A CLOSER LOOK AT JOLS: MENDING A METACOGNITIVE ILLUSION

A Thesis
Presented
to the Faculty of
California State University, Chico

In Partial Fulfillment
of the Requirements for the Degree
Master of Arts
in
Psychology

by
Debbie Anne Magreehan
Summer 2012
A CLOSER LOOK AT JOLS: MENDING A METACOGNITIVE ILLUSION

A Thesis

by

Debbie Anne Magreehan

Summer 2012

APPROVED BY THE DEAN OF GRADUATE STUDIES
AND VICE PROVOST FOR RESEARCH:

_________________________________
Eun K. Park, Ph.D.

APPROVED BY THE GRADUATE ADVISORY COMMITTEE:

_________________________________
Neil H. Schwartz, Ph.D., Chair

_________________________________
Eddie Vela, Ph.D.
DEDICATION

I am really grateful to have had the love and support of my family and friends throughout this whole journey. I don’t think I would have been as successful without all of the help that they have provided for me. Being a part of the Chico State community has also been a large part of the journey and I am thankful for all of the benefits received from being a graduate student at California State University at Chico.
ACKNOWLEDGMENTS

I would like to give my greatest appreciation to Dr. Neil Schwartz. Without his mentoring, I would not have been able to complete this work with such great precision; without his support, I would have not have been to complete this work with such great spirits; without his desire, I would not have learned the beauty and value of research. From Dr. Schwartz, I have learned things that will be carried forward with me into my future and career. I would also like to thank the Learning, Cognition, and Instruction Research Group for all of the insight provided throughout the journey of the study. I would like to thank all of the great minds at TU-Dresden for their help with the design of this study. Finally, my greatest thanks to Dr. Susanne Narciss who provided endless help and support toward the design of this study.
# TABLE OF CONTENTS

**PAGE**

Dedication................................................................................................................... iii

Acknowledgments ...................................................................................................... iv

List of Tables.............................................................................................................. vii

List of Figures............................................................................................................. viii

Abstract....................................................................................................................... ix

**CHAPTER**

I. Introduction............................................................................................................. 1

   Metacognition and Learning................................................................. 1
   Foresight Bias ............................................................................... 7
   Present Investigation .................................................................... 9
   Hypotheses ................................................................................... 10

II. Literature Review............................................................................................... 12

   Introduction .................................................................................. 12
   Feelings of Knowing Judgments .................................................. 14
   Judgments of Learning ................................................................. 17
   Study Time Allocation ................................................................... 22
   Retrospective Confidence Judgments........................................... 25
   How Metacognitive Monitoring is Done...................................... 31

III. Methodology ..................................................................................................... 34

   Design ........................................................................................... 34
   Participants .................................................................................... 34
   Experimental Materials .................................................................. 34
   Procedure ........................................................................................ 42
# LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Word List</td>
<td>38</td>
</tr>
<tr>
<td>2. Judgment of Learning for Each Type of Word Pair</td>
<td>44</td>
</tr>
<tr>
<td>3. Accurate Recall of Each Type of Word Pair</td>
<td>45</td>
</tr>
<tr>
<td>4. Associations Between Time Spent Studying and JOLs</td>
<td>46</td>
</tr>
<tr>
<td>5. Associations Between Time Spend Studying and Final Recall</td>
<td>47</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Predicted and Actual Recall</td>
<td>48</td>
</tr>
<tr>
<td>2. Relative Monitoring Accuracy for Word Pairs in Disfluent Font</td>
<td>49</td>
</tr>
</tbody>
</table>
ABSTRACT

A CLOSER LOOK AT JOLS: MENDING A METACOGNITIVE ILLUSION

by

Debbie Anne Magreehan

Master of Arts in Psychology

California State University, Chico

Summer 2012

This study aimed to look closer at a specific aspect of metacognitive monitoring—judgments of learning (JOLs). Learners allocate study time according to JOLs so it is essential that JOLs are accurate. Unfortunately, learners are not always accurate when making JOLs. In order to improve accuracy of monitoring, one must understand what cues learners use to form JOLs. The focus of this study is on experience-based JOLs. It looks at a particular aspect, encoding fluency. In order to provide evidence for encoding fluency used as a cue, this study manipulated the fluency of material to alter learner’s JOLs. By manipulating this process, it provides theoretical support and also serves as an application for future learning techniques. All material was manipulated within bounds of what can be used in actual classroom settings. Results of this study showed that the fluency of processing items is a central cue for making JOLs.
CHAPTER I

INTRODUCTION

Metacognition and Learning

The judgments learners make of how well they will be able to learn new material is an important variable to consider in learning and cognition, because the judgments influence the way learners deploy their cognitive skills and the amount they actually learn. Learners who are better judges of their own learning tend to do a better job monitoring and controlling the learning strategies they deploy; learners with better judgments also show better performance on new learning tasks (Son & Metcalfe, 2005). The monitoring and control of learning processes are metacognitive skills that regulate learning activities and influence learning performance. Judgments of learning reveal the adequacy of these monitoring processes.

Metacognitive monitoring can be examined either at the time learning material is encoded, or at the time learning material is retrieved. The investigation reported here focused on the time of encoding.

One of the pertinent questions about encoding is whether judgments of learning can be improved when the target material appears harder to learn. The assumption is that when material appears harder to encode, learners are more inclined to spend more time monitoring their learning processes during encoding—a condition called encoding fluency. Therefore, this investigation was designed to manipulate encoding
fluency of specific to-be-learned material to improve the accuracy of judgments of learning (JOLs).

JOLs are made by participants at the end of a learning trial and estimate the likelihood of remembering the acquired information on a later memory test (Koriat, 1997). T. O. Nelson and Dunlosky (1991) pointed out that the accuracy of JOLs is critical because if the judgments are inaccurate, the allocation of subsequent study time will be correspondingly less than optimal. Results of several studies suggest that participants allocate study time according to their JOLs (Mazzoni & Cornoldi, 1993; Mazzoni, Cornoldi, & Machitelli, 1990; T. O. Nelson & Leonesio, 1988).

JOLs, as an index of metacognitive monitoring, are often inaccurate in mirroring actual knowledge so it is important to understand how the JOLs are formed. Indeed, accurate monitoring of learning supports more effective control decisions, and effective control decisions support better learning and retention (Dunlosky, Hertzog, Kennedy, & Thiede, 2005; Lipko et al., 2009; Winne & Hadwin, 1998). Monitoring refers to evaluations of learning, and control refers to making study decisions (e.g., deciding what needs to be studied further or what has already been learned). Therefore, monitoring is crucial for control. However, before the process of monitoring can be adequately understood, it is critical to understand the cues that drive JOLs. Thus, this investigation aims to examine metacognitive monitoring by manipulating a process of fluency to mend a metacognitive illusion of overconfident judgment making and to provide theoretical support for how JOLs are made.

A recent study by Koriat, Bjork, Sheffer, and Bar (2004) used a dual process model to attempt to find support for how JOLs are made. According to Kahneman’s
(2003), the two processes refer to a system 1 and a system 2. As Kahneman (2003) explained it:

the operations of system 1 are typically fast, automatic, effortless, associative, implicit, and not emotionally charged. The operations of system 2 are slower, serial, effortful, more likely to be consciously monitored and deliberately controlled; they are potentially rule governed. (p. 697)

Thus, the dual process model highlights the distinction of experience-based JOLs—System 1, and theory-based JOLs—System 2. Experience-based processes are cued by a to-be-learned item’s properties; for example, the item’s associability and concreteness, including the fluency with which the item comes to mind, (e.g. an internal cue specific to the learner). Theory-based processes, on the other hand, are guided by a learner’s metacognitive beliefs. A metacognitive belief is a learner’s theory, or mental model, of how memory works. An example of a metacognitive belief would be: learning with pictures results in better memory. Previous research has found evidence for both types of JOLs—experience-based (Begg, Duft, Lalonde, Melnick & Sanvito, 1989; Koriat, 1997), and theory-based (Dunning, Johnson, Ehrlinger, & Kruger, 2003).

To determine whether experience or theory is a better predictor of how learners actually make JOLs, Koriat et al. (2004) used retention intervals. Retention intervals were used to cue learners’ system 2 theory that memory of items will be worse overtime. The learners were asked to estimate how well they would recall a target word at one of three retention intervals: 10 minutes, 1 day, and 1 week later. Koriat et al. (2004) reasoned that if the learners made JOLs on experience-based information (system 1), they should be indifferent to the retention intervals because they would be ignoring the theory that memory declines over time. On the other hand, if the learners were using reasoning
based on system 2, they would judge that words would have a lower recall after the 1-week interval.

The first experiment used a between-subjects design in which learners were given word pairs and then asked to make their judgments of learning after each pair. No significant differences were found between each of the retention intervals—a finding, which supported JOLs, based on experience (system 1). This finding indicates that learners were making judgments based on their experience during processing of the word pairs, rather than using the theory that memory declines overtime as a function of the retention intervals.

In a second experiment, Koriat et al. (2004) used a within-subject design in which learners were asked to make JOLs relative to all of the retention intervals at the same time. In the between subjects design, the learners were asked to predict the number of words they would recall at each different retention interval. The question presented in the second experiment was posed to learners about other learners. Specifically, the participant learners were asked to judge how many word pairs would be recalled on average by other learners after ten minutes, after a day, and after a week. Results from experiment 2 revealed that participant learners estimated progressively fewer items remembered for the other learners at one week, one day, and ten minutes, respectively. The results supported the position that JOLs are made by learners on the basis of theory (System 2), not experience (System 1).

Unfortunately, there are several methodological problems in Koriat et al.’s (2004) second experiment. One is that the JOLs made by participant learners were relative and comparative rather than singular and independent. The second problem is
that participant learners were asked to make JOLs of other learners rather than themselves. Having all retention intervals judged together, participant learners might have been making their judgments of each interval relative to the other intervals by comparing the three with one another. Judgments made for others are not judgments of self; thus, other-person judgments, by definition, can be regarded as a theory or belief, but cannot be assumed to be operative by self during learning or be expected to control the cognition of self. Finally, a third problem with experiment 2 was that the participant learners made their judgments in the absence of exposure to the actual learning materials. Therefore, support for theory-based JOLs was not established; only experience-based judgments received empirical support.

While Koriat et al. (2004) established that experience-based judgments seem to explain learners’ JOLs, the experiments did not make a distinction as to whether the learners made their judgments based on encoding or retrieval cues. The cues refer to particular attributes of the to-be-learned stimuli. Encoding fluency is the ease with which the items are mastered during study in the presence of the items, and retrieval fluency is the ease with which these items come to mind after study when the items are removed. The current investigation focused on experience-based JOLs, specifically from the perspective that the fluency of encoding or retrieving target-items at study provides a basis for JOLs. In short, encoding and retrieval fluency make independent contributions to JOLs (Koriat & Bjork, 2005).

Most research has focused on retrieval fluency because it is believed that delaying JOLs until some point after study results in greater JOL accuracy than when the judgments are made immediately after study (Begg et al., 1989; Benjamin & Bjork, 1996;
Benjamin, Bjork, & Schwartz, 1998; Kimball & Metcalfe, 2003; Koriat, 1997; T. O. Nelson & Dunlosky, 1991, 1992; T. O Nelson & Narens, 1994; Spellman & Bjork, 1992). Indeed, T. O Nelson and Dunlosky (1992) found that delaying a JOL for improved accuracy highlighted the importance of retrieval. In an original study, T. O Nelson and Dunlosky (1991) compared immediate to delayed JOLs of 60 word pairs. JOLs were made immediately for half of the word pairs and made at a delay for the other half. The accuracy of JOLs was increased from .38 in the immediate condition to .90 in the delayed condition. In a follow-up study, T. O Nelson and Dunlosky (1992) had learners make their judgments again at a delay, but this time with both the stimulus and response words present. The delayed JOL effect of improved accuracy disappeared; that is, the original effect was only present when learners made judgments while presented with the stimulus word alone without the response word. These findings revealed that retrieval attempts are crucial for making accurate JOLs.

Son and Metcalfe (2005) conducted a study to test the retrieval-only hypothesis. This study assessed JOLs within three different conditions; condition 1—learners were asked to make a JOL once they had practiced retrieving the target (covert); condition 2—learners were asked to make a JOL once they had practiced retrieving the target and to say the to-be-remembered word aloud (overt); condition 3—learners were asked to make a JOL without any direction of retrieval. The retrieval hypothesis claims that, except for the time needed to express responses, reaction time should be the same for all three conditions (overt, covert, and none).

Results did not support the retrieval hypothesis—JOLs in the “none” condition differed from the other two conditions. This suggests that learners do not
spontaneously attempt retrieval when making all JOLs, and less accurate JOLs may be made by a process other than target retrieval. Therefore, a 2-stage model for processing fluency is supported for encoding and retrieval. The inaccuracy of JOLs seems to occur at the encoding stage.

During the encoding stage, inaccurate JOLS may be due to the speed with which the judgments are made—the faster the speed, the better learners expect they will perform at recall, and the reciprocal. These expectations are based on what Begg et al. (1989) described in the ease of learning hypothesis, “what is easiest to do now will be remembered best later.” The ease of learning hypothesis has received much attention and has been further explained by the encoding fluency hypothesis (Hertzog, Dunlosky, Robinson, & Kidder, 2003). The encoding fluency hypothesis states that the more quickly an encoding process is concluded, the greater the subjective probability that encoding will result in retention and final recall. The difference between Beg et al.’s (1989) and Hertzog et al.’s (2003) hypotheses is that the encoding fluency hypothesis focuses on the cognitive processing operations of the to-be-remembered items, rather than simply the item’s apparent difficulty.

Foresight Bias

The illusion of quick judgments made at the encoding stage has been termed “foresight bias” (Koriat & Bjork, 2005). Foresight bias occurs when a question and an answer appear very natural or obvious when they are presented together, but when the question is later presented on a test without the answer the answer is harder to bring to mind than originally predicted—in short, an illusion of competence.
Castel, McCabe, and Roediger (2007) demonstrated this illusion of competence using identical word pairs. He showed that encoding fluency caused the foresight bias—that learners would overestimate recall of identical word pairs (scale – scale) compared to strong associates (scale – weight) or unrelated pairs (mask – scale). The effect of encoding fluency was examined through self-paced study time. Results indicated that learners spent the least amount of time studying identical word pairs. Thus, the overconfidence during study was an illusion due to the item similarity, even though the learners’ recall for these items was lower. Dunlosky and Matvey (2001) revealed similar findings. Perceptual and semantic similarity in identical pairs caused learners to engage in short amounts of study time and to assign high JOLs.

Foresight bias studies have been studied using paired associate learning, as well (Koriat & Bjork, 2006; Rhodes & Castel, 2008). Word pairs can be presented in different directions, either as forward or backward associates. Forward associated pairs are measured by the strength of association from the presented word to its associates (Umbrella—presented word  ➔ Rain—associate). Backward associated pairs are measured by the strength of association from the associate to the presented word (Rain—associate  ➔ Umbrella—presented word). The backward associated pairs are what are predicted to cause inflated JOLs and have a larger dissociation between actual and predicted recall (D. L. Nelson & Zhang, 1999). In the present investigation, strong forward and weak backward associations were manipulated to create a foresight bias.
Present Investigation

This study aims to lower the inflated JOLs by manipulating the encoding fluency. In order to manipulate encoding fluency, it must be disentangled from item difficulty. Undorf and Erdfelder (2011) claimed we are still lacking conclusive evidence for encoding fluency used as a cue for making JOLs because it has always been confounded with item difficulty in previous studies. A study was conducted with the attempt to disentangle these two; this was done using symbols rather than actual words. We do not believe this is conclusive evidence because symbols can still be more or less difficult than one another. In addition, this study used a method in which judgments of learning were made for others rather than self-judgments.

In order to manipulate encoding fluency while keeping the item difficulty constant, the current investigation uses a font manipulation. Therefore, the same items were used. This study was comprised of a fluent font condition—easy to read times new roman 12-point black font, and a disfluent font condition—difficult to read italicized times new roman 10-point gray font. The font was manipulated within bounds of variations of fonts that can be used in the classroom.

A study by Rhodes and Castel (2008) provided support for the use of font manipulation in our current investigation. They investigated whether the font size of “to-be-remembered words” influenced predicted memory performance. It was predicted that the larger font would appear more fluent and thus have a higher JOL than the smaller font. Results showed that the 48-point font did create an illusion of competence. A further experiment found the effect of font size still apparent when semantic relatedness was implemented. This suggests that perceptual features of stimuli belong in the class of
intrinsic cues that are attended to at the expense of other, more diagnostic information that is not readily available from perception. Further, this cue can strongly influence JOL, likely via encoding fluency. Therefore, we can still ask our question whether these cues are strong enough to mend the illusion of the foresight bias by improving the inaccuracy of judgments of learning.

Hypotheses

The goal of this study is to alter encoding fluency of to-be-learned material in order to improve monitoring accuracy and show, theoretically, that encoding fluency plays a role in learners JOLs; evidence from this study can be applied toward classroom settings. We hypothesize that judgment of learning ratings will be lower when learners are making judgments for items presented in the disfluent font compared to the same items presented in the fluent font. This font manipulation should only have an affect on the predicted judgment rather than actual recall; this suggests it slows down the system 1 of the dual process model rather than activates system 2 (more analytical reasoning). Therefore, we further hypothesize that font will not have an effect on final recall; recall should be better for forward associated word pairs regardless of the font condition. A lesser dissociation between predicted and actual recall is predicted for the word pairs in the disfluent font condition, specifically for the backward associated word pairs (foresight bias). As supporting evidence for encoding fluency, measured by time spent studying each word pair, we predict that learners will have a lower predicted judgment of learning when they spend more times studying the word pairs—regardless of the type of font.
This study sheds light onto the processes of metacognitive monitoring; looking for supporting evidence for encoding fluency used as a cue when making JOLs. This is central for finding ways to improve processes. As well as improving processes, it provides implications for future learning techniques. The disfluency manipulation is within bounds of what can be used in actual classroom settings, which will allow teachers to apply this font manipulation to material that may appear very natural or obvious during learning but is actually harder to retain in memory because students do not pay enough attention due to inaccurate monitoring.
CHAPTER II

LITERATURE REVIEW

Introduction

Metacognition, thoughts about one’s own thoughts, is a relatively recent field of study in psychology. John Flavell coined this term in 1979. Even though this is a recent field of study, metacognition is something that has been happening around us for as long as we can recall. The state of recall itself is a metacognitive act. For example if you are unable to recall something but were sure that you knew it you are engaging in metacognition. Specifically, this is called a tip-of-the tongue phenomena. This type of phenomena is just one of many metacognitive experiences.

Metacognition can be studied and investigated under certain aspects. Metacognitive knowledge, metacognitive monitoring, and metacognitive control are the three main areas to be studied.

Metacognitive knowledge pertains to people’s believe about cognition. This is also known as declarative knowledge. An example of this would be when one has the belief that they are not good at reading comprehension so would invest more time in a reading comprehension task in order to receive a higher score. This is a specific belief; but, the belief can also be more general. An example of a general belief is that if one studies a word list with images then they will be more successful. This introduces a later
The research topic in metacognition—the use of theory based or experience based knowledge. This topic will be addressed further in the review.

Metacognitive monitoring refers to assessing the current state of a particular cognitive activity. This type of metacognitive activity is beneficial for learners assessing how well they are learning one item relative to another item. For example, this can be done during a math test. Learners may realize that they have not learned one type of problem as well as another problem so they will invest more time into the one that they assess a lower understanding of. This is a very important aspect because it leads to the control part of metacognition. A controversial topic in this area is if monitoring leads to control or if control leads to monitoring.

Metacognitive control is defined in the literature as regulating an ongoing cognitive activity. This can be done by either: stopping the activity, continuing the activity, or changing what has been done midway through the activity. An example of this is if during the monitoring of learning math items one decides to continue studying one item for longer because they have assessed this as less well learned. In addition, they will terminate the studying of another item because they have assessed this as well learned. This shows how important metacognitive control is because if one does not choose the appropriate cognitive control then they will not learn the material to the best of their ability. In other words, a learner can use monitoring to make a decision about how to allocate study time. My research will focus on a specific aspect of metacognitive monitoring—judgments of learning.

There are various different types of metacognitive monitoring that have been researched and discussed in the literature. They all serve a different control purpose. The
main research questions in this area are based upon how people monitor ongoing thought processes; and if this monitoring is accurate or if it causes distortion for the learner because of inaccurate assessment. A further topic is when these learners are controlling cognition, do they do this in an effective manner. An example of this last topic is the allocation of study time. In order to answer these questions, a researcher must assess the metacognitive judgments of a learner to measure monitoring. Each type of judgment is made during a different stage of learning or retrieval. Monitoring judgments include: feeling-of-knowing judgments, judgments of learning, and retrospective confidence judgments. On the other hand, to measure control processes self-paced study time and retrieval times for recall can be looked at.

Feelings of Knowing Judgments

To understand this, one must first understand how it originated from the recall-judge-recognition paradigm. In 1965, a graduate student at Stanford University, Joseph Hart, devised this. The way this worked is how the paradigm sounds- participants were given questions in which they first tried to recall the answer, then they judged how well they would be able to recognize the answers that they recalled incorrectly in a later multiple choice test. This judgment can be thought of as a feeling-of-knowing judgment. This paradigm had findings that supported the accuracy of feeling-of-knowing judgments for recognition tests. In order words, the recognition was in fact higher if they predicted that they would be able to recognize the answer correctly than if they predicted they would not.
What feeling-of-knowing judgments tell us is that we are aware of what is in our memory even though we are not able to accurately recall the information. The question is how this happens; there are two theories to how this is done. It can be described through the target-strength account or a more contemporary heuristic-based solution.

The target-strength account was described by Hart (1967). This deals with both a recall threshold and a feeling-of-knowing threshold. The recall threshold accounts for when one is able to accurately recall the target. A feeling-of-knowing threshold is when the target is somewhat activated. Therefore, if a target falls below the feeling-of-knowing threshold then one will judge that they will not be able to recognize the answer correctly on a later test. On the other hand, if a target falls below the recall threshold but somewhat above the feeling-of-knowing threshold then a learner would judge that they in fact would be able to recognize the target. A study by Connor, Balota, and Neely (1992) tried to support this theory but ended up supporting the fact that rather than tapping the strength of a particular target, feeling-of-knowing judgments are based on familiarity within a topic domain.

The domain-familiarity hypothesis is the leading heuristic-based account of feeling-of-knowing judgments. This is similar to the target strength theory but has a main underlying difference. The main distinction between the two is that this approach explains that the feeling-of-knowing arises due to the target’s existence based on another factor. This other factor could be familiarity to a cue in the question. Basically, the target itself is not elicited. This theory uses a heuristic which is an easy-to-use rule. The heuristic may be either cue familiarity or target accessibility. Experimental evidence is
not readily available for this theory but there is indirect evidence (Costermans, Lories, & Ansay, 1992).

Some indirect evidence for the cue familiarity hypothesis can be explained by first understanding the difference between omission and commission errors during incorrect recall. An error of omission is when nothing has been recalled at all; whereas, an error of commission is when the incorrect information has been recalled. A study by Krinsky and Nelson (1985) found that feeling-of-knowing judgments after commission errors were higher than omission errors. The reason for this may be that commission errors tend to occur when learners are more familiar with the domain which indirectly supports the cue familiarity hypothesis.

There is another study by Metcalfe (1993) that supports cue familiarity hypothesis. This was a study using paired-associates. Paired associates are word pairs that participants are given. An example of this would be dog-cat. In this particular example, dog is the cue word and cat is the target word. There were different groups in this study; both groups received two lists of paired-associates. The lists were either: (1) AB, AB- meaning that they were exactly the same; or (2) AD, AB- meaning that the cues were the same on both lists but the targets were different in both lists; or (3) CD, AB- meaning that the cues and the targets were not the same in either list. The hypothesis was supported that the feeling-of-knowing for the first two groups was the higher than that of the third group.

On top of the theories of how feeling-of-knowing judgments are made, there are brain bases that have been established. Even though the neural circuitry that underlies memory monitoring is unknown, research has attempted to image the brain’s functioning
when feeling-of-knowing-judgments are being made. The main hypothesis in this area is that metacognitive monitoring is controlled by the frontal lobes.

A recent study (Schnyer et al., 2004) showed that the frontal lobe is, in fact, key for accurate feeling-of-knowing judgments. Patients with specific frontal lobe lesions were assessed on a feeling-of-knowing task. There were 14 patients involved in this study. The patients had damage to the frontal cortex; they were compared with control participants who had no frontal cortex damage. Participants were asked to make feeling-of-knowing predictions of episodic memory performance. Hypotheses were supported that the frontal lobe damaged patients were impaired at both recall and recognition performance. In addition, what is most interesting is that they were also impaired in feeling-of-knowing accuracy. Although, the research has been sparse a review by Pannu and Kaszniak (2005) claims that “neurological populations are consistent with the conclusion that the frontal lobes play a central role in the production of accurate metamemory judgments” (p. 120).

Overall, the main function of feeling-of-knowing judgments is to select the best strategy for answering a question or to inform the termination of retrieval. Reder (1988) developed a method to show that feeling-of-knowing drives strategy selection. In addition, it was found that preliminary feeling-of-knowing judgments drive strategy choice that is based on cue familiarity and not on the access to the actual responses.

Judgments of Learning

The most important aspect of judgments of learning is the accuracy of these judgments. The reason for this is that this memory monitoring is what leads to allocation
of study time. There are things to be taken into account to understand the accuracy of judgments of learning; the two main factors influencing judgments of learning are—number of study trials and the timing of the judgment.

It is understood amongst most people that practicing something will lead to improve. Also, within cognitive psychology, extra study time often leads to better memory. The question that is not readily understood is what extra study time will do to memory monitoring. A study by Koriat (1997) was initiated to answer this question. The influence of multiple study-test trials on monitoring accuracy was examined. The procedure that was put forward in this study was a study-JOL-test procedure; which was repeated to create multiple trials. The finding in this study was that along with memory, the relative accuracy also did increase across trials.

The study by Koriat (1997) had another surprising finding that relates to monitoring accuracy. Even though, the relative accuracy improved across trials the absolute accuracy decreased. This is known as the underconfidence-with-practice effect. On the first study trial, participants were showing an overconfidence in the size of their JOLs. What they predicted was higher than the actual recall. This led to the fact that they predicted lower than the actual recall in the next trial. This has been found across various studies; it is most robust when participants are actually given feedback about their recall. This underconfidence with practice effect provides an opportunity for insight in the field. It has not quite been answered what is causing this illusion.

The timing of the judgment of learning is another thing that plays a part in the accuracy of the judgment. This variable has found improvement in both relative and absolute accuracy.
T. O. Nelson and Dunlosky (1991) were the first to find an effect with delaying the timing of the judgment of learning. They found that most studies prior to their investigation asked for judgments of learning immediately after studying the items. They compared these immediate judgments of learning to delayed judgments of learning that they asked for in their study. The study consisted of a list of 60 paired-associate items. Half of the items were paired with immediate judgments of learning; the other half of the items had several minutes pass before the participants were asked to make their JOL. They found substantially higher accuracy for the items with delayed judgments of learning. Like previous studies, the immediate JOLs had relative accuracy of about .38 whereas the delayed JOLs had a relative accuracy at .90.

This delayed judgment of learning effect has been replicated many times; although, there are some boundaries which lead to deeper underlying mechanisms for how judgments of learning are made. This comes down to what participants are presented with when they are asked to make the judgment of learning. For example, in the study with T. O. Nelson and Dunlosky (1991) participants were given just the cue word when asked to make the judgment of how well they would recall the second word. In a follow up study (T. O. Nelson and Dunlosky, 1992) found the delayed judgment of learning effect disappear. Instead of presenting participants with just the cue word when making the judgment of learning, they were presented with both the cue and target word. When presented with both words at a delayed time, participants did not have accurate judgments of learning. This shows that it is more than the timing of the JOL that causes greater accuracy.
A hypothesis in the literature is that it is the attempt at retrieval. This is known as the monitoring-retrieval assumption. There are different hypotheses for this assumption. Two common hypotheses are the monitoring-dual-memories and the self-fulfilling-prophecy hypothesis. The monitoring-dual-memories hypothesis was proposed by T. O. Nelson and Dunlosky (1991). This hypothesis claims that when making a JOL, information is monitored about the to be recalled item in both long-term and short-term memory. This explains why delayed JOL would have greater accuracy; when JOL is made immediately, the item is still in short term memory which disrupts the individual from monitoring whether the item is also in long-term memory. On the other hand, when an individual is making the judgment at a delayed point in time there is no interference with both short and long-term memory.

Spellman and Bjork (1992) had another hypothesis, the self-fulfilling prophecy hypothesis, that still involved retrieval but argued against the monitoring-dual-memories hypothesis. They claimed that this was a memory phenomenon. It is the idea of boosting memory of an item to be retrieved when judgment is made at a delay. This is because when it is made immediately because all responses can be easily retrieved which is artificial. Thus, the good accuracy that is ensured at a delay is a self-fulfilling prophecy. Investigations have been designed in support of each hypothesis but still none has gained complete support. It is important to note that neither is mutually exclusive and the shared sameness is that they both involve the attempt of retrieval when making accurate judgments.

Furthermore, a study by Son and Metcalfe (2005) undermined the retrieval only hypothesis. They found that there is more than retrieval attempts involved when
learners are making judgments of learning. They had various conditions; in one learners were told to make a retrieval attempt but in another they were not told to make an attempt at retrieval. If it is the case that learners will always try retrieval before making a judgment of learning then the reaction time should be no different for both of the conditions. They actually found the opposite. There was a significant different in reaction time for both of the conditions. This study found that people often produce very fast JOLs when they predict that they will not recall a response.

A hypothesis proposed early on was overlooked but may be the cause of judgments of learning. Begg et al. (1989) claimed that JOLs are based on the ease of processing of an item immediately prior to making the judgment. This hypothesis means that judgments are made based on heuristics. Using this approach, it explains how judgments may be accurate at times but it can also lead to inaccuracy. For example, the ease of processing a French-English translation such as singe-monkey. One may be quick to imagine a monkey swinging on a tree. With this immediate reaction, they could form a high prediction of recalling the translation. In reality, at the time of recall this image of the monkey swinging may still be clear in their memory. On the other hand, it may be the case that they can not remember what animal it was that was swinging on the tree and make an incorrect recall; such as, chimpanzee. This shows how processing ease can be misleading because sometimes the things that are easiest to process are actually the hardest to remember.

Hertzog et al. (2003) also examined this ease of processing hypothesis for making judgments of learning. The main idea for constructing this study was because ease of processing was not directly measured in Begg et al.’s original experiments. To
directly measure the ease of processing, they used interactive imagery. In other words, they asked participants to study paired associates (such as, “cake- girl”) and develop an image while studying these word pairs. In this case, a participant may imagination a girl eating a piece of cake. Once they had developed the image they were asked to press a key. The amount of time between when the word pairs were presented and the key-press was measured; the latency was the measure of ease of processing for that particular paired-associate. Longer latencies meant that the items were not as easy to process as shorter latencies. The finding of this study was as predicted—JOLs were negatively related to the latency of imagery generation. This supports the claim that the ease of processing while studying an item serves as a cue for people making a judgment of learning.

It is clear that the underlying mechanisms involved for making a judgment of learning are still controversial among researchers. What is known in the literature is that these judgments serve a function. Because of the functional role of judgments of learning, it is essential that research moves forward in the direction of recognizing what is influencing these judgments. All monitoring plays a different role in controlling a specific as of learning or retrieval. The role of JOL’s in learning control is to determine which items to study and how long to study them. This leads to the next section of study time allocation.

Study Time Allocation

Monitoring accuracy is key to the control of study. More importantly, the relative accuracy of metacognitive monitoring is essential for self-paced study. Study
time may be allocated to the wrong items when they are judged inaccurately. From the previous section, it makes sense that it is more useful for a learner to use delayed JOLs rather than immediate when deciding which items need to be re-studied. This reasoning is due to the increased relative accuracy found with delayed JOLs.

The theory of self-regulated study explains how metacognitive monitoring is used in pacing study time. The norm of study is essential to understanding self-regulated; this is what is expected to influence how much time you study. This concept differs for everyone. For example, if one has a higher norm of study for their math class than their science class then they would spend more time studying for math. The norm of study is altered when other factors come into play, for example—having an overall goal or certain time pressures. Therefore, further hypotheses explain how learners allocate study time across materials. The two main competing hypotheses in study time allocation are: the discrepancy-reduction model and region of proximal learning hypothesis.

The discrepancy-reduction model takes into account both on-line monitoring of items and norm of study; it focuses on the interplay of these two things. There is a type of “threshold” for each individuals norm of study. This threshold is used along with the monitoring of learning- as soon as the individual believes that they have reached the norm of study for a particular item then they will move on to the next. This model explains that one reduces discrepancy between what is the perceived current degree of learning and what is the norm of study (Dunlosky & Thiede, 1998). If this model is used, one can assume that the items judged as more difficult will be studied for longer than the less difficult ones because it takes longer and more effort to reduce a higher discrepancy between perceived and norm in difficult cases.
An interesting phenomenon arises within the discrepancy-reduction model—shift-to-easier-materials (STEM) effect. This effect has been shown across different materials and has been attributed to students developing a plan to restudy items in a way that will be most beneficial (Dunlosky & Thiede, 2004). Adaptive decision making is something that may occur when allocating study time which is exactly what this effect is showing. A common theme for students is to wait until the night before to study for a big exam. It can be assumed that mastering all the material the night before an exam is not possible. With this limited amount of time to gain knowledge on the materials, students may choose to focus on the easier items because they have a better chance of being able to at least remember these; if they spent more time on the harder items then they would be risking the possibility of not being able to get as high of a score. This counters the discrepancy-reduction model because of the third variable that is introduced.

Metcalfe (2002) introduced a new hypothesis, region-of-proximal-learning (RPL) hypothesis, to further argue that the discrepancy reduction hypothesis does not adequately explain how people use monitoring to allocate study time. In this view, discrepancy reduction is described as a time waster. Students would be spending too much time on items that were not going to get mastered because of this large discrepancy. The region-of-proximal-learning is more realistic to that of the actual learner because it states that study time is allocated to items in a region just beyond the grasp of the learner. So rather than spending time on the hardest items, the learner will first focus on what is in their regions of learning and if they have additional time will focus on what is just beyond their region. This two hypothesis are completely opposite in their viewpoints—discrepancy is hardest to easiest and region of proximal learning is easiest to hardest.
A study by Metcalfe and Kornell (2005) supported to RPL hypothesis. This was a simple experiment investigating how students with different levels of expertise learned a foreign language. The focus was on Spanish-English word translations. The results were in line with this hypothesis; the students that already had experience in Spanish spent more time on the different items but the beginners spent more time on the easier items. This finding can demonstrate how study time was allocated differently due to the region-of-proximal learning for each of the groups.

These are just two of the main competing hypotheses in the field. It is a goal of the field to continue to introduce general hypotheses about the allocation of study time. There are many factors that should be included rather than just time pressure and goal setting. For example, the individual differences in domain interest and sense of self-efficacy.

Retrospective Confidence Judgments

Confidence judgments can go alongside any answer we give. An everyday life example is if someone asks for directions. An initial answer is given for how to get to where the destination may be; along with this answer, the person will have a certain amount of confidence that the directions they have given are correct. These judgments are called retrospective confidence (RC) judgments. Even though these judgments are retrospective, they still serve the same purpose as other metamemory judgments in which they regulate memory. In reality, the confidence that one has in their memory is central to if they will share the information or if others will be convinced that the information they have given is in fact correct.
A problem with confidence judgments is when an individual is sure that they are correct but the information is entirely incorrect. Loftus and Ketcham (1991) describe this type of situation. This is the case of Howard Haupt who was accused of kidnapping a young boy. This kidnapping had a particular witness, John Picha, who judged from a photo lineup who committed the crime. Howard Haupt was not among these photographs but Picha still identified the criminal among these photos. After this first man was released, Picha choose another photo and claimed he was 90% confident that this was the suspect who committed the crime. Again, after two months Picha choose Howard Haupt as the criminal but was not very confident in this decision. Picha had the chance to see Haupt in person and after this rated his confidence in choosing Haupt from the photo lineup very high—9 out of 10. Haupt was also released after a few days. This brings up an interesting question; what is causing the overconfidence of eyewitnesses? Before answering this question, one must understand the accuracy of RC judgments.

Various factors influence the accuracy of confidence judgments. The judgments are rarely accurate but what is interesting is that the show similar patterns in either under or overconfidence. These effects have been coined as the overconfidence effect and the hard-easy effect. Understanding these effects, the goal is to move forward by improving the calibration of RC judgments.

There was an early review by Lichtenstein, Fischhoff, and Phillips (1982) that reported data showing outcomes from confidence measures. These studies presented participants with general information questions. Confidence judgments were made for all of the answers that they provided for the questions. The findings show that participants were overall more confident than they should have been in their answers. What was most
prominent about this is that the questions were all general knowledge; this indicates that people are overconfident with general-knowledge items.

Not only did this study by Lichtenstein and colleagues (1982) find overconfidence, they also found instances when participants were underconfident in the correctness of the answer that they provided. The phenomena of underconfidence is described as the easy-hard effect. Trying to explain how this effect and also the overconfidence effect work has troubled researchers. One view is that people actually can be accurate in making their confidence judgments so rather than some underlying mechanisms causing a distortion in accuracy, it is an artifact of how the experiment was designed. With this view at hand, it becomes a question of if it really is the methods used to examine how judgments are made or if there are cognitive biases for poor judgments.

Because of the debate of how confidence judgments are made, a goal of the area has been to try debiasing techniques to improve calibration as a way to look closer at how the judgments are made. There are two important classes of debiasing techniques (Keren, 1991). One class is a process-oriented modification—this attempts to change the internal processing of the judged event in a way to produce a more realistic judgment. The way in which the information is processed and represented is changed to represent the relevant information that is the basis of the judgment. Rather than the relevance of the particular judged event, the other class of techniques directly influences how the judge uses the rating scale for a single task. A participant may always be 20% overconfident in all of their answers; if this were the case, this participant would be told to reduce their confidence by 20% when making a judgment rating. There are many limitations that have been noted about this response-oriented modification. For example, the judge is not
necessarily forced to understand the structure of the problem. There have been many different techniques within these two viewpoints, some successful but many scrutinized poorly.

A technique that has been evaluated is to generate reasons for answers (Koriat, Lichtenstein, & Fischhoff, 1980). If the proposal that when making an RC judgment, people consider the reasons that confirm with why their answers are correct and ignore the reasons for why their answers are incorrect then this technique should actually reduce the overconfidence in judgments. To test this, participants had to answer to questions that were general knowledge and had only two alternatives as an answer. One group of these participants had to generate an argument in support of their answer before making an RC judgment; the other group was forced to generate an argument against their answer and then make an RC judgment. The finding confirmed that the group that had to generate an argument against their answer had improved RC judgment accuracy—overconfidence was reduced.

Allwood and Granhag (1996) attempted to replicate the results found by Koriat. This study introduced a new group. The third group that they had were given an argument against their answer. It was expected that even though the were given this argument and did not have to generate it themselves, there would still be an increase in RC judgment accuracy. The results were surprising, there was no improvement in RC accuracy for the group that were given the argument or, in fact, for the group that generated the argument themselves. So they could not replicate what Koriat found, “generating arguments reduces the overconfidence effect” (p. 118).
More promising debiasing techniques have been found with giving feedback (Stone & Opel, 2000). There are various forms of feedback but they all can debias RC judgments in a different way. Performance feedback is giving feedback on the overall accuracy of RC judgments. This feedback is given directly after answering a set of questions; for example, an individual may be informed right away that they were overconfident in all of the RC judgments. Giving this type of feedback does reduce overconfidence across a proceeding set of questions and answers. This technique has more support in the literature than the former.

Theories have proposed on why people are so often overconfident in judging knowledge. These theories can be separated by pessimism and optimism (Tversky & Kahneman, 1974). The pessimistic view is that the people making the judgments are the ones to be blamed for this overconfidence; people are inherently poor at judging their internal states. On the other hand, the optimistic view believes that people are inherently good at judging their internal states; psychological experiments are set up to make people look bad. Tversky and Kahneman (1974) are leading for the pessimistic camp. They claim that people make decision based on heuristics that distort judgments. Gigerenzer, Hoffrage, & Kleinbolting’s (1991) probabilistic mental model is leading for the optimistic camp. This model claims that the misleading questions used by experimenters make what would be good judgments appear inaccurate.

There is evidence to show that heuristics do lead to biases. Heuristics that may influence human judgment under uncertainty have been intensely scrutinized (Gilovich, Griffin, & Kahneman, 2002). Some examples are the availability heuristic and the anchoring-and-adjusting heuristic. The availability can actually produce accurate
judgments at times but it provides inaccuracy when the ease of retrieval does not reflect the probability of occurrence. The idea underlying this heuristic is that judgments are made based on the ease of retrieval of an event coming to mind. Furthermore, the anchoring-and-adjusting heuristic includes more steps than the availability. People will have an initial value to judge the occurrence of an event; after this initial value, they will adjust the likelihood of the event occurring. This shows more accuracy.

Ecological approaches, by the optimists, also have evidence in the literature. Gigerenzer et al. (1991) tested the prediction that learners were actually good judgers but it is deflated in the way that the experiment is set up. A study was set up that developed two sets of questions. One set of questions concerned which of two German cities (the largest 65 were used) had the greatest population. The other set of questions were selected from common questions used in previous studies—i.e., “Who was the first born? (a) Buddha or (b) Aristotle.” It was as predicted, the calibration accuracy was greater for the first set of questions than the second set. Although, what was more prominent for Gigerenzer et al. was to show that the difference in calibration was not due to content but the way that the questions were set up. To demonstrate this, Gigerenzer et al. choose items from the representative set that matched the selected set in difficulty. He found that now that they were matched for difficulty, there was no difference in the calibration of the two sets. This finding supports the view that people are not necessarily poor judges but it is an artifact of the experimental procedure.

Debates continue about which view, the heuristic-based or ecological-based, best demonstrates confidence judgment in answers. A mistake of this two debating camps is to think that they are completely divergent. There is some overlapping reasoning in
both viewpoints; a more mutually inclusive view may be more promising for future research.

Despite how these judgments are made, retrospective confidence judgments do serve an important function. It plays a role in whether you will or will not volunteer a response. So a function is to help us decide when it is appropriate to offer a response or when what we believe to be true should be kept to ourselves.

How Metacognitive Monitoring is Done

A debate in the field has been if metacognitive monitoring is driven by theory or experience. According to the dual process model, monitoring can arise from one of two processes: (1) subjective experience; and (2) domain-specific knowledge retrieved from memory (Koriat et al., 2004). Basically, the question is whether or not judgments are based on one’s knowledge or belief or the mnemonic feedback from the subjective experience of the task at hand. The main question that little research has focused on is when each style is adapted by learners.

Koriat et al. (2004) conducted a study investigating each aspect of the model. This was a study using paired-associates. Participants were asked to study word pairs and then eventually to recall the words that they were given. Judgments of how well they would remember the words were also recorded. This study used retention intervals as a possible cue for how learners would make their judgment of recall. The different intervals were: 10 minutes, 1 day, and 1 week. There were various experiments within this study.

The first experiment that was conducted was a between subjects experiment. This means that subjects were given the word lists and then presented with the retention
interval of 10 minutes, 1 day, or 1 week. No one subject saw all of the different retention intervals. Subjects were given the word pairs and then asked for their judgment of learning. There was no difference between the groups. This supports the experience based side of the dual process model—meaning that subjects were making their judgments based on the subjective feeling of the content rather than the theory that they will remember less overtime. The interval of time had no effect on the learner’s judgment.

Another experiment was conducted in order to tap into the theory based view. Subjects were again asked to make a recall prediction at different retention intervals. The difference with this study is that the retention intervals were conducted in a within-subject design. Subjects were presented with the following statement: “How many word pairs were recalled on average by each group; after 10 minutes: ___, 1 day: ___, and 1 week: ___”. Results showed that there was in fact a significant difference between the different intervals. Subjects rated that subjects would do significantly worse after 1 week rather than 10 minutes.

From the second experiment, Koriat et al. (2004) concluded that theory-based predictions were used. In other words, participants were using the theory that overtime it is harder to remember more items. There are confounding problems with this study which deteriorates the validity of the authors claim. One issue is that it may be a comparative effect that is causing the difference. Participants are being swayed in a direction that they should choose different numbers and it is an obvious distinction that they should get worse with memory overtime. In addition, judgments are being made for how others would do rather than for the self. Self-judgments are not always the same as other
judgments. Further studies were implemented to try to control for these confounding problems but were unable to find conclusive evidence for theory based predictions.

This study provides more promising evidence for experience based predictions rather than theory based. The problem with experience based predictions is that they may be producing faulty output. What is easily processed may not always be most easily recalled; this metacognitive illusion of competence is addressed in the current investigation. The fluency of processing an item can be done through encoding or retrieval fluency.

Experience based predictions may arise through processing fluency. This is the view that perceiving or retrieving a target at study will provide a basis for making a monitoring judgment. Most research in the past has focused on retrieval. When learners use the cue of retrieval they in fact have quite accurate judgments relative to what they actually remember. The issue is that learners are not always using retrieval fluency which was shown by the study previously described by Metcalfe (2005) that undermined the retrieval only hypothesis.

There is another type of processing fluency other than retrieval that gives rise to a subjective feeling from information. This other stage is the encoding stage; at this stage learners experience the ease at which items that are being studies are processed. This is when a problem may arise. This has been described at the illusion of competence which the current study will attempt to undermine.
CHAPTER III

METHODOLOGY

Design

Two factors, direction of word association and font, were combined to yield four experimental cells. The resulting design was a two word association (forward vs. backward) X 2 font (fluent vs. disfluent) analysis of variance with repeated measures on both factors.

Participants

Eighty undergraduate volunteers (58% female, 42% male, mean age = 23, $SD = 6.9$ years) were sampled from a midsized university in the western United States, and randomly assigned to one of the four experimental groups. Demographic data of the participants revealed that their primary language was English (93%); and 72% of the participants reported Caucasian as their ethnicity. The average GPA of the participants was 3.19 with a standard deviation of .5.

Experimental Materials

The experimental materials used in this investigation consisted of a list of 64 word pairs, a judgment of learning scale, a visual patterns test, and memory recall test.
Word List

Two word-pair lists were used for this investigation—a 4-item list of word pairs comprised of a random order of forward and backward associates used for training, and 60-item list of word pairs used for study. The study word list had 30 forward and 30 backward associate word pairs chosen from the University of South Florida free association, rhyme, and word fragment norms (D. L. Nelson & Zhang, 2004). The norms are free association response probabilities—the likelihood that one word can cue another word to come to mind—calculated for 5,019 stimulus words across 6,000 participants. Forward associative strength is the response probability from the stimulus to the target word; backward associative strength is the response probability from the target to the stimulus word. Word pairs in the current investigation were selected from the norms if the forward associative strength was above .50 and backward associative strength was below .10. Both training and study word pairs were presented in both directions—forward (e.g. Umbrella → Rain = forward associative strength of .70) and backward (e.g., Rain → Umbrella = backward associative strength of .04).

The word pair lists were also presented in fluent or disfluent fonts. Fluency is the ease with which a word is processed. In order to manipulate fluency, a gray Times New Roman italicized 10-point font was used (font sample) for the disfluent font. Black Times New Roman italicized 12-point font was used (font sample) for the fluent font. These fonts have been used across a number of investigations designed to detect effects of disfluency (cf. Alter, Oppenheimer, Epley, & Eyre, 2007), and are within the bounds of what can be considered appropriate for classroom use.
Norming Word Pairs for Difficulty and Disfluency.

It was essential to ensure that the target words were harder to recall when presented as a backward association rather than as a forward association. It was also essential to determine that all words were harder to read when disfluent, regardless of their presentation as a target or a stimulus. Thus, 20 participants were sampled from a psychology subject pool and randomly assigned to one of four experimental groups in a pilot study.

The pilot study tested the 60 word pairs—each pair presented as forward and backward associates in both the fluent and disfluent font. The associative directions and type of font were alternated between groups, yielding 4 possible combinations of word pair types: Fluent/Forward (e.g., umbrella → rain), Fluent/Backward (e.g., rain → umbrella), Disfluent/Forward (e.g., umbrella → rain), and Disfluent/Backward (e.g., umbrella → rain). To control for order effects, the associates were presented in a block design.

Participants were directed to study each of the 60 word pairs individually on a computer screen for as long as necessary for a later test of recall. After studying, participants filled out a demographic data sheet, followed by a testing phase. The testing phase directed participants to type in the second word of the pair following presentation of the first word. Each stimulus word was presented alone on a screen with a blank box for the participant’s target word response. At the end of the testing phase, participants made a fluency/disfluency rating on a 5-point likert scale (1 = very easy to read; 5 = very difficult to read). In order to control for possible order effects, half of the respondents
rated the disfluent first, then the fluent font; the other half rated the fluent, then the disfluent font.

A paired samples t-test showed that forward associated word pairs had a greater percentage recall ($M = .87; SD = .07$) than the backward associated word pairs ($M = .46, SD = .14$), $t(20) = 15.98, p < .001$. Results confirmed that the backward associated word pairs were harder to recall than the forward associated word pairs.

In addition, a paired samples t-test showed that the disfluent font was harder to read ($M = 2.48; SD = .78$) than the fluent font ($M = 1.29; SD = .78$), $t(20) = 5.29$, $p < .001$. However, the mean for the disfluent font confirmed that it was not too difficult to read, suggesting that it was within the bounds of normal variations of what can be used in classrooms.

The final sets of word pairs used in the present investigation are listed in Table 1.

**Judgment of Learning Ratings**

Judgments of learning—the estimate of learners’ likelihood of successfully recalling a target given its cue—were assessed immediately after each item and again as an aggregate after all items had been seen. Item-by-item judgments were made on a 6-point likert scale given the statement, “How well do you think you will be able to recall the second word that you just read when presented with the first word alone” (1= will definitely not be able to recall; 6= will definitely be able to recall). Aggregate judgments were assessed with a single question, “Out of the 60 items just studied, how many do you think you will be able to recall correctly?” For this estimate, learners were asked to type in a number from 0 to 60.
### Table 1

**Word List**

<table>
<thead>
<tr>
<th>Cue Word</th>
<th>Target Word</th>
<th>Forward Strength</th>
<th>Backward Strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lamp</td>
<td>Light</td>
<td>.77</td>
<td>.02</td>
</tr>
<tr>
<td>Antler</td>
<td>Deer</td>
<td>.62</td>
<td>.07</td>
</tr>
<tr>
<td>Slacks</td>
<td>Pants</td>
<td>.64</td>
<td>.01</td>
</tr>
<tr>
<td>Weary</td>
<td>Tired</td>
<td>.61</td>
<td>.00</td>
</tr>
<tr>
<td>Rectangle</td>
<td>Square</td>
<td>.72</td>
<td>.00</td>
</tr>
<tr>
<td>Row</td>
<td>Boat</td>
<td>.74</td>
<td>.02</td>
</tr>
<tr>
<td>Assist</td>
<td>Help</td>
<td>.84</td>
<td>.02</td>
</tr>
<tr>
<td>Conclude</td>
<td>End</td>
<td>.53</td>
<td>.00</td>
</tr>
<tr>
<td>Tangerine</td>
<td>Orange</td>
<td>.73</td>
<td>.05</td>
</tr>
<tr>
<td>Crib</td>
<td>Baby</td>
<td>.84</td>
<td>.00</td>
</tr>
<tr>
<td>Faucet</td>
<td>Water</td>
<td>.62</td>
<td>.04</td>
</tr>
<tr>
<td>Eight</td>
<td>Nine</td>
<td>.63</td>
<td>.00</td>
</tr>
<tr>
<td>Meow</td>
<td>Cat</td>
<td>.84</td>
<td>.00</td>
</tr>
<tr>
<td>Scissors</td>
<td>Cut</td>
<td>.88</td>
<td>.03</td>
</tr>
<tr>
<td>Bank</td>
<td>Money</td>
<td>.80</td>
<td>.02</td>
</tr>
<tr>
<td>Occupation</td>
<td>Job</td>
<td>.68</td>
<td>.03</td>
</tr>
<tr>
<td>Tuna</td>
<td>Fish</td>
<td>.81</td>
<td>.00</td>
</tr>
<tr>
<td>Illness</td>
<td>Sick</td>
<td>.73</td>
<td>.02</td>
</tr>
<tr>
<td>Duplicate</td>
<td>Copy</td>
<td>.65</td>
<td>.07</td>
</tr>
<tr>
<td>Gums</td>
<td>Teeth</td>
<td>.70</td>
<td>.08</td>
</tr>
<tr>
<td>Cue Word</td>
<td>Target Word</td>
<td>Forward Strength</td>
<td>Backward Strength</td>
</tr>
<tr>
<td>----------</td>
<td>-------------</td>
<td>------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Wick</td>
<td>Candle</td>
<td>.84</td>
<td>.05</td>
</tr>
<tr>
<td>Chirp</td>
<td>Bird</td>
<td>.78</td>
<td>.00</td>
</tr>
<tr>
<td>Incorrect</td>
<td>Wrong</td>
<td>.67</td>
<td>.05</td>
</tr>
<tr>
<td>Broth</td>
<td>Soup</td>
<td>.81</td>
<td>.04</td>
</tr>
<tr>
<td>Rye</td>
<td>Bread</td>
<td>.79</td>
<td>.00</td>
</tr>
<tr>
<td>Bloom</td>
<td>Flower</td>
<td>.59</td>
<td>.03</td>
</tr>
<tr>
<td>Sob</td>
<td>Cry</td>
<td>.76</td>
<td>.07</td>
</tr>
<tr>
<td>Spotless</td>
<td>Clean</td>
<td>.63</td>
<td>.04</td>
</tr>
<tr>
<td>Wardrobe</td>
<td>Clothes</td>
<td>.78</td>
<td>.03</td>
</tr>
<tr>
<td>Stump</td>
<td>Tree</td>
<td>.67</td>
<td>.05</td>
</tr>
<tr>
<td>Joke</td>
<td>Laugh</td>
<td>.61</td>
<td>.05</td>
</tr>
<tr>
<td>Umbrella</td>
<td>Rain</td>
<td>.70</td>
<td>.04</td>
</tr>
<tr>
<td>Flame</td>
<td>Fire</td>
<td>.51</td>
<td>.05</td>
</tr>
<tr>
<td>Jet</td>
<td>Plane</td>
<td>.66</td>
<td>.04</td>
</tr>
<tr>
<td>Icing</td>
<td>Cake</td>
<td>.81</td>
<td>.05</td>
</tr>
<tr>
<td>Sneaker</td>
<td>Shoe</td>
<td>.63</td>
<td>.05</td>
</tr>
<tr>
<td>Prim</td>
<td>Proper</td>
<td>.60</td>
<td>.04</td>
</tr>
<tr>
<td>Quench</td>
<td>Thirst</td>
<td>.82</td>
<td>.09</td>
</tr>
<tr>
<td>Option</td>
<td>Choice</td>
<td>.64</td>
<td>.03</td>
</tr>
<tr>
<td>Hilarious</td>
<td>Funny</td>
<td>.81</td>
<td>.04</td>
</tr>
<tr>
<td>Scent</td>
<td>Smell</td>
<td>.63</td>
<td>.03</td>
</tr>
<tr>
<td>Cue Word</td>
<td>Target Word</td>
<td>Forward Strength</td>
<td>Backward Strength</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
<td>------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Racquet</td>
<td>Ball</td>
<td>.69</td>
<td>.00</td>
</tr>
<tr>
<td>Starving</td>
<td>Hungry</td>
<td>.72</td>
<td>.00</td>
</tr>
<tr>
<td>Astrology</td>
<td>Star</td>
<td>.66</td>
<td>.00</td>
</tr>
<tr>
<td>Hello</td>
<td>Goodbye</td>
<td>.64</td>
<td>.00</td>
</tr>
<tr>
<td>Rake</td>
<td>Leaves</td>
<td>.62</td>
<td>.04</td>
</tr>
<tr>
<td>Automobile</td>
<td>Cars</td>
<td>.71</td>
<td>.00</td>
</tr>
<tr>
<td>Shake</td>
<td>Quiver</td>
<td>.62</td>
<td>.01</td>
</tr>
<tr>
<td>Chapter</td>
<td>Book</td>
<td>.61</td>
<td>.00</td>
</tr>
<tr>
<td>Rays</td>
<td>Sun</td>
<td>.62</td>
<td>.00</td>
</tr>
<tr>
<td>Forgive</td>
<td>Forget</td>
<td>.64</td>
<td>.01</td>
</tr>
<tr>
<td>Stallion</td>
<td>Horse</td>
<td>.63</td>
<td>.00</td>
</tr>
<tr>
<td>Noisy</td>
<td>Loud</td>
<td>.67</td>
<td>.00</td>
</tr>
<tr>
<td>Algebra</td>
<td>Math</td>
<td>.66</td>
<td>.06</td>
</tr>
<tr>
<td>Hare</td>
<td>Rabbit</td>
<td>.73</td>
<td>.03</td>
</tr>
<tr>
<td>White</td>
<td>Black</td>
<td>.66</td>
<td>.05</td>
</tr>
<tr>
<td>Skeleton</td>
<td>Bone</td>
<td>.63</td>
<td>.05</td>
</tr>
<tr>
<td>Communicate</td>
<td>Talk</td>
<td>.65</td>
<td>.05</td>
</tr>
<tr>
<td>Pal</td>
<td>Friend</td>
<td>.77</td>
<td>.08</td>
</tr>
<tr>
<td>Stumble</td>
<td>Fall</td>
<td>.71</td>
<td>.00</td>
</tr>
</tbody>
</table>
**Visual Patterns Test (VPT)**

The VPT, a measure of visual short-term memory (Della Sala, Gray, Baddeley, & Wilson, 1997), was used as an interpolated task to preclude recall from short-term memory. The instrument is a 42-item assessment comprised of matrixes of black and white squares in which respondents must memorize patterns on the squares as the patterns increase in complexity. The matrices are shown in grids progressing in size from smallest, 2X2 matrix, to largest, 5X6 matrix. Each block pattern is presented for the seconds and then a blank matrix appears. Participants are asked to reproduce the patterns by clicking on the appropriate boxes that were black in the original. Test-retest reliability has been reported at $r = .75$, with a one-week interval for 50 British adults (Della Sala, Gray, Baddeley, Allamano, & Wilson 1999). Criterion-related validity was also established by Della Sala et al. (1999), in which 345 British adolescents’ and adults’ VPT scores were correlated with number of years of formal education and age. Results revealed $r = -.55$, $p < .001$ between VPT and age, and $r = .42$ between VPT and number of years of formal education. The test takes 10 minutes to complete.

**Memory Recall Test**

The memory-recall test, a 60-item assessment comprised of all cue words from the study phase, was used as a measure of accurate recall of the target words. The test, presented via computer, was structured so that the cue word was presented to the left of the screen with a blank box to the right. Participants were directed to type the corresponding target word in the blank box.
Procedure

Participants entered a campus research lab with nine computers and were asked to choose their seats. One of the four study conditions was randomly placed on each computer. Participants were first presented with an informed consent sheet, from which they either clicked to participate or to leave. Next, directions were given for the training stage; participants were asked to click to proceed or ask the experimenter if they had questions. The training stage consisted of the four training word pairs with a judgment of learning scale following each pair. After the training stage, participants were then directed to begin the study stage wherein they were to read the word pairs then left-click when ready to make their judgment of future recall. When finished with the 60 items, participants gave their overall judgment of how many items they would be able to recall correctly. Following their overall judgment, participants were directed to complete the VPT, after which they completed the memory recall test. Finally, participants concluded the procedure by responding to several demographic questions: age, GPA, sex, ethnicity, and primary language spoken. When done, they were debriefed, thanked for their participation, and excused.
CHAPTER IV

RESULTS

Judgment of Learning Ratings

We expected the judgment of learning ratings to be lower for word pairs with disfluent font than fluent font, regardless of the direction of the word pair. This hypothesis was supported. Item by item judgment of learning ratings were summed across all items and entered into a 2 Font Type (fluent vs. disfluent) X 2 Word Association Direction (forward vs. backward) analysis of variance, with repeated measures on both variables. The analysis revealed only a main effect for type of font

\[ F(1, 79) = 11.26, MSerr = 61.3, p = .001, \text{Cohen's } d = .19, \text{ effect size } = .09. \]

As expected, disfluent font \((M = 64.3; SD = 15.5)\) produced significantly lower judgment of learning ratings than fluent font \((M = 67.3; SD = 14.9)\). No main effect was found for the direction of the word pairs, \(F(1,79) = 3.5, MSerr = 38.4, p = .07.\) And, the two variables showed no significant interaction, \(F(1, 79) = .001, MSE = 52.9, p = .97.\) See Table 2 for the means and standard deviations for the analysis.

Word Recall

We expected recall for target words to be better for forward associated word pairs than backward associated word pairs, regardless of the type of font. This hypothesis was also supported. Target word recall was scored as either correct or incorrect, with the
Table 2

*Judgment of Learning for Each Type of Word Pair*

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fluent Forward</td>
<td>67.93</td>
<td>14.16</td>
<td>80</td>
</tr>
<tr>
<td>Disfluent Forward</td>
<td>65.01</td>
<td>15.18</td>
<td>80</td>
</tr>
<tr>
<td>Fluent Backward</td>
<td>66.66</td>
<td>15.71</td>
<td>80</td>
</tr>
<tr>
<td>Disfluent Backward</td>
<td>63.70</td>
<td>15.09</td>
<td>80</td>
</tr>
</tbody>
</table>

sum of the corrects entered into a 2 Font Type (fluent vs. disfluent) X 2 Word Association Direction (Forward vs. Backward) analysis of variance, with repeated measures on both variables. The analysis revealed a main effect for word direction, $F(1, 79) = 495.6, MSerr = 7.5, p < .001$, Cohen’s $d = 2.4$, with forward associated word pairs ($M = 12.9; SD = 2.24$) significantly better recalled than backward associated word pairs ($M = 6.1; SD = 3.35$). No main effect was found for the type of font, $F(1, 79) = 2.6, MSerr = 3.1, p = .12$, or the interaction between the two variables, $F(1, 79) = .47, MSerr = 2.9, p = .49$. See Table 3 for the means and standard deviations for this analysis.

**Time Spent Studying (Encoding) with Judgments and Recall**

In order to determine if the time students spent studying the word pairs was associated with their judgments of learning, study time was correlated with learning judgments. If learners spent longer studying the word pairs, then their judgments of learning should have been lower, because the learners would be expected to judge these
Table 3

*Accurate Recall of Each Type of Word Pair*

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fluent Forward</td>
<td>12.96</td>
<td>2.29</td>
<td>80</td>
</tr>
<tr>
<td>Disfluent Forward</td>
<td>12.78</td>
<td>2.19</td>
<td>80</td>
</tr>
<tr>
<td>Fluent Backward</td>
<td>6.29</td>
<td>3.52</td>
<td>80</td>
</tr>
<tr>
<td>Disfluent Backward</td>
<td>5.84</td>
<td>3.17</td>
<td>80</td>
</tr>
</tbody>
</table>

items to be more difficult to encode. The results of the analysis supported this logic, yielding significance only for the word pairs presented in the fluent font, $r(80) = -.22$, $p = .05$. Word pairs presented in the disfluent font failed to be significantly associated with study time, $r(80) = -.01$, $p = .97$. Thus, the time spent studying the items was only associated with the learning judgments when learners had no perceptual difficulty encoding the items. Perceptual cues of the disfluent font seemed to override the perception of time spent studying the disfluent items. Indeed, an analysis of the difference between study time for each font type failed to reach statistical significance, $t(80) = 1.50$, $p = .13$. Specifically, learners spent just as much time studying word pairs in the fluent font ($M = 42; SD = 18$) as they did studying disfluent-font word pairs ($M = 44; SD = 17$). This suggests that the disfluency of word font had no influence in study time of either type of word pairs. See Table 4 for the bivariate correlation matrix.

In order to determine if the time spent studying the word pairs in both font conditions was associated with *actual* recall, study time was correlated with target-word
Table 4

*Associations Between Time Spent Studying and JOLs*

<table>
<thead>
<tr>
<th></th>
<th>Disfluent Time</th>
<th>Fluent Time</th>
<th>Disfluent JOL</th>
<th>Fluent JOL</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Correlation</strong></td>
<td>1</td>
<td>.783**</td>
<td>-.005</td>
<td>-.034</td>
</tr>
<tr>
<td><strong>Sig. (2-tailed)</strong></td>
<td>.00</td>
<td>.00</td>
<td>.968</td>
<td>.765</td>
</tr>
<tr>
<td><strong>N</strong></td>
<td>80</td>
<td>80</td>
<td>80</td>
<td>80</td>
</tr>
<tr>
<td><strong>Fluent Time</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Correlation</strong></td>
<td>.783**</td>
<td>1</td>
<td>-.191</td>
<td>-.22*</td>
</tr>
<tr>
<td><strong>Sig. (2-tailed)</strong></td>
<td>.00</td>
<td>.00</td>
<td>.089</td>
<td>.05</td>
</tr>
<tr>
<td><strong>N</strong></td>
<td>80</td>
<td>80</td>
<td>80</td>
<td>80</td>
</tr>
</tbody>
</table>

*Note.* ** indicates *p* < .001; * indicates *p* = .05

recall. Results indicated significance only for word pairs presented in the backward direction. The amount of time spent studying word pairs in the disfluent font was positively correlated with the number of correctly recalled word pairs presented in the backward direction, \( r (80) = .265, p = .018 \); in addition, the amount of time spent studying the word pairs in the fluent font was also positively correlated with the number of correctly recalled word pairs in the backward direction, \( r (80) = .241, p = .031 \). Target recall for word pairs presented in the forward direction failed to be significantly associated with the amount of time spent studying the items both in the disfluent font, \( r (80) = .026, p = .821 \) and the fluent font, \( r (80) = .167, p = .139 \). See Table 5 for bivariate correlation matrix.
Table 5

*Associations Between Time Spend Studying and Final Recall*

<table>
<thead>
<tr>
<th></th>
<th>Disfluent Time</th>
<th>Fluent Time</th>
<th>Forward Recall</th>
<th>Backward Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forward Recall</td>
<td>Correlation</td>
<td>.026</td>
<td>.167</td>
<td>1</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.821</td>
<td>.139</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>N</td>
<td>80</td>
<td>80</td>
<td>80</td>
<td>80</td>
</tr>
<tr>
<td>Backward Recall</td>
<td>Correlation</td>
<td>.265*</td>
<td>.241*</td>
<td>.479**</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.018</td>
<td>.031</td>
<td>.00</td>
<td>.05</td>
</tr>
<tr>
<td>N</td>
<td>80</td>
<td>80</td>
<td>80</td>
<td>80</td>
</tr>
</tbody>
</table>

*Note.* ** indicates \(p < .001\); * indicates \(p < .05\)

Overall Predicted Versus Actual Recall

In order to test the hypothesis that learners are overconfident when predicting overall actual recall—a condition dubbed “the overconfidence hypothesis”—learners’ overall predicted judgment of learning (0-60) was compared with their actual recall (0-60). This hypothesis was rejected. A paired samples T-test revealed that learners predicted they would recall less words (\(M = 33.4; SD = 14.7\)) than they actually did recall (\(M = 37.8; SD = 8.7\)), \(t(80) = 2.4, p = .017\), Cohen’s \(d = .36\) which suggests underconfidence (see Figure 1).

Monitoring Accuracy

Word pairs presented in the backward condition lead to a foresight bias—an erroneous estimation in which target words that appear obvious to recall during study are
Figure 1. Predicted and actual recall.

actually harder to recall during testing. Word pairs in a forward condition do not produce the bias. Thus, it was important to determine if the foresight bias was mended based on the manipulation of font in only the backward condition. We predicted more accurate judgments of learning for target words presented in the disfluent, rather than fluent, font in the backward direction. In order test this prediction, we used gamma correlations and tested the difference in these correlations between the two font conditions. T. O. Nelson
(1984) proposed that the gamma correlation—the correlation between judgment and final recall, be used to measure the relative accuracy of feeling of knowing judgments. Thus, the gamma correlation has since become the standard for measuring the relative accuracy for all metacognitive judgments. Results of the analysis revealed no differences between the two font groups, \( t(80) = -.86, p = .39 \) (see Figure 2). Learners’ predictions of the amount they would recall relative to the amount they actually recalled was uninfluenced by the font of the word pairs.

*Figure 2*. Relative monitoring accuracy for word pairs in disfluent font.
CHAPTER V

DISCUSSION

In this investigation, learners were asked to study a list of 60-paired associates; these paired associates had a strong forward association and a weak backward association. Word pairs were presented with the cue and target word in both the first and second position. When word pairs were presented with the first word as a target word for the weak backward association this created a foresight bias. The foresight bias is a metacognitive illusion of competence in which learners predict that they will do better than they actually do perform (Koriat & Bjork, 2005).

As an attempt to override the foresight bias, encoding fluency of the items was manipulated. Each learner studied word pairs presented in both fluent and disfluent font types. After studying each individual word pair, learners made a JOL on a likert scale (1-6) of how likely they would be to recall the target word when given the cue word on a later recall test. In addition, learners also made an overall judgment of how many words they thought they would recall correctly at the end.

The investigation was designed to manipulate the encoding fluency of items to find theoretical support for how JOLs are formed and to attempt to improve the accuracy of JOLs. In the main, results from the present study met these objectives. JOLs appear to be formed on the basis of the perceptual cues that are indigenous to the to-be-learned material in addition to the amount of time learners spend studying that material. In terms
of JOL accuracy, results borne from the gamma correlations failed to demonstrate that JOL accuracy was improved. However, word pairs appearing in the disfluent font yielded lower judgments of learning, decreasing the inflated JOLs.

Encoding fluency was measured as time spent studying each word pair. It was expected that when learners spent longer studying the word pairs, the learners would predict that they would remember less at recall (and the reciprocal). This expectation was based on what Begg et al. (1989) described, in the ease of learning hypothesis, as “what is easiest to do now will be remembered best later” (p. 611). In addition, the more quickly an encoding process is concluded, the greater learners’ subjective estimate of final correct recall. Results supported Begg et al.’s hypothesis. Learners show lower JOLs when they spend more time studying word pairs, but only when the word pairs are presented in a fluent font.

When word pairs were presented in the disfluent font, the time learners spent studying the word pairs did not influence JOLs. This indicates that perceptual cues appear to override the cue of study time; learners seem to use the cue of the font manipulation rather than the cue of time to make the JOL. It can be reasoned that learners are already inclined to make lower JOLs when word pairs are presented in the disfluent font. The finding provides evidence that different cues are more prominent in different situations for making JOLs. Apparently, there is no universal approach that learners take to make their judgments of learning—even during the encoding stage. In short, learners appear to make their judgments based on the combination of cues available to them.

Thus, learners do not always use the same cues to make judgments—a conclusion that adds support to the findings of Rhodes and Castel (2008). He found that
perceptual cues strongly influence JOLs via encoding fluency. The present investigation yielded findings very close to the same.

The current investigation expected to find no differences in JOLs based on the direction of the word pairs. If a difference had been detected, based on the direction of the word pairs, then a retrieval attempt would have been made before learners made JOLs. This investigation follows the evidence from Son and Metcalfe (2005) that learners make spontaneous JOLs based on cues other than retrieval attempts. Results supported Son and Metcalfe’s (2005) claim—specifically, learners had lower predicted JOLs when word pairs were presented in the disfluent font, rather than the fluent font, but there was no difference between the type of word association and learners’ JOLs. The fact that there was no difference for the word association supports the contention that learners did not practice retrieval when making JOLs.

It was expected that the font manipulation would not have an influence on final recall. This was contrary to expectations from Alter et al. (2007), based on the dual process model that perceptual cues would trigger system 2 processing and support better retention. The current investigation was based on the idea that the dual process model is not on a continuum but separate systems (Kahneman, 2003). As expected, the font manipulation only had an affect on the predicted judgment rather than actual recall. This suggests that the disfluent font slowed down system 1 processes of the dual process model rather than activating processes associated with system 2 (e.g., more analytical reasoning).

Furthermore, it was expected that overall word recall would be better for the forward associated word pairs regardless of the font condition—following the concept
obviated above. Results supported that learners have a more difficult time remembering the target words when they are presented in the direction of backward association (foresight bias). The fact that backward associated word pairs are harder to remember demonstrates that foresight bias is a pivotal issue for metacognitive monitoring. Learners expect that they will do well on a later recall test for the to-be-learned target words presented in the backward direction, but at the time of testing learners have a difficult time correctly recalling the target word.

This investigation aimed to improve the JOL accuracy for foresight bias by decreasing the inflated JOLs. Following the attempt to eliminate the foresight bias, it was expected that the JOLs would be most accurate when the word pairs were presented in the disfluent font, specifically for the word pairs presented in the backward direction. This finding was not supported which may have been due to the 6-point likert scale. With a larger range, there may have been more room for variation, potentially yielding significant gamma correlations—the correlation between the predicted and actual recall for each item.

The failure to detect accurate JOLs in the presence of words with disfluent font may have been due to contributions of anchoring and adjusting. The concept of anchoring suggests that participants rely on a fixed anchor point low on the JOL scale to make their JOLs (England & Serra, in press). With a 6-point scale, there is not much room for adjusting the fixed point, which makes it more difficult to detect differences between groups of word pairs, if the differences are there.

Even though results indicated no elimination of the foresight bias, the encoding manipulation had significant effects on judgments and recall. In addition, there
was an interesting finding with the amount of time spent studying the word pairs and final recall. When learners spent a longer time studying the word pairs presented in the backward direction, the overall correct recall of target words was better. This shows the opposite of what learners predict when making judgments of learning. Longer time spent on word pairs results in greater overall recall. On the contrary, time spent studying did not influence the recall of target words when word pairs were presented in the forward direction, which may be because overall recall is already at its peak.

The present investigation used the overall predicted JOL score and compared it with the overall number of correctly recalled items in order to test underconfidence or overconfidence. Underconfidence in judgments of learning has been found to be pervasive across multiple study-test trials as suggested by Koriat et al. (2002). On the other hand, the overconfidence effect has been found to occur for difficult items. Previous findings test these hypotheses over multiple test trials. The current investigation only used a single study trial with one JOL score. Setting an anchor at a low score leads to the expectation that learners would be underconfident in the JOL. Results supported this expectation—learners predicted a lower number of recall items than the actual recall. Learner’s that made the overall JOL prediction preceding a word pair in the disfluent font, rather than a word pair in the fluent font, predicted a lower number of items to be correctly recalled. This finding highlights the fact that the encoding manipulation of disfluency played a significant role in learners’ judgments.
REFERENCES


