significant, X^2 (8, N = 158) = 4.32, $p > .05$. There was no difference in the proportion of participants who reported that the keyword learning strategy was worth the effort, was not worth the effort, or might be worth the effort to improve comprehension test performance across all the conditions.

Table F12

Frequency of answer options for Question 12 by Learning Strategy Order and keyword at Trial 3 in Experiment 2

	Yes	No	Maybe
Multi Mix KW-E	22	3	8
Multi Mix E-KW	23	2	6
Multi Same KW-KW	22	1	10
Multi Same E-E	17	4	7
Single Trial	21	2	10

Question 13. Was the explanation learning strategy worth the effort to improve test performance? A count for "Yes", "No", and "Maybe" responses was tabulated; frequency for each answer option is presented in Table F13. A Chi-square test of independence was performed to examine the differences in participant's self-report of whether the keyword learning strategy was worth the effort to improve their test performances. The differences in count across the conditions was not significant, X^2 (8, N = 106) = 11.02, $p > .05$. There was no difference in the proportion of participants who reported that the explanation learning strategy was worth the effort, was not worth the effort, or might be worth the effort to improve comprehension test performance across all the conditions.

Table F13

Frequency of answer options for Question 13 by Learning Strategy Order and keyword at Trial 3 in Experiment 2

	Yes	No	Maybe
Multi Mix KW-E	7	11	15
Multi Mix E-KW	14	10	7
Multi Same KW-KW	2	0	2
Multi Same E-E	16	7	10
Single Trial	3	0	2

As a conclusion on the findings of participants' responses to all the post-experiment interview questions, it can be inferred that participants were not aware of the manipulations of the experiment in any way. Additionally, the proportion of responses for each answer option is similar between the multi and single trials as well as amongst the multi trial conditions. It can be concluded that participants' responses were not in any way a reaction to the awareness of the experimental manipulations.