USING SUBJECTIVE CONFIDENCE TO IMPROVE METACOGNITIVE MONITORING ACCURACY AND CONTROL

A Dissertation

by

TYLER MICHAEL MILLER

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2012

Major Subject: Psychology

Using Subjective Confidence to Improve Metacognitive Monitoring Accuracy and Control

USING SUBJECTIVE CONFIDENCE TO IMPROVE METACOGNITIVE MONITORING ACCURACY AND CONTROL

A Dissertation

by

TYLER MICHAEL MILLER

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Lisa Geraci
Committee Members,	Steven M. Smith
	Terrence M. Barnhardt
	Louis G.Tassinary
Head of Department,	Ludy T. Benjamin

August 2012

Major Subject: Psychology

ABSTRACT

Using Subjective Confidence to Improve Metacognitive Monitoring Accuracy and Control. (August 2012)

Tyler Michael Miller, B.A., Buena Vista University;

M.S., Emporia State University

Chair of Advisory Committee: Dr. Lisa Geraci

Metacognition is defined as a person's awareness of the capabilities and vulnerabilities of their own cognition and also encompasses the actions that a person takes as a result of that awareness. The awareness and actions that a person takes are known as monitoring and control respectively. The relationship between accurate monitoring and improved control and performance has been borne out in multiple research studies. Unfortunately, people's metacognitive judgments are far from perfect; for low performers, that inaccuracy is most often in the form of overconfidence. Attempts to improve metacognitive monitoring and control have led to mixed results. The purpose of the experiments here was to examine whether participants could use confidence in their predictions to recalibrate subsequent performance predictions and to determine if improved metacognitive monitoring would confer benefits to metacognitive control. Would participants become less overconfident and would they then decide to study longer to improve performance? In three experiments, participants made predictions about their upcoming memory performance and reported their confidence that their

predictions were accurate. Participants then adjusted their predictions so that they could be more confident the prediction was accurate. Experiment 1 served as a proof of concept – it established that confidence judgments could be used to improve metacognitive monitoring accuracy. Experiment 2 explored the boundary conditions of the calibration improvement effect. The results revealed that continuous improvement in performance predictions was possible after reporting confidence. And finally, Experiment 3 showed that participants' improved monitoring accuracy did not influence metacognitive control, which in this study was allocation of study time. One possible reason why reporting confidence did not affect metacognitive control was that participants required feedback about the benefits of confidence judgments before the improved calibration effect would influence their decisions to allocate study time. Future research will examine the influence of reporting confidence and other interventions to improve calibration and performance.

ACKNOWLEDGEMENTS

I would like to extend my gratitude to my committee chair and academic mentor, Lisa Geraci, as well as committee members, Terry Barnhardt, Steve Smith, and Louis Tassinary for their guidance throughout the course of this research.

Thanks also to my friends and colleagues and the department faculty and staff for making my time at Texas A&M University a great experience. I also want to thank the research assistants that helped me collect data, specifically Katy Cooney, Aubrey Kirchoff, and Kelley Lobpries.

Finally, thanks to my mother, father, and my brothers for their encouragement and to my wife for her patience and love.

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

1. INTRODUCTION AND LITERATURE REVIEW*

If a student in an anatomy course were required to memorize the bones of the human body to do well on a final exam, it is likely that the student would recognize the difficulty in memorizing these items without extensive study. Because of this awareness, the student would choose to study the bones. That student's awareness of the difficulty of memorizing the to-be-remembered information and the student's decision and action to study make up two related processes of *metacognition*. Metacognition is the term used to refer to a person's awareness of the state of their own cognition in addition to the capabilities and vulnerabilities of cognitive processes—the awareness of the difficulty of memorizing the bones of the human body. Secondly, metacognition refers to the actions a person takes as a result of that awareness—the prudent decision to study instead of attempting to remember them without aid. The awareness and actions a person takes are known as monitoring and control respectively. Records of people's awareness of the capabilities and

---

This dissertation follows the style of *Journal of Experimental Psychology: Learning, Memory and Cognition*.

vulnerabilities of cognitive processes date back to at least the time of Ancient Greece when orators used the method of loci (associating items with places in the environment) to deliver long speeches from memory. Then and now, people actively use mnemonic devices because they are aware of memory's vulnerabilities and are also aware of the benefits mnemonic devices provide to scaffold, or augment normal cognitive capabilities.

Metacognition has been a topic of interest and research throughout the history of modern psychology. For example, Wilhelm Wundt wrote about metacognition even though his research goals were not related to metacognition per se. He wrote that "In psychology, the person looks upon himself as from within and tries to explain the interrelations of those processes that this internal observation discloses" (Wundt, 1873). Hermann Ebbinghaus knew that using existing words in his memory experiments could confound the results and thus he strategically chose to use nonsense syllables to remove the contamination of previous experience on new learning (Fuchs & Milar, 2003). More examples of interest in metacognition can be found throughout the history of psychology. Of course, research and theory on metacognition has made significant progress since the time of Ancient Greece and early German psychology.

Today, research in metacognition is wide-ranging. From judgments-of-learning (Leonesio & Nelson, 1990), to reality monitoring (Johnson & Raye, 1981), to source monitoring (Johnson, Hashtroudi, & Lindsay, 1993), to feeling-of-knowing judgments (Hart, 1965), to second-order judgments (Dunlosky, Serra, Matvey, & Rawson, 2005), and more; there are literally thousands of examples of research studies with the express

purpose of investigating some aspect of metacognition. The whole spectrum of research in metacognition is beyond the scope of this dissertation. Rather, I will focus on one niche of metacognitive research, that is, ways in which researchers have attempted to improve individuals' metacognitive ability and allocation of study time. I will first describe theories of metacognition, methods to measure metacognition, the neurological bases of metacognition, and systematic distortions in metacognition.

1.1 Theories of metacognition

Although clearly people were aware of metacognition and were studying the concept earlier, it was not defined as we know it today until 1979 when John Flavell defined *metacognition* as "knowledge and cognition about cognitive phenomena" (p. 906). Flavell was a developmental psychologist who was responsible for some of the first studies of metamemory in children (e.g., Flavell, Friedrichs, & Hoyt, 1970). In 1979, he outlined four classes of phenomena that make up metacognition. The four classes he outlined were 1) metacognitive knowledge, which allows an individual to compare his or her own cognitive abilities to others' cognitive abilities, 2) metacognitive experience, which includes the "sudden feeling that you do not understand something another person just said," 3) goals or tasks which refer to "the objectives of cognitive enterprise", and 4) actions or strategies which refer to the type(s) of cognitive activity that will be used to reach the goals (pgs. 906-908). A later definition of metacognition according to Dunlosky and Metcalfe (2009) is somewhat more expansive. They wrote that metacognition involves "*any* reflection or judgment made upon an internal representation" (pg. 145, emphasis is the authors').

Flavell (1979) described the first theory of metacognition in the modern era and went on to present a slightly modified system almost a decade later (Flavell, 1987). In that system, Flavell elaborated on the key concepts of his metacognitive taxonomy. He wrote that within metacognitive knowledge there were person, task, and strategy variables that all provide some information to the metacognitive observer (i.e., the individual). Interestingly, he wrote about his awareness that his taxonomy of metacognition was insufficient and that "deeply insightful," "detailed proposals" have yet to be proposed (p. 28). In the same volume, Brown (1987) provided an assessment of metacognition as a concept and as an area of study more than a decade on. In her review of the existing literature at the time, Brown questioned whether the diversity of research areas claiming to be metacognition should in fact be under the metacognition heading. As Flavell predicted, these theories are seen as mostly descriptive and problematic in terms of generating testable hypotheses for later research.

One theory of metacognition that has generated a significant amount of research and is now widely accepted was proposed by Nelson and Narens (1990) and has three critical features. First, it assumes that individuals are self-reflective and that they model their environment. Second, their theory splits cognitive processes into two interrelated parts, one is the object-level (e.g., memory) and the other is the meta-level. The individual's meta-level contains a model of the cognitive process. Third, their system requires a dominance scheme in communication such that the meta-level is informed by the object-level, via monitoring, and the meta-level acts on the object-level, via control processes. In other words, information about cognitive activity feeds forward into the

individual's meta-level and this information allows the individual to monitor their cognitive activity. Once the person's meta-level of cognitive activity is updated it is compared to an ideal state, which is known as the model. From there, through control or regulatory processes, the person's meta-level modifies the object-level depending on the how the current state of activity compares to the model state. The modification could include initiating an action, continuing an action or terminating an action. From the previous example, the student attempting to memorize the bones of the human body, as the student attempts to memorize the bones, the meta-level receives information about the ongoing cognitive activity through monitoring and makes comparisons to the model. Based on these comparisons, the student is able to control their study by initiating a different kind of study strategy, continuing or discontinuing study.

Nelson and Narens' model of metacognition does not presuppose that monitoring processes provide veridical accounts of object-level activity. In fact, the exact mechanism by which individuals monitor object-level activity has yet to be determined. Two classes of mechanisms, the direct-access view and inferential views have been offered. In the direct-access view, individuals make monitoring judgments based on features of the target(s) that they can access (Schwartz, 1994). In contrast, the inferential view of monitoring processes maintains that cues and heuristics guide monitoring judgments (Schwartz & Metcalfe, 1992). For example, if a person studied a Swahili-English word pair, and was asked to indicate the likelihood of remembering the English word given the Swahili word as the cue, that person would base their judgment on their familiarity with the cue. There are other examples of the inferential view of monitoring

processes (e.g., the accessibility hypothesis; Koriat, 1993). Evidence has supported both views and it is most likely that monitoring processes are served by both mechanisms (Metcalfe, 1999). One study suggests that direct-access mechanisms may take precedence during encoding and inferential mechanisms take precedence during retrieval (Schwartz). Furthermore, other mechanisms, that have yet to be identified, could better characterize how individuals make monitoring judgments.

1.1.1 Summary

Not everyone is in agreement about a complete theory of metacognition, one that encompasses the diversity of research in the area. Indeed, metacognition as an area of research has been criticized on this point (Brown, 1987; Schraw, 2000). Fortunately there are commonalities among many theories of metacognition and agreement among most researchers about the main components of metacognition. These broad areas of agreement, namely monitoring and control, are exemplified in the Nelson and Narens (1990) model of metacognition and research has supported the monitoring and control distinction (Nelson, Dunlosky, Graf, & Narens, 1994).

1.2 Methods to measure metacognition

Because the Nelson and Narens' (1990) framework of metacognition is so well accepted it is useful to think about the variety of methods to measure metacognition and how the processes involved correspond to monitoring and control during acquisition, retention, and retrieval stages of cognition. Multiple methods to measure metacognition have been identified during acquisition and retrieval. Methods to measure metacognition during retention are scarcer.

1.2.1 Measuring metacognitive monitoring

Even in advance of learning, an individual can make metacognitive judgments. From the previous example, the student could make a monitoring judgment to determine how easy or difficult she believes memorizing the entire list of bones will be; this is known as an *ease-of-learning* (EOL) judgment (Underwood, 1966). During on-going learning, while she attempts to memorize the bones, she could make item-by-item *judgments of learning* (JOLs), in which she determines how well she believes she has learned the information.

Another type of monitoring judgment one can make that is similar to a JOL is a performance prediction. In both instances, with JOLs and performance predictions, a person attempts to evaluate how well they have learned something. A major point of distinction is the goal of the assessment. In the case of a JOL, the assessment is an end in itself. In contrast, with a performance prediction, the person not only makes a JOL, but that person also has to translate the JOL into a prediction about how about well they will perform on an upcoming test. Therefore, dissociations could exist between these JOLS and performance predictions when information about the test influences performance predictions but not JOLs (c.f. Miller and Geraci, 2011a).

When it comes to retrieving the bones of the human body on the test, the student can make a *source-monitoring judgment*, in which she attempts to remember the context or source of information for that particular fact (Johnson et al., 1993). The source of one's memory—did the professor describe the bone in class or did a student in her study group describe the bone—is important for evaluating the reliability of the information.

Following a response, the student could determine her confidence that the response she made is correct; these judgments are known as retrospective confidence judgments (Lichtenstein, Fischoff, & Phillips, 1982). Moreover, a person could also make a *postdiction* by indicating after the exam if she believed the response she made was accurate (c.f., Pierce and Smith, 2001). Results from studies on postdictions have shown that postdictions are significantly more accurate than predictions. If the student is not able to remember an answer, she could be prompted to make a *feeling-of-knowing judgment* (FOK). An affirmative FOK indicates that the person is sure they could recognize the correct answer if provided with a list of possible answers (Hart, 1965).

Another measurement theme is whether to use relative and/or absolute accuracy measures. When people make monitoring judgments on an item-by-item basis, relative accuracy or resolution is measured by computing a correlation coefficient, typically a Goodman-Kruskal gamma correlation (Nelson, 1984). In contrast, when people make monitoring judgments about a large number of items, absolute accuracy or calibration is measured by the degree to which the prediction corresponds to the actual level of performance by creating a calibration curve (Dunlosky & Metcalfe, 2009).

Relative and absolute monitoring accuracy represent different dependent variables. For example, in an experiment using a measure of relative accuracy, a participant assigns a JOL to a specific item and the question is whether or not items that received high JOLs were recalled with a greater probability than items receiving lower JOLs? If so, the participant is said to have high resolution. In contrast, for absolute accuracy, a person has accurate calibration if their actual recall level matched their JOL.

To clarify these two measures, consider a situation in which a participants assigns a particular item an 80% JOL. For relative accuracy, the question is whether or not that item is recalled more often than an item given a lower JOL (e.g., 20%). If the same person provided an 80% JOL for the entire list of items, the question for absolute accuracy would be whether or not the person recalled 80% of the items. Two related measurement themes in metacognitive accuracy are global and local monitoring accuracy. In a study by Nietfeld, Cao, and Osborne (2005), participants made item-by-item confidence judgments about the accuracy of their answer, what the authors referred to as local monitoring judgments, and they made overall confidence judgments about their accuracy on the entire test, referred to as global monitoring judgments. The authors then averaged the local monitoring judgments and compared them to the global judgments. On three different exams, the global monitoring judgments were more accurate than the local judgments.

1.2.2 Measuring metacognitive control

Much of the research into how individuals control metacognitive processes uses item selection for restudy and, by extension, which items to quit studying. This research has compared two distinct theories of item selection – the discrepancy-reduction model and the region-of-proximal learning model (Dunlosky & Hertzog, 1998; Metcalfe & Kornell, 2005). In the discrepancy-reduction model, the learner has a goal in mind, known as the norm-of-study. For example, if the norm-of-study was mastery, the learner would continue studying items until she believed that the she had memorized all of the material. In other words, the goal of study is to reduce the discrepancy between what is

known and what the norm-of-study happens to be (Dunlosky & Hertzog). This model does not specify the order of item-selection for further study. In contrast, the region-of-proximal learning model of study-time allocation states that, items that remain for further study will be prioritized from the subjectively easiest to the hardest. This latter model also accounts for how learners terminate study – so long as the learner believes they are learning they will continue study (Metcalfe & Kornell).

Learners can also use metacognitive control at retrieval. One way in which learners control their retrieval is by deciding what answers to report and what answers to withhold. Koriat and Goldsmith (1996) manipulated the incentives in a forced-choice or free-report memory test. In the moderate-incentive condition, correctly recalling an item was worth the same as the penalty for reporting an incorrect item, about $0.50. In the high-incentive condition, the penalty for false alarms ($5.00) was much larger than the incentive for correct recall ($0.50). The incentive manipulation had a measurable impact on the quantity of items participants reported in the free-report test format. That is, the quantity of recall in the high incentive/high penalty condition was significantly reduced relative to the free-report test format, which indicated that participants were able to effectively withhold low-confidence answers. Furthermore, participants' tendency to report an item was highly correlated with their subjective confidence that they had learned the item.

1.2.3 Summary

Measurements in metacognition are differentiated by when they occur in the acquisition, retention, retrieval stages of cognition and also by what process they reflect

– either monitoring or control processes. One method that has been used extensively is the judgment-of- learning (JOL). JOLs can be made by participants on an item-by-item basis and are assessed by correlating the JOL with recall or JOLs can be made by participants on a global basis and can be assessed by calculating the difference between the prediction and the performance. The numerous methods for measuring monitoring (e.g., EOLs, JOLs, and FOKs) and control (item selection) are beyond the scope of this dissertation but as described, one type of judgment affects the other. The interconnectedness of metacognitive monitoring and control cannot be emphasized enough. Indeed, Brown (1987) remarked that a source of confusion and tension among researchers who study metacognition are the attempts to separate the two processes.

1.3 Neurological bases of metacognition

Recently, researchers have attempted to identify the neural correlates of metacognitive processes. Research using imaging techniques has allowed researchers to gain a better understanding of the brain areas that are associated with metacognitive processes. Because of the top-down control processes associated with metacognition and its similarities to executive control, the frontal lobes seemed to be a good candidate location for metacognition. Indeed, the results from the overwhelming majority of studies indicate areas in the prefrontal cortex (PFC) as being associated with metacognition (Pannu & Kasniak, 2005; Schwartz & Bacon, 2008). Other research with special and patient populations has examined the influence of aging and different types of brain injury on metacognitive processes. Other studies have examined the influence of drugs on metacognition.

Very few studies measuring metacognitive processes have used imaging techniques. Those that have, indicate the important role of areas in the PFC for accurate metacognitive monitoring. In one such study, participants were asked to view pictures and predict future memory performance by making a 2-choice JOL (i.e., will recognize or will-not recognize) while event-related fMRI data were collected (Kao, Davis, & Gabrieli, 2005). The fact that brain activity was measured during learning is an important component of this experiment because brain activity would likely be different if it were measured following learning but at the same time JOLs were made. Following learning and after making a JOL, participants took a recognition test. There were four main conclusions from the study relating to predicted and actual encoding success. First, although medial temporal lobe activity was associated with encoding success (correctly recognizing the item at test), it was not associated with predicted encoding success. Therefore, the MTL area does not support the individual making JOLs. Second, medial pre-frontal cortex (PFC) activity was associated with JOL processing, that is, when a participant reported they would recognize a scene later, areas in the PFC were active. Third, individuals with greater ventro-medial PFC activity reported more accurate JOLs than those with less activity in that region. Finally, actual encoding success and predicted encoding success, JOL accuracy, was associated with lateral PFC activity.

Another interesting metacognitive process that could be occurring before conscious awareness is an event-related brain potential known as the error-related negativity (ERN). When a participant commits an error on a trial in a Stroop or Flanker task paradigm, a negative-going brain potential occurs approximately 100ms after the

response (e.g., Gehring & Fencsik, 2001; Luu, Flaisch, & Tucker, 2000). Topographical maps of electroencephalogram (EEG) activity consistently highlight the frontal lobe as the area most active during the ERN, but the signal originates in the anterior cingulate cortex (Simons, 2010). Error trials and correct trials are not differentiated by EEG activity among individuals with damage to the lateral PFC, a finding that is consistent with research indicating the importance of this area for metacognitive processing (Kao et al., 2005).

Individual differences in brain structure among healthy participants are also correlated with metacognitive ability. For example, (Fleming, Weil, Nagy, Dolan, & Rees, 2010) found that introspective accuracy on a perceptual task correlated with gray matter volume and white matter microstructure in the anterior prefrontal cortex. Individuals with more volume were more aware of their success and failures leading the researchers to suggest "a central role for anterior and dorsolateral PFC in metacognitive sensitivity" (Fleming et al., p. 1543).

More recently, repetitive transcranial magnetic stimulation (rTMS) has been used to support the connection between areas in the PFC and accurate monitoring (Rounis, Maniscalco, Rothwell, Passingham, Lau, 2010). Although this study used a slightly idiosyncratic monitoring paradigm in the visual domain, it remains the only rTMS experiment investigating metacognitive processes that this author is aware of. Participants in the experiment were required to identify the spatial location of two objects on a computer screen and to report the visibility of the objects as either "clear" or "unclear." Participants completed the task at baseline and after a session of bi-lateral

theta-burst stimulation (TBS) to areas in the dorsolateral prefrontal cortex (DLPFC) to depress cortical activity in that area. The results indicated that even though participants could complete the task as well after the rTMS as they had before rTMS, participants' self-reported visibility – the metacognitive judgment – decreased after rTMS. In other words, participants were not as metacognitively aware of their performance after rTMS as they were before the procedure.

There are also studies examining metacognitive processes in special populations. In one study, Hertzog, Sinclair, and Dunlosky (2010) collected and compared JOL resolution for paired-associate items from participants of all ages (ages 18-81). Their regression analyses revealed a significant increase in resolution across the lifespan; older adults were more likely to exhibit monitoring accuracy than younger adults. Importantly, there was also a significant decline in overall recall performance with age. Therefore, better monitoring accuracy of older adults was a result of a reduction in reported JOLs over time. In contrast, other studies have indicated that younger and older adults have equivalent JOL resolution (e.g., Hertzog, Kidder, Powell-Moman, & Dunlosky, 2002). The different conclusion from the previous two studies, with one study indicating older adults have superior monitoring ability and the other indicating equivalent monitoring, could be a result of the older adult sample used. The older adult sample used in Hertzog et al. (2010) was a full cross-sectional sample, meaning all ages were represented whereas Hertzog et al. (2002) used an extreme age-groups cross sectional design. Another possible source of the discrepancy is the type of list-learning participants engaged in – participants either used mnemonic devices (i.e., interactive imagery) while

learning the words (Hertzog et al., 2010) or the relatedness of the words was manipulated (Hertzog et al., 2002). A common finding of the two studies though is that monitoring accuracy does not decline with age. In light of typical declines in cognitive abilities across the lifespan (Salthouse, 2004; Singh-Manoux et al. 2011), these results and others (e.g., Connor, Dunlosky, & Hertzog, 1997), indicate that metacognitive ability is spared with age. Even older university-aged students are less likely than their younger counterparts to be overconfident (Grimes, 2002). In fact, older adults' spared monitoring abilities have been used as one way to improve test performance (Dunlosky, Kubat-Silman, & Hertzog, 2003). Older adult participants who were taught to regulate their study by identifying less well-learned items, a monitoring process, in order to restudy them were more likely to have improved memory test performance compared to control groups who used other study strategies or no strategies at all (Dunlosky et al., 2003; see also Hertzog & Dunlosky, 2011).

Pharmacologic manipulations and assessments of monitoring ability in other situations corroborate the robustness of metacognitive processes in healthy individuals. For example, in one study participants inhaled nitrous-oxide (N_2O) during study of paired associates and were asked to make item-by-item JOLs (Dunlosky, Domoto, Wang, Ishikawa, Roberson, Nelson, & Ramsay, 1998). N_20 was used because it has a clear detrimental effect on learning and memory and because its influence is temporary. Analysis of relative and absolute accuracy of participants' JOLs indicated that even though overall recall performance was low relative to a placebo group, N_2O inhalation did not inhibit accurate monitoring for immediate or delayed JOLs. A similar pattern of

results is observed for metacognitive accuracy under alcohol intoxication. That is, even when participants were under the influence, metacognitive processes were intact (Nelson, McSpadden, Fromme, Marlatt, 1986).

In contrast to the robustness of metacognition shown in some pharmacologic research, other research suggests that metacognitive processes are vulnerable to brain damage and hypoxia. For example, patients with dementia of the Alzheimer's disease type have frontal-executive impairments and often the earliest symptom of these deficits is the lack of insight into their impaired abilities (Mendez & Cummings, 2003). The argument that AD patients have little to no metacognition has been challenged though. For example, Moulin (2002) argued that traditional accuracy measures indicate AD patients have severely impaired metacognition. But sensitivity measures, which measure ongoing metacognitive processes during encoding, indicate AD patients do in fact exhibit metacognitive processes. Patients with vascular dementia have relatively preserved insight (Mendez & Cummings). This dissociation in awareness would likely be explained by the differences in neuronal cell death in Alzheimer's disease versus vascular dementia. Whereas neocortical frontal lobe atrophy is common in Alzheimer's disease (Salat, Kaye, & Janowsky, 2001) and even in frontotemporal dementia (FTD), frontal lobe atrophy is not always present in vascular dementia.

The fidelity of metacognitive processes also suffers in oxygen deficient states. Oxygen deficiency, or hypoxia, is a well-known problem for humans at altitude. One study examined the effects of hypoxia on metacognitive processes at various stages of a trek to the summit of Mt. Everest – in Kathmandu, at base-camp before ascent, at 6,500

or 7,100 m, at base-camp after ascent, and again in Kathmandu (Nelson, Dunlosky, White, Steinberg, Townes, & Anderson, 1990). Participants attempted to answer 34 general knowledge questions from the FACTRETRIEVAL2 battery (Wilkinson & Nelson, 1984). An example of the general knowledge question in the battery was "What is the capital of Finland? (Helsinki)" For items participants could not answer, they made feeling-of-knowing judgments and selected responses from an 8-item multiple choice format recognition test. Participants' recall at the testing locations was not different, however participants' mean FOK judgments were. Participants reported lower FOK judgments at the 3 highest altitudes even when the accuracy of recognition (after the failed recall attempt) did not differ between altitudes.

1.3.1 Summary

The quantity of research using imaging techniques to investigate aspects of metacognition is limited; however, there is some consensus among the research findings that are available. At least during study, areas in the brain that support metacognitive processes (i.e., JOLs) are located in the ventro- and dorso- medial PFC and some individual differences can be accounted for by activation in the ventro-medial PFC (Kao et al. 2005). Furthermore, the left lateral PFC is highly associated with accurate JOLs. Repetitive-TMS corroborates these latter findings by indicating that a depressed activity in the dorso-lateral PFC inhibits accurate metacognitive judgments. Research with patient populations, like those with dementia, corroborates these findings and suggests that damage to the frontal (and the temporal lobe in the case of FTD) impairs metacognitive processes. Moreover, a hypoxic environment, like one might find on Mt.

Everest, has clear effect on metacognitive monitoring. On the other hand, some pharamacologic manipulations (e.g., N_20 inhalation) and research with healthy older adults suggests that metacognitive processes are robust.

1.4 Systematic distortions in metacognitive monitoring

Although metacognitive processes can be robust into late adulthood and under unusual circumstances, the general sentiment about the accuracy of metacognitive monitoring in healthy individuals is that they are merely at above-chance accuracy – far from perfect. It is it useful to think about ways in which metacognitive processes are inaccurate. Countless studies show that people make inaccurate self-assessments. When people are asked to predict their future performance on a test, when they are asked to predict how long a task will take to complete, or even when they are asked to describe their driving skills, their assessments are inaccurate. And there are several systematic distortions, or cognitive biases that can explain these monitoring errors. The focus for the remainder of this section will be on one distortion, overconfidence, which is seen extensively, both in the research literature and in naturalistic settings.

Overconfidence, or sometimes referred to as the better-than-average effect (Alicke, 1985), is one of the most widely studied distortions in metacognitive research. Research indicates that more often than not, it is the lowest performers that are the most overconfident (e.g., Bol, Hacker, O'Shea & Allen 2005; Burson, Larrick, & Klayman, 2006, Hacker, Bol, Horgan, & Rakow, 2000; Kelemen, Winningham & Weaver, 2007; Kruger & Dunning, 1999; Krueger & Mueller, 2002; Miller & Geraci, 2011b.) For example, low performers on a given test think they have learned the information much

better than they actually have and so they predict that they will perform better on the test than they actually do.

Much research has been completed to determine why low performers are overconfident. One provocative idea is that low performers suffer from a "double curse," which is the idea that low performers not only struggle with learning the material they will be tested on, they also do not know that they are struggling and they make overly optimistic metacognitive judgments (Kruger & Dunning, 1999). It follows from the double-curse account that if low performers lack knowledge and awareness then, in addition to making inaccurate performance predictions, they would also be unduly confident in these predictions. Indeed, Dunning (2005) likened low performers' inflated self-assessments to a form of brain damage (i.e., anosognosia), and suggested that "people performing poorly cannot be expected to recognize their ineptitude" and that "the ability to recognize the depth of their inadequacies is beyond them" (pg. 15).

The double-curse characterization of low performers' inaccuracy has been challenged though. One study in particular examined whether low performers were entirely unaware of their deficits (Miller & Geraci, 2011c). To answer this question, we measured metacognition using what some people have referred to as a meta-meta judgment. We asked participants to predict their upcoming performance and also to indicate the confidence they had that the prediction was accurate. The judgments we asked participants to make highlights the distinction between two forms of confidence. Herein, we refer to errors of overestimating one's ability – predicting that one will perform better than they do as functional overconfidence and errors of overcertainty –

being overly certain of one's predictions as subjective overconfidence. Measuring subjective confidence has limited precedence in the literature. As far as I am aware, there is only one study that has examined subjective confidence associated with predictions of performance (Dunlosky et al., 2005). In this study, participants made JOLs to indicate the likelihood that they would remember unrelated noun pairs. For each JOL, participants made a second-order judgment (SOJ) indicating their confidence in the JOL. Results showed that JOLs and SOJs were functionally distinct from each other, displaying a U-shaped curvilinear relationship with higher SOJs at extreme JOLs. In addition the curve was asymmetrical, showing that SOJs associated with high JOLs were much greater than SOJs associated with low JOLs.

Returning to a description of the Miller and Geraci (2011c) study, results showed the standard effect for low performers; that is, low performers predicted that they would perform much better than they actually performed. Thus, low performers were functionally overconfident. But, interestingly, low performers were subjectively underconfident relative to high performers in that they were less confident that their predictions were accurate. In two studies on three different exams, there was a consistent dissociation between functional and subjective overconfidence. We found this dissociation regardless of whether participants predicted their scores as a letter grade or as a percentage, whether participants could earn incentives for accuracy or not, and regardless of whether their predictions were for the first exam or the final exam in the course. The fact that low performing students were less subjectively confident in their predictions than high performing students supports the notion that low performing

students may have some awareness of their ineptitude. Furthermore, this pattern of data provides evidence against the strongest version of the double curse account, which suggests that low performers overestimate performance because they are unaware of their lack of metacognitive knowledge (see Kruger & Dunning, 1999).

Given the conclusion that low performers might have some awareness of their metacognitive errors, one might ask: Why do low performers consistently exhibit more metacognitive errors than high performers? Some hypotheses about this discrepancy are that low performers are motivated to be overconfident (Gramzow, Willard, & Mendes, 2008), that it is a result of attributional style (Hacker, Bol, & Bahbahani, 2008a), that they wish to "look good" to an experimenter (Ehrlinger, Johnson, Banner, Dunning, & Kruger, 2008), or that perhaps it is simply an issue of flawed reasoning about the content of the upcoming test (Miller & Geraci, 2011a). Other possibilities are that low performers engage in the "wrong" kinds of study behaviors (e.g., highlighting, re-reading, etc.), which then may lead to overconfidence, or that the overconfidence is merely a measurement artifact (Krueger & Mueller, 2002).

Gramzow and colleagues (2008) argued that there could be adaptive benefits to overestimation because it serves to encourage future behavior. Participants in the study were interviewed about their academic performance while autonomic nervous system activity was also measured. One measure, respiratory sinus arrhythmia (RSA) is an indicator of cardiac vagal tone, or more specifically heart rate acceleration and deceleration during the respiratory cycle. Low RSA versus high RSA suggests negative emotionality like anxiety. The results indicated that those who exaggerated their grade

point average (GPA) did not exhibit low RSA which, is a typical cardiovascular reaction of lying. In fact, RSA for participants who exaggerated their GPA the most actually increased during the interview. This finding, led the authors to suggest exaggerators' equanimity while they "lied" about their academic history may be adaptive. Moreover, increased RSA during the interview was significantly positively correlated with GPA improvement (Gramzow et al., 2008).

Others have examined students' attributions for monitoring inaccuracy (Hacker et al., 2008a). After participants had taken the exam for which they made performance predictions, they were told how inaccurate their predictions were. Then, all participants completed an attributional style questionnaire that included task-centered questions (e.g., "The instruction wasn't really helpful in preparing us for the test"), student-centered testing questions("I usually get really anxious while taking tests"), student-centered studying questions ("I didn't study as much as I should have"), and social-centered questions ("My interactions with other students in class influenced my judgments"). Participants answered each of the 20 questions on a 5-point Likert scale with the degree to which they believed the question explained the discrepancy between their performance prediction and their actual performance. Results indicated that low-performing students, who also made overconfident performance predictions, attributed the discrepancy between their prediction and actual performance to external factors (e.g., "The instruction wasn't really helpful in preparing us for the test.") significantly more so than high-performing students.

Another reason participants might remain overconfident is because they wish to "look good" in front of the experimenter. To circumvent this potential confound, researchers have provided monetary incentives to increase metacognitive accuracy (Ehrlinger et al., 2008). Participants were contestants in a trap and skeet competition. Participants in the experimental (incentive) condition were offered double their money, from $5.00 to $10.00, if they gave accurate predictions on a gun safety test. Control condition participants received no such incentive, just the $5.00 base for participation. Overall, participants' accuracy did not improve when they were given incentives to be accurate and surprisingly, the low performers on the safety test actually became more overconfident in the incentive condition whereas the high performers slightly recalibrated. Questioning whether $10.00 was enough for participants to set aside self-presentation concerns, the researchers also offered undergraduate participants up to $100 for predictive accuracy on a logical reasoning task. Still, the monetary incentive had no influence on predictions for high or low performers.

Another source of metacognitive inaccuracy, typically overconfidence, could be an individuals' understanding about what will be on the test (Miller & Geraci, 2011a). Students in the classroom are aware that only a portion of the course material will be tested on the exam. Believing that only a portion of the course material can be tested may influence students to study less for less time than if they thought all of the material would be tested. In a recent study, we asked participants to study Swahili-English paired-associates in a self-paced situation and told them that either 25% of the material would be tested or 100% of the material would be tested. Results indicated that high-

performers (defined using their grade point average) were not affected by probability information whereas low performers studied significantly less time in the 25% condition compared to low performers in the 100% condition. In this case, the metacognitive inaccuracy is borne out of a failure to monitor the task appropriately. For example, In the extreme case, if a student believes that 50% of the material covered in the course will be on an upcoming exam, the student may reason that mastering 50% of the material will yield a good score and thus they will make a performance prediction that is likely much too high. A second study corroborated the laboratory results showing that low-performing students reported attempting to know the same amount of material that they thought would be tested, whereas high performers attempted to know more than they thought would be tested.

A statistical artifact account of the overconfidence effect has also been offered (Krueger & Mueller, 2002). The authors of this account argued that a regression artifact in addition to a better-than-average effect can explain the overconfidence effect without relying on cognitive, metacognitive, or motivational theories. But this statistical artifact hypothesis of overconfidence was challenged by Kruger and Dunning (2002). For example, Krueger and Mueller argued that test unreliability is one determinant of miscalibration, but calibration asymmetries (i.e., lower performers are much more miscalibrated than high performers) do not disappear when test unreliability is controlled. We have also suggested that low performers might not know how they will so they just guess an average grade, which is higher than their eventual grade (Miller & Geraci, 2011c).

So far, the focus has been on the finding that people, and particularly, low performers, tend to be overconfident, but there are other systematic distortions that occur under specific situations and with certain types of metacognitive judgments. For example, occasionally people are underconfident in their abilities, especially after practice. The underconfidence-with-practice (UWP) effect occurs when participants make JOLs after multiple study opportunities. The participants believe they have not learned as much as they actually have and report JOLs that underestimate recall performance (Koriat, Sheffer, & Ma'ayan, 2002). Another distortion is the hard-easy effect that occurs when retrospective confidence judgments for hard items are overestimated and confidence judgments for easy items are underestimated (Lichtenstein et al., 1982).

1.4.1 Summary

When people monitor their cognitive activity, by making JOLs and other judgments, there are several systematic distortions. Some of the distortions include overconfidence, the hard-easy effect, and the underconfidence-with-practice effect. But the reasons for metacognitive inaccuracy are not totally understood. Clearly, from the reasons presented here, there could be multiple sources of metacognitive inaccuracy. However, it is fair to conclude that the strongest version of the double-curse explanation of low performers' metacognitive inaccuracy lacks empirical support (Miller & Geraci, 2011c) as does the statistical artifact account (Kruger & Dunning, 2002). However, motivational issues, failures to understand probability information, study strategy use and possibly other reasons may contribute to metacognitive inaccuracy.

1.5 Methods to improve metacognition

Because of the benefits of accurate metacognition and the clear differences that exist between individuals' metacognitive ability in normal populations, researchers have attempted to improve or "train" metacognition. These attempts have resulted in varying degrees of success. One common outcome that researchers have referred to as the "Matthew effect," which occurs when an intervention designed to benefit low achieving students has a greater benefit for high achieving students (Kelemen et al., 2007; Hacker, et al., 2000). Although improving metacognitive monitoring alone is theoretically interesting, an applied research goal has been to improve monitoring with the goal of improving control and future educational outcomes.

1.5.1 Improving metacognitive monitoring accuracy in the laboratory

Attempts to improve metacognition in the laboratory have produced mixed results. Even before Flavell defined the concept of metacognition, researchers were attempting to improve participant's metacognition by training calibration ability. For example, in one early attempt to train metacognition, participants viewed word pairs with various encoding tasks (i.e., specifying the word pairs' relationship as synonyms, antonyms, or as unrelated) and then they were told to rate their confidence that their answers on the encoding task were correct on five successive days of sessions (Adams & Adams, 1958). Participants in the experimental condition were given feedback after each session that indicated the discrepancy between their confidence and performance for each item. The control condition was only shown a distribution of their confidence judgments. The calibration results indicated that the experimental condition participants

showed moderate improvement from session one to five while control condition participants did not.

In a slightly different training paradigm, participants read two-choice general knowledge questions, answered them, and also rated their confidence that their answer was correct in 23 1-hour sessions (Lichtenstein & Fischoff, 1980). Prior to the first session, participants were given detailed information about calibration. Calibration was illustrated by provided examples of well-calibrated and poorly-calibrated individuals. They were also told the goal of the experiment, which was to determine if calibration ability could be improved with multiple sessions. Moreover, after each session, summary reports that included each participant's performance, calibration, resolution and other information were given to the participant and explained by the experimenter for up to 20min. Results of Experiment 1 indicated that the participants began overconfident but most were able to improve their calibration; with the majority of the improvement occurring between the first and second sessions of feedback.

In another study, participants improved their calibration with repeated practice without feedback but only for a portion of the participants (Kelemen et al., 2007). In this study, participants were instructed to study different Swahili-English word pairs on five occasions and to indicate the likelihood that they would remember the English words. Results showed that by session 5, participants' predictions improved significantly relative to previous sessions. Therefore in this lab study, repeated practice making performance predictions did improve calibration, but importantly, it was only the high

achieving students (as measured by participant's SAT scores) who were able to increase their calibration.

While the previous two experiments were completed in multiple sessions across multiple days, other studies have asked participants to make predictions and postdictions over multiple sets of questions, but all in the same session (Pierce & Smith, 2001). A main finding of this research indicated that postdictions were significantly more accurate than predictions. Another finding that may be more relevant to the current section on improving metacognition is that there was no improvement from the first time making either monitoring judgments to the final time making the judgments.

These previous studies (Adams & Adams, 1958; Lichtenstein & Fischoff, 1980; Kelemen et al., 2007; and Pierce & Smith, 2001) all included multiple opportunities for participants to make monitoring judgments. In the majority of these studies, participants made judgments on multiple days. Another, commonality is that most used intense practice regimens. Lichtenstein and Fischoff commented that the training involved for Experiment 1 of the study was "both arduous and expensive," so much so that they completed a second experiment to determine if similar improvement would be seen with a shortened training program, from 23 1-hour sessions down to 11 sessions (pg. 166). Similar improvements in monitoring accuracy were also observed with the shortened program. Furthermore, at least in two of the three studies that showed improvement, intense feedback was involved. But given that Kelemen et al. did not use feedback and still showed improvement, the role of feedback cannot be determined. One component of

many of the previous studies and some to follow is the role that achievement plays in participants' success or failure in adjusting their performance predictions.

There are also other methods used to improve metacognitive calibration in the laboratory that are more easily accomplished. For example, in the laboratory one easy way to improve calibration is to simply increase the time between when the subject finishes study and when the prediction, or judgment-of-learning (JOL), is made (Nelson & Dunlosky, 1991). This improvement is known as the delayed-JOL effect. In the study that identified the accuracy superiority of delayed-JOLs, participants studied paired associates and were told that they would take a test 10 minutes after study on the paired associates. For half of the paired-associates, participants made an immediate JOL and on the other half they made the JOLs more than 30 seconds after studying (in order to exceed the length of short-term memory) with some intervening paired associates. Calibration, or absolute accuracy, for the delayed-JOLs was significantly better than calibration for the immediate JOLs. The authors concluded that delayed JOLs extremely accurate, which was in stark contrast to other research on JOLs. The advantage of delayed-JOLs over immediate JOLs has recently been confirmed through a meta-analysis involving more than 40 studies and more than 100 effect sizes (Rhodes & Tauber, 2011). The authors suggested that delaying a JOL confers an advantage over an immediate JOL because when participants made the immediate JOL, they were monitoring the contents of short-term memory in addition to the contents of long-term memory. The authors named this hypothesis the "monitoring-dual-memories principle" (MDM). The problem with monitoring the contents of both short- and long-term

memory of course is that participants were only able to tap the contents of long-term memory at the time of the memory test, therefore monitoring short-term memory adds "noise" to the judgment and could be a source of overly optimistic self-assessments (Nelson & Dunlosky). The reason delaying making a JOL improves resolution then is because the individual making the JOL is able to more accurately monitor the contents of long-term memory.

1.5.2 Improving metacognitive monitoring accuracy in the classroom

In classroom studies, metacognitive monitoring has proven difficult to modify. In a review of the literature, Hacker, Bol, and Keener (2008b) identified several attempts that have been made to improve metacognitive accuracy in the classroom using a variety of methods including giving students practice tests, practice making predictions, incentives, feedback, training and more. Yet, very few studies achieved their goal of improving students' monitoring accuracy. In one particular classroom study, student participants were asked to predict exam scores on each of three mid-term exams and one final comprehensive exam (Nietfeld et al., 2005). Students made both local (i.e., item-by-item) and global (for the entire exam) performance predictions. After each exam, students were encouraged to review their predictions, although no feedback or formal monitoring training was provided. Results showed that global monitoring was more accurate than local monitoring but that both types of monitoring actually decreased from exam 1 to 2. Based on the pattern of data, the authors concluded that self-directed feedback was not a sufficient intervention to improve students' metacognitive calibration.

Others have tried to improve metacognition in the classroom by providing practice and specific types of training on the value of accurate self-assessment. In one study, students made prediction and postdictions on each of three different exams (Hacker et al., 2000). Students were encouraged to make accurate self-assessments and were informed about the value of accurate self-assessments. They also completed practice exams prior to each exam to obtain more accurate feedback on the status of their knowledge. After the exams they were advised to reflect on their predictions and develop a plan to improve their accuracy. Under these conditions, students' predictions improved across exams while postdictions remained stable and consistently more accurate than predictions. When students were split into high and low performance groups (based on the percentage of total items answered correctly), results showed that the prediction improvement was carried by the high-performing students. Notably, even though prediction accuracy improved for the high-performing group, overall exam performance did not.

The authors offered reasons why high- but not low-performing students were able to improve their prediction accuracy. First, they suggested that students' use of feedback may vary according to the extent to which they externalize negative outcomes. When poor students receive negative feedback about the accuracy of an exam prediction, they might either use the feedback to recalibrate or attribute the outcome to an external factor such as bad instruction or a poorly constructed exam. Second, the authors suggested that the incentive used (motivation to graduate) may have only been effective for high-performing students. In a subsequent study, Hacker, Bol, and Bahbahani

(2008a) examined the role of attributional style and incentives for accuracy, this time providing increased course credit for more accurate judgments. But again, results showed that while postdictions improved predictions did not.

The reasons why classroom studies have shown that poor students cannot improve their metacognitive accuracy are unknown. One possibility is that low performing students do not improve their metacognitive accuracy because the nature of the feedback was simply too general for them to use. For example, in Hacker and colleagues (2008a), participants in the reflection condition were instructed to reflect on the accuracy of their judgments after receiving their calibration scores, but poor students may not be able to make use of this type of feedback or instruction. To address this issue, we attempted to improve metacognitive accuracy and exam performance for low and high performing students by providing tangible extra-credit incentives and concrete feedback for students. Our hypothesis was that providing immediate and tangible incentives in conjunction with concrete and specific feedback regarding how students could bring their predictions in line with their performance would lead both high- and low-performing students to improve their metacognitive accuracy (Miller & Geraci, 2011b). In both studies, participants were asked to make global predictions regarding the outcome of 4 different mid-term exams. We examined prediction calibration for each exam and whether calibration improved throughout the semester. Note that in previous work, improvements from the first exam to the second are not always evaluated for methodological reasons (see Hacker et al., 2008a), even though one might expect the biggest improvements early in the course (Lichtenstein & Fischoff, 1980). We also

examined whether students' performance improved. We predicted that giving students practice and concrete feedback predicting their own grades would lead them to become more proficient at self-monitoring and possibly better students. We also asked students to complete a questionnaire at the end of the course to determine whether students were using the feedback appropriately and what their general strategies were for incorporating the feedback they received.

In Study 1 feedback that was provided to students about their prediction accuracy was minimal but it served as a baseline for Study 2 in which we used the same extra credit incentives for accurate predictions but also provided more explicit, concrete feedback to students regarding their prediction accuracy. The common result from both studies was that students were overconfident, low-performing students even more so than high-performing students. As such, the findings were consistent with the literature showing that people are mostly overconfident in their self-assessments (e.g., Dunning, Heath, & Suls, 2004; Kelemen et al., 2007; Kruger & Dunning, 1999). Study 1 showed that when students had the opportunity to earn extra credit for accurate predictions and were given feedback regarding their performance, they were not able to improve their metacognitive calibration. In Study 2, when feedback was made more explicit and concrete, low-performing students improved their calibration from exam 1 to exam 2. However, we did not see any improvement in exam performance. Post-exam questions indicated that students used the feedback appropriately, suggesting that the failure to find improved exam performance was not a result of students failing to attend to the feedback.

There are even easier ways to improve metacognitive monitoring that have been identified in the retrospective confidence literature that could be applied to performance predictions in the classroom or JOLs in the laboratory. These simple techniques to debias retrospective confidence judgments are known as response-oriented modifications. For example, given that low performers are consistently overconfident, one method to debias their judgments would be to simply tell them to lower their predictions. All that participants are required to do in order to improve their accuracy is to artificially downgrade their prediction (Keren, 1990). But these sorts of debiasing techniques do not force the individual to think critically about errors in monitoring judgments. In contrast, process-oriented modifications, encourage participants to re-think the way such judgments are made. As such, the hope is that improvements made as a result of process-oriented modifications are likely to persist and generalize to other situations whereas response-oriented modifications would not.

1.5.3 Improving metacognitive monitoring accuracy to improve metacognitive control effectiveness

In order for improved metacognitive monitoring to be useful to the individual in applied settings, the improved monitoring must also lead to improved metacognitive control. Indeed, one reason accurate metacognitive monitoring is beneficial is because monitoring and self-regulated learning are intimately connected such that better monitoring leads to more effective control and better performance (Nelson et al., 1994). One clear example of this relationship was shown when accurate metacognition was associated with better academic performance (Everson & Tobias, 1998). In this study,

researchers assessed the monitoring ability of incoming college freshman students and compared monitoring ability to their GPA and the end of the semester. In the assessment, students were first asked to identify words they knew and did not know from a word list and then were asked to take an objective test on the same words. Results indicated that most calibrated students also had the best GPAs.

Even stronger evidence for the link between accurate monitoring and more effective control and improved test performance comes from experimental research. For example, Thiede, Anderson, and Therriault (2003) manipulated monitoring accuracy by asking participants to generate keywords about expository texts immediately after reading, after a 5-min delay or not at all. Afterwards, all participants took a comprehension test and then were allowed to self-select and reread texts of their choice; rereading was followed by another test. Participants who generated keywords after a delay had better monitoring accuracy and were better able to regulate their study by choosing and rereading texts appropriately. This improved control also conferred an advantage on the test for the high monitoring group. Similarly, Nelson et al. (1994) showed that accurate monitoring leads to effective control. In this case, control was measured as allocation of study time. Participants studied 36 Swahili-English word pairs and made item-level delayed-JOLs following the study period. Following study and JOL trials, 18 (out of 36) of the original items were restudied. The between subjects manipulation determined the 18 items that were restudied. Participants either restudied the subjectively best-learned items (i.e., the items they had given the highest JOLs), the subjectively worst-learned items (the lowest JOLs), the objectively most difficult items

(based on normative ratings) or 18 items that the individual participants chose to restudy. Following the first restudy session, participants took a memory test on all of the items. Following this first cycle, each participant completed another 5 restudy-test trials. An important feature of this study is that the 18 items that were selected for restudy originally were restudied throughout the session. Because allowing the participants to select what items should be restudied led to improved recall, the authors concluded that participants' original JOLs were accurate. Choosing the subjectively worst-learned items (the lowest JOLs) was also significantly more effective for guiding participants' allocation of study time and recall performance compared to restudying the normatively most difficult items.

Others have manipulated monitoring accuracy in the classroom to improve performance (Nietfeld, Cao, & Osborne, 2006). Participants in the experimental classroom were given instructions to complete a monitoring worksheet after each class, one class per week for 16 weeks. The purpose of the worksheets was so that participants could assess their own understanding of the material, identify concepts they found difficult to understand and what they would do to understand these difficult concepts, and finally, the worksheet contained three practice questions for students to answer. For the three practice items, participants also reported a confidence judgment regarding their answer. Participants in the control classroom did not complete these monitoring worksheets. The results indicated that performance predictions on the exams became more calibrated with time for participants in the experimental condition but not the

control condition. Importantly, exam performance also improved for experimental condition participants.

Another strong piece of evidence linking increased accuracy of monitoring judgments to more effective self-regulation of study time and improved test performance was completed by Thiede (1999). Participants in this study studied Swahili-English word pairs and, after viewing all of the word pairs, made a JOL for each pair. Following the JOL, participants took a cued-recall memory test. Participants then reported JOLs for each item again and were given the opportunity to restudy as many items as they wanted of their choosing. Participants were told that the experiment would end only when all 36 word-pairs were recalled. Restudy was followed by another memory test. The JOL-restudy-test cycle was repeated until all 36 word pairs were recalled. The results indicated that the participants with the most accurate monitoring (strong positive correlation between JOL and recall) and the most effective control (strong negative correlation between JOL and restudy) also had the best recall test performance.

Recent work has also suggests that testing may be a critical factor for learning because of the metacognitive information that it can provide. Karpicke and Roediger (2008) showed that multiple retrieval opportunities enhanced participants' long-term retention of Swahili-English word pairs. In their study, asking participants to take a test, or practice retrieving word meanings from memory, enhanced long-term retention even more so than additional study – a finding that is commonly referred to as the "testing effect." The testing effect is relevant here because if students tested themselves while they studied for an exam it would provide valuable information about how well they

knew the material. Students could then use this information to inform their decisions about what material to restudy and what material to discontinue studying (see Dunlosky, Rawson, & McDonald, 2002). Unfortunately though, most students do not practice retrieval and most are not even aware of the benefits of practice retrieval (Karpicke, Butler, & Roediger, 2009). Furthermore, when students are given the option to practice retrieval, they most often choose not to (Karpicke, 2009).

When participants are trained to regulate their study by testing themselves, their monitoring and performance improves (Dunlosky et al., 2003). In this study, older adults, whose monitoring ability is spared or sometimes better than younger adults, were trained over multiple sessions about how to use self-testing or how to use other study strategies (i.e., imagery) to help them learn paired-associates. As predicted, the older adults who were trained to use self-testing while learning had superior memory performance compared to older adults who were trained to use other study strategies while learning.

1.5.4 Summary

Accurate monitoring is associated with improved educational outcomes, in part because accurate monitoring leads to more effective control. Attempts to improve monitoring have been carried out in the laboratory and in naturalistic settings. Results from both types of studies have yielded mixed results, classroom studies even more so than laboratory studies. Although the methods vary and results are mixed there are at least a few common themes among most the previous studies. For example, attempts to improve metacognition in the classroom represent relatively long interventions. In our

study, the intervention involved efforts across an entire college semester (Miller & Geraci, 2011b). Other common themes are multiple opportunities to predict performance, feedback, and incentives. In contrast, for improved monitoring accuracy to occur in the lab, somewhat less intensive methods have been utilized (i.e., delayed-JOLs). Methods that improve monitoring via delayed-JOLs do so by limiting the focus of the judgment to long-term memory (Nelson & Dunlosky, 1991).

1.6 Conclusion

Metacognition has been a topic psychologists have been interested in since the beginning of experimental psychology in the late 19th century. In fact, given that the early Greek orators were aware of the limitations of their cognitive process, one could say that metacognition is an ancient topic. Today there are countless research areas in metacognition. In the area of metacognitive predictions about future performance, Nelson and Narens' (1990) theory of metacognition implied that an individual's metacognitive control can only be effective when the information it receives via monitoring processes is accurate. And research has indicated that more accurate monitoring leads to more effective control (e.g., Thiede et al., 2003). Unfortunately, many people's monitoring accuracy is biased (e.g., Alicke, 1985; Buehler, Griffin, & Ross, 1994; Burson et al., 2006; Kelemen et al., 2007; Knouse, Bagwell, Barkley, & Murphy, 2005; and others). Some have even claimed that overconfident low performers suffer a double-curse, or that their behavior is akin to individuals with brain damage (Dunning, 2005, Kruger & Dunning, 1999). More often than not, poor self-regulation leads to deficient performance, as in the case above, poor regulation of study behavior

leads to poor academic performance (cf. Everson & Tobias, 1998). Of course the opposite scenario is also true; people who have accurate metacognitive monitoring are better equipped to control their study which then leads to improved performance (Nelson et al., 1994, Thiede et al., 2003).

Because of the link between accurate monitoring, effective control, and improved performance, many researchers have attempted to improve the first link in the causal chain. That is, researchers have attempted to improve metacognitive monitoring with the hope that doing so would also benefit control and performance (Hacker et al., 2000; Nietfeld et al., 2006; and others). However, laboratory and classroom studies have indicated that metacognitive monitoring is resistant to intervention (e.g., Lichtenstein & Fischoff, 1980; Miller & Geraci, 2011b). Two factors that appear to influence monitoring positively are multiple opportunities to make monitoring judgments over long periods of time with explicit feedback (Miller & Geraci). Incentives for accuracy and instructions for participants to reflect on their judgments have led to mixed results (Hacker et al., 2008a). More recently, Miller and Geraci (2011c) showed that even though low performers' monitoring judgments are less accurate than high performers, they have some awareness that their predictions are inaccurate. Low performer's awareness of their inaccuracy reveals a possibility for improving metacognition that has yet to be tested – whether or not participants can use subjective confidence as a guide to recalibrate and improve their metacognitive judgments.

2. EXPERIMENTS

Nelson and Narens' (1990) theory of metacognition implied that the quality of metacognitive control is contingent on the quality of metacognitive monitoring processes. Research has indicated that more accurate monitoring leads to more effective control (e.g., Thiede et al., 2003). The purpose of current research is to examine whether participants can use confidence in their predictions to recalibrate subsequent performance predictions. The purpose of Experiment 1 was to establish that metacognitive monitoring improvement could be accomplished using subjective confidence as a guide. To do this, participants made performance predictions about an upcoming memory test and reported their confidence that the prediction was accurate. Participants then adjusted their performance predictions so that they could be more confident their prediction was accurate. Calibration in this condition was compared to calibration in the control condition in which participants did not rate their confidence in their predictions—they simply made the prediction twice. The purpose of Experiment 2 was to determine if continuous improvement in performance predictions was possible by using subjective confidence as a guide. That is, not just improvement from the first to second prediction but from the second to third and so on. Participants adjusted their performance prediction 3 times – each time indicating their level of confidence that their prediction was accurate. Calibration in this condition was compared to calibration in a control condition in which participants made repeated predictions but did not indicate confidence in these predictions. And finally, the purpose of Experiment 3 was to

examine whether participants' improved performance predictions would also influence, and improve, control of study time and performance. Participants studied paired associates with the goal of remembering at least 15 out of 20. Following study, participants made predictions about future memory performance and rated their confidence that the prediction was accurate. Then participants made the decision to restudy the items or take the memory test. The total number of times and total time participants chose to study the paired associates in this condition was compared to the control condition where participants made predictions about future memory performance but did not rate confidence.

2.1 Experiment 1 – improving monitoring accuracy

Experiment 1 was designed to serve as proof of concept – to examine whether or not participants could use subjective confidence reports as a guide to recalibrate performance predictions.

2.1.1 Method

Design. Experiment 1 used a 2 Condition (Experimental and Control) X 2 Performance Prediction (Original and Adjusted1) mixed randomized repeated model. Condition was the randomized between subjects independent variable and Performance Prediction was the repeated measures independent variable. The dependent variable of interest was calibration score for the original performance prediction and the adjusted performance prediction. Calibration was calculated by subtracting recall performance from prediction. As such, positive values indicated overconfidence and negative values indicate underconfidence. Other variables, including demographic variables and

vocabulary ability were also measured using the Shipley vocabulary test (Zachary, 1986).

Participants. Participants were 172 undergraduate students from Texas A&M University who participated for partial course credit. The sample of undergraduate student participants was 55% female and largely from the freshman and sophomore class (*M* education in years = 13.52, *SE* = 0.10). Ethnicities represented in the sample were European American (72%), Hispanic (17%), African American (4%), Asian (5%) or other (2%). The mean age of participants was 19.57 (*SE* = 0.11) years and the mean vocabulary score was 29.81 (*SE* = 0.28). There were no between group differences in age ($F(1, 168) = 0.11$, $MSE = 2.02$, $p = .74$, $\eta^2_p < .01$) or education ($F(1, 170) = 0.65$, $MSE = 1.57$, $p = .42$, $\eta^2_p < .01$). However, although there was less than a two-item mean difference in vocabulary ability, experimental condition participants (*M* vocabulary = 30.51, *SE* = 0.39) had significantly greater vocabulary scores than control condition participants ($M = 29.10$, $SE = 0.39$, $F(1, 170) = 6.45$, $MSE = 13.10$, $p = .01$, $\eta^2_p = .04$). When participants' vocabulary scores are included as a covariate in the main analyses, the interaction term, which is the critical effect in Experiment 1, remains statistically significant.

Materials. Swahili-English paired associates were taken from Nelson and Dunlosky (1994) (see Appendix for the sample of paired associates used in the following experiments). All of the paired associates produced less than 52% correct recall after 3 learning trials, the English words had a frequency of occurrence of 50-100 words per million, and the Swahili words were rated as a 2-3 in wordlikeness on a 1-5 scale where

1 meant “not like a word at all” and 5 meant “very like a word” (see Nelson & Dunlosky).

Procedure. After giving consent to participate, participants studied 20 Swahili-English paired associates that were presented via computer and were on screen for 6 seconds per paired associate. Following study all participants were told that they would take a cued-recall test in approximately five minutes and that they should make a performance prediction as a percentage (i.e., 0-100%). After making the prediction, participants in the experimental condition rated their confidence that their performance prediction was accurate on a scale of 1-10, where 10 indicated absolute confidence the prediction was accurate and 1 indicated no confidence the prediction was accurate. Participants in the control condition did not rate their confidence that the prediction was accurate. After making the performance prediction (and rating confidence in the experimental condition), participants made a second performance prediction. Participants were instructed to make a second prediction with verbal and written instructions that the second prediction should be as accurate as possible but could go up, down, or stay the same. In the experimental condition, participants also made a second confidence judgment and received and instruction that their confidence in the second performance prediction should increase. After participants made their second prediction there was a 5-min retention interval. During the retention interval participants took a vocabulary test (Zachary, 1986).

After the retention interval participants took a cued-recall memory test in which the Swahili word was provided and they wrote the English equivalent. Participants were

told they would have 5-min to complete the memory test but that additional time would be given if needed. Following the memory test, participants completed a demographics questionnaire and answered 3 post-questions where they reported 1) why they thought they were asked to adjust their performance prediction, 2) what their thought processes were while adjusting their prediction, and 3) the reason why they lowered, raised, or kept their performance prediction.

2.1.2 Results

Original and adjusted calibration scores were calculated by subtracting each participant's performance from their original prediction. From this calculation, positive values indicate overconfidence and negative values indicate underconfidence. My prediction that original calibration scores between conditions would be equivalent was verified ($F(1, 170) = 3.51$, $MSE = 432.33$, $p = .06$, $\eta^2_p = .02$) as was the prediction that performance between conditions would be equivalent ($F(1, 170) = 1.86$, $MSE = 165.54$, $p = .18$, $\eta^2_p = .01$, see Tables 1 for means and standard errors).

Related to the purpose of the current experiment – to improve metacognition – results from the mixed randomized repeated measures ANOVA indicated main effects of calibration ($F(1, 170) = 9.32$, $MSE = 92.35$, $p = .003$, $\eta^2_p = .05$) and condition ($F(1, 170) = 7.16$, $MSE = 777.24$, $p = .008$, $\eta^2_p = .04$) as well as a significant interaction ($F(1, 170) = 4.13$, $MSE = 92.35$, $p = .04$, $\eta^2_p = .02$, see Figure 1).

That is, participants' adjusted performance predictions (M = 14.74, SE = 1.64) were significantly more accurate than their original performance predictions (M = 17.90, SE = 1.60). Participants in the experimental condition were also significantly more calibrated (M = 12.30, SE = 1.94) than control condition participants (M = 20.34, SE = 2.29) in general. But both main effects were qualified by the significant interaction effect. Planned comparisons showed that the accuracy of participants' predictions in the experimental condition improved significantly from original (M = 14.93, SE = 2.14) to adjusted prediction (M = 9.66, SE = 2.14, $t(85) = 2.92$, $p = .004$, $d = 0.32$) whereas participants in the control condition had equivalent original (M = 20.87, SE = 2.34) and adjusted predictions (M = 19.81, SE = 2.36, $t(85) = 1.04$, $p = .30$, $d = 0.11$). Participants in the experimental condition had higher confidence in their adjusted performance predictions (M = 9.12, SE = 0.17) than their original predictions (M = 6.20, SE = .20, $t(85) = 15.46$, $p < .001$, $d = 1.67$).

Table 1.

Experiment 1 mean recall, predictions, calibration for all predictions, and confidence by condition

Condition	Recall	Original Prediction	Adjusted Prediction	Original Calibration	Adjusted Calibration
Control	13.72 (1.42)	34.59 (2.29)	33.54 (2.26)	20.87 (2.34)	19.81 (2.36)
Experimental	16.40 (1.35)	31.44 (1.94)	26.17 (2.04)	14.93 (2.14)	9.66 (2.14)
Confidence		6.20 (0.20)	9.12 (0.17)		

Note. Recall and predictions are expressed as a percentage of the total items studied (20). Participants reported confidence on a scale of 0-10. Calibration is calculated by subtracting performance from prediction. Standard errors shown in parentheses.

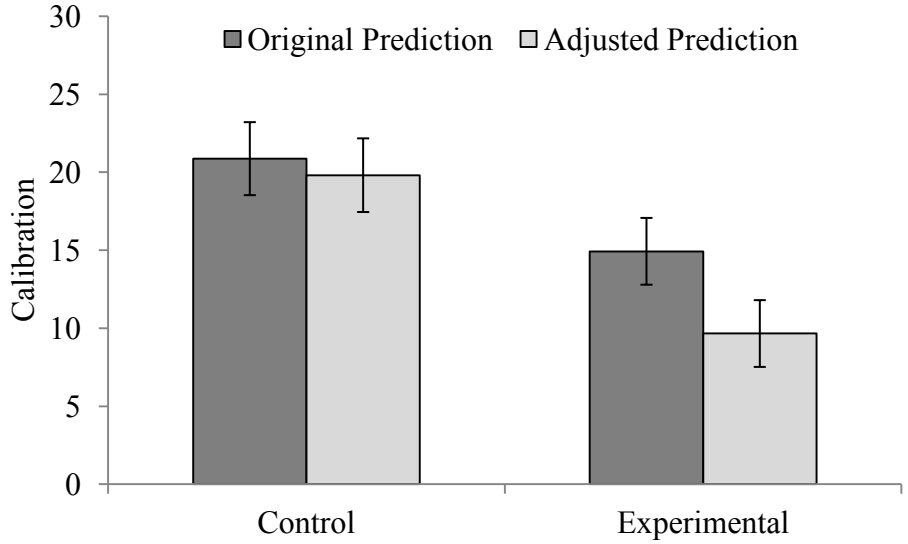

Figure 1. Experiment 1 calibration scores for original and adjusted predictions by condition.

Calibration results were also examined in separate mixed randomized repeated measures ANOVAs as a function of low and high performing groups based on recall performance. High performing participants' calibration scores improved regardless of condition ($F(1, 84) = 6.28$, $MSE = 64.98$, $p = .01$, $\eta^2_p = .07$) as indicated by a significant main effect of calibration. The condition main effect was also significant due to experimental condition participants being more calibrated in general ($F(1, 84) = 5.15$, $MSE = 913.36$, $p = .03$, $\eta^2_p = .06$). The interaction term was not significant ($F(1, 84) = 1.18$, $MSE = 64.98$, $p = .28$, $\eta^2_p = .01$). On the other hand, the main effect of calibration for low performing participants' was statistically non-significant ($F(1, 84) = 3.73$, $MSE = 121.30$, $p = .06$, $\eta^2_p = .04$). Similarly, the interaction term did not reach significance ($F(1, 84) = 2.92$, $MSE = 121.30$, $p = .09$, $\eta^2_p = .03$). Finally, the between-subjects condition main effect was statistically non-significant ($F(1, 84) = 2.53$, $MSE = 540.57$, $p = .12$, $\eta^2_p = .03$). At least on visual inspection of the low performers' calibration results, it appears that low performers were able to gain some calibration accuracy in their adjusted predictions after having first reported subjective confidence.

2.1.3 Summary

The important finding from Experiment 1 is that when participants were instructed to use subjective confidence about the accuracy of their original prediction when making a second (adjusted) prediction, their predictions became significantly more accurate. In contrast, control condition participants' adjusted predictions were not more or less accurate than their original predictions. Given that participants were only able to adjust their predictions one time, it is not known whether participants were maximally

calibrated or if there could be continuous improvement in calibration given the opportunity to adjust predictions multiple times.

2.2 Experiment 2 – continuous improvement

The purpose of Experiment 2 was to identify whether or not the calibration improvement seen in Experiment 1 would continue when participants were asked to adjust their performance predictions more than once using subjective confidence as a guide. Given this purpose, the design, materials, and procedure were similar to Experiment 1 with a few key exceptions described below.

2.2.1 Method

Design. Experiment 2 used a 2 Condition (Experimental and Control) X 4 Performance Prediction (Original, Adjusted1, Adjusted2 and Adjusted3) mixed randomized repeated model. As in Experiment 1, Condition served as the randomized between subjects independent variable and Performance Prediction served as the repeated measures independent variable. The dependent variable of interest was the calibration score for the original performance prediction and the calibration scores for the subsequent adjusted performance predictions. Other variables, including demographic variables and vocabulary ability were also measured.

Participants. Participants were 140 undergraduate students from Texas A&M University who participated for partial course credit. The sample of undergraduate student participants was 61% female and largely from the freshman and sophomore class (*M* education = 12.77, *SE* = 0.07). Ethnicities represented in the sample were European American (72%), Hispanic (2%), African American (9%), Asian (1%) and other (14%),

three participants did not report ethnicity. The mean age of participants was 18.82 years (*SE* = 0.08) and the mean vocabulary ability was 29.34 (*SE* = 0.30). To anticipate, there were no between group differences in age ($F(1, 138) = 3.12$, $MSE = 0.83$, $p = .08$, $\eta^2_p = .02$) or vocabulary ($F(1, 138) = 0.51$, $MSE = 12.78$, $p = .48$, $\eta^2_p < .01$). Participants in the control condition had higher education levels ($M = 12.99$, $SE = 0.10$) than participants in the experimental condition ($M = 12.55$, $SE = 0.10$) ($F(1, 138) = 9.83$, $MSE = .70$, $p = .002$, $\eta^2_p = .07$). When years of education were included as a covariate in the main analyses, the pattern of findings is the same compared to when no covariates are included.

Materials and procedure. The materials used in Experiment 2 were exactly the same the materials used in Experiment 1. The procedure was similar but included a key modification to ask participants to make more than one adjusted performance prediction. Participants studied 20 Swahili-English paired associates that were presented via computer and were on screen for 6 seconds per paired associate. Following study, all participants were told that they would take a cued-recall test in five minutes and were asked to make a performance prediction as a percentage (i.e., 0-100%). Participants in the experimental condition were then be asked to rate their confidence that the performance prediction is accurate. These participants were then asked to change their performance prediction so that they could be more confident their prediction was accurate. They were also given the additional instruction that their "adjusted performance predictions should be as accurate as possible and can go up, down, or stay the same, but if you are following the instructions, your confidence in the new prediction

should increase." After this additional instruction participants made an adjusted performance prediction and rated their confidence in the new prediction. This cycle, (i.e., prediction-confidence) was repeated two more times for participants in the 3 adjustment condition. Participants in the control condition were asked to make an original performance prediction and three adjusted performance predictions without having made any confidence ratings. Note that they were also told "adjusted performance predictions should be as accurate as possible and can go up, down, or stay the same."

2.2.2 Results

The purpose of Experiment 2 was to determine if improved calibration that was seen in Experiment 1 continues when participants adjust their predictions multiple times. Thus, the principal hypothesis was that participants' adjusted performance predictions would be significantly more calibrated than their original performance predictions in the experimental but not the control condition.

As before, calibration scores were calculated by subtracting each participant's performance from their prediction. Thus, positive calibration scores indicate overconfidence. Results indicated no differences between groups for original calibration scores ($F(1, 138) = 0.05$, $MSE = 513.42$, $p = .83$, $\eta^2_p < .01$) or recall performance ($F(1, 138) = 1.67$, $MSE = 146.86$, $p = .20$, $\eta^2_p = .01$, see Tables 2 and 3). To the point of the experiment, a repeated measures ANOVA indicated a significant Calibration X Condition interaction ($F(1.72, 237.61) = 6.18$, $MSE = 91.78$, $p < .004$, $\eta^2_p = .04$; see Figure 2). Main effects of calibration ($F(1.72, 237.61) = 2.24$, $MSE = 91.78$, $p = .12$, $\eta^2_p = .02$) and condition ($F(1, 138) = 0.74$, $MSE = 2034.02$, $p = .39$, $\eta^2_p = .01$) were non-

significant. Note that due to a violation of sphericity, the error term degrees of freedom were adjusted using the Huynh-Feldt adjustment (Huynh & Feldt, 1976) for the previous results and future results in Experiment 2. To follow up the significant interaction, separate repeated measures ANOVAs for the control and experimental conditions were run. For the experimental condition, the simple effect of calibration was significant ($F(1.55, 106.62) = 7.87$, $MSE = 101.10$, $p = .002$, $\eta^2_p = .10$). Experimental condition participants' final predictions were significantly more accurate than their first and second predictions ($p = .002$ and $p = .02$ respectively). Furthermore, participants' second and third predictions were significantly more accurate than their first predictions ($p = .005$ for both comparisons). In contrast, the simple effect of calibration for control condition participants was not significant ($F(1.74, 120.37) = 0.49$, $MSE = 88.35$, $p = .59$, $\eta^2_p = .01$). Confidence among experimental condition participants was also significantly greater for each subsequent prediction ($F(1.88, 129.89) = 125.61$, $MSE = 1.19$, $p < .001$, $\eta^2_p = .65$).

Table 2.

Experiment 2 mean recall, original and all adjusted predictions, and confidence for the experimental condition

Condition	Recall	Original Prediction	Adjusted Prediction 1	Adjusted Prediction 2	Adjusted Prediction 3
Control	15.64 (1.70)	34.02 (2.67)	34.76 (2.84)	35.02 (2.86)	35.46 (3.09)
Experimental	13.00 (1.14)	32.20 (2.55)	28.95 (2.48)	27.94 (2.53)	26.50 (2.55)
Confidence		7.11 (0.30)	8.93 (0.30)	9.39 (0.20)	9.68 (0.20)

Note. Recall and predictions are expressed as a percentage of the total items studied (20). Participants reported confidence on a scale of 0-10. Standard errors shown in parentheses.

Table 3.

Experiment 2 calibration for original and all adjusted predictions

Condition	Original Calibration	Adjusted Calibration 1	Adjusted Calibration 2	Adjusted Calibration 3
Control	18.38 (2.85)	19.11 (2.98)	19.38 (2.92)	19.81 (3.10)
Experimental	19.20 (2.56)	15.95 (2.57)	14.94 (2.65)	13.50 (2.69)

Note. Calibration is calculated by subtracting performance from prediction. Standard errors shown in parentheses.

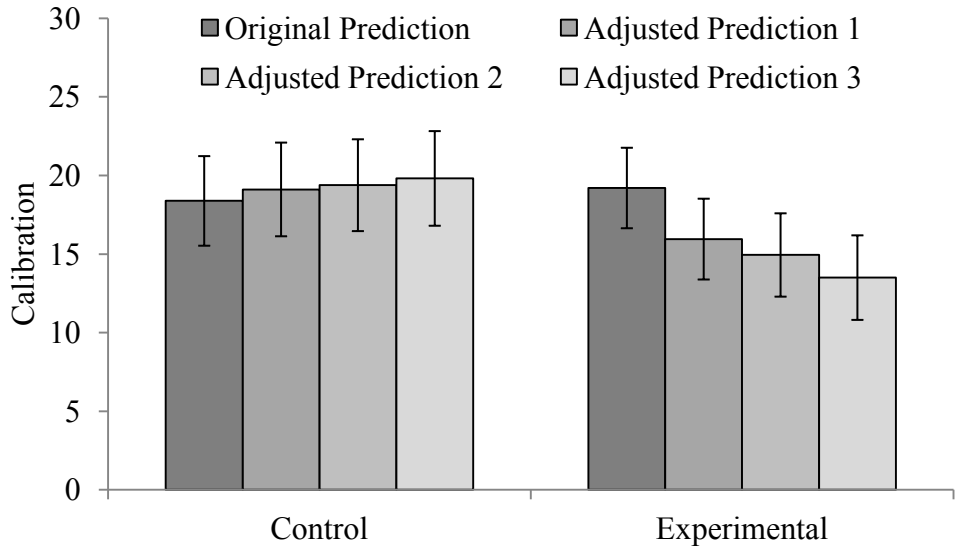

Figure 2. Experiment 2 calibration scores for original and all adjusted predictions by condition.

Again, one can also analyze the data as a function of low and high performance groups using separate repeated measures ANOVAs. For high performing participants there was no effect of condition ($F(1, 68) = .17$, $MSE = 2318.82$, $p = .68$, $\eta^2_p < .01$) or calibration ($F(1.80, 122.53) = 2.43$, $MSE = 111.92$, $p = .10$, , $\eta^2_p = .04$). However, there was a significant interaction ($F(1.80, 122.53) = 6.28$, $MSE = 111.92$, $p = .003$, $\eta^2_p = .09$), such that experimental condition participants improved their prediction accuracy but control condition participants did not. Follow-up analyses on the significant interaction indicated a significant effect of calibration ($F(1.69, 57.50) = 8.13$, $MSE = 120.15$, $p = .001$, $\eta^2_p = .19$) in the experimental condition. And post-hoc analyses confirmed that participants' final adjusted predictions were more accurate than the original and second performance predictions ($p = .005$ and $p = .02$ respectively) and participants' third and second predictions were more accurate than the original prediction ($p = .005$ and $p = .01$).

For low performers there were no main effects of condition ($F(1, 68) = .72$, $MSE = 1676.94$, $p = .40$, $\eta^2_p = .01$) or calibration ($F(1.64, 111.71) = .19$, $MSE = 63.94$, $p = .78$, $\eta^2_p < .01$). Nor was there a significant interaction ($F(1.64, 111.71) = .82$, $MSE = 63.94$, $p = .42$, $\eta^2_p = .01$). These results confirm that low performers, even when they used subjective confidence as a guide to adjust their performance predictions, could not become more metacognitively accurate.

2.2.3 Summary

Using the same materials and procedures as Experiment 1 with the key modification that participants made 3 adjusted performance predictions rather than 1, the

results showed that continuous improvement in metacognitive calibration is possible when participants focus on their subjective confidence as a guide to recalibrate their predictions. The largest calibration improvement occurred between the first prediction and second predictions, but the final prediction was also more accurate than the second prediction. When participants were split into low and high performing groups based on recall, high performing participants seemed to be more capable of using subjective confidence as a guide to recalibrate than low performing participants.

2.3 Experiment 3 – influence on control

The purpose of the final experiment was to examine the influence of improved metacognitive monitoring on metacognitive control. The Nelson and Narens (1990) model of the metacognition posits that metacognitive control is downstream of metacognitive monitoring. The effectiveness of metacognitive control is contingent on the accuracy of metacognitive monitoring. Therefore, given that the results from the previous two experiments indicated that participants could improve their metacognitive monitoring by making subjective confidence judgments, one would predict that there would be corresponding improvements in metacognitive control. In the current design, metacognitive control was operationalized as the decision to restudy to-be-remember material and overall allocation of study time. When participants decided to restudy the list, they were required to restudy the whole list.

2.3.1 Method

Design. Experiment 3 used a between subjects design in which participants were randomly placed in either the Experimental or Control condition. The dependent

variables of interest were the number of times participants decided to restudy the TBR material, overall study time, and recall performance. Participants' performance prediction calibration and other variables, including demographic variables and vocabulary ability were also measured.

Participants. Participants were 140 undergraduate students from Texas A&M University who participated for partial course credit. The sample of undergraduate student participants was 51% female and largely from the freshman and sophomore class (*M* education = 13.28, *SE* = 0.08). Ethnicities represented in the sample were European American (70%), Hispanic (20%), African American (3%), Asian (6%) and other (1%), one participant did not report ethnicity. The mean age of participants was 19.01 years (*SE* = 0.09) and the mean vocabulary ability was 29.37 (*SE* = 0.32). There were no between-group differences in age, vocabulary, or education.

Materials and procedure. Again the materials in Experiment 3 were the same as the materials used in the previous two experiments. The procedure was modified significantly. Prior to study, participants were informed that their goal of study, also known as the norm-of-study, was to be able to remember at least 15 out of 20 items on the memory test. Participants were also told that they should only stop studying when they believed they could recall at least 15 items on the memory test. To motivate participants to comply with this request, participants were told they would be given an unspecified prize at the end of the experiment for reaching the goal. Similar to previous experiments, participants in Experiment 3 studied the paired associates at a pace of 6 seconds per paired associate. After each time a participant studied the paired-associates,

he or she reported a performance prediction on a scale of 0-20 items. Participants in the experimental condition also rated their confidence that their performance prediction was accurate. Following the performance prediction (and confidence rating in the experimental condition) all participants were given the choice to restudy the paired-associates or to continue to the memory test. If the participant chose to restudy the associates, he or she studied all items in the study-list with the added option to manually advance the study list or allow the presentation software to automatically advance after 6secs, whichever occurred first. Following restudy, participants were given the same choice as before, to either restudy or take the test. There was no maximum limit on the number of restudy sessions. When participants chose to take the test, he or she took the memory test after an approximately 5-min retention interval (as in the previous experiments).

2.3.2 Results

The purpose for Experiment 3 was to examine how reporting subjective confidence after a performance prediction would influence metacognitive control. The principal hypothesis was that participants in the experimental condition, who rated their subjective confidence in addition to making a performance prediction, would choose additional study time more frequently than participants in the control condition. Given the increased study time, I also predicted participants in the experimental condition would have higher recall performance on the memory test compared to control condition participants.

There were no differences between groups for participants' first predictions (F(1, 138) = .10, MSE = 1.21, p = .75, η^2_p < .01, see Table 4). Note that in the previous experiments, I compared participants' first calibration scores, but in Experiment 3, because participants were able to study the items as many times as they chose to, calibration would not be the most appropriate comparison to ensure that the groups were similar to begin the experiment. The main analyses of Experiment 3 indicated no effect of condition on participants' decision to continue study. Participants in the experimental condition (M study sessions = 3.36, SE = 0.23) did not study any more than participants in the control condition (M = 3.61 SE = 0.26; F(1, 138) = .53, MSE = 4.37, p = .47, η^2_p < .01, see Table 5). Accordingly, total time studying was not different between the experimental condition (M study time (min) = 5.66, SE = 0.39) and the control condition (M = 6.09, SE = 0.41; F(1, 138) = .59, MSE = 11.26, p = .45, η^2_p < .01). Likely because there was very little difference in total number of study sessions or study time, there were no between-group differences in memory performance between the experimental condition (M recall = 9.36, SE = 0.67) and the control condition (M = 10.26, SE = 0.70; F(1, 138) = .87, MSE = 32.50, p = .35, η^2_p < .01).

Table 4.
Experiment 3 mean recall, original and final predictions, confidence for the experimental condition, with calibration for original and final predictions

Condition	Recall	Original Prediction	Final Prediction	Original Calibration	Final Calibration
Control	51.29 (3.49)	39.00 (2.05)	12.83 (0.47)	-12.29 (4.06)	12.86 (3.09)
Experimental	46.79 (3.33)	38.07 (2.12)	12.63 (0.41)	-8.71 (4.10)	16.36 (2.94)
Confidence		5.99 (0.26)	6.54 (0.22)		

Note. Recall, original prediction, and final prediction are expressed as a percentage of the total items studied (20). Calibration is calculated by subtracting performance from prediction. Standard errors are shown in parentheses.

Table 5.
Experiment 3 total number of study sessions and total study time (min)

Condition	Study Sessions	Study Time (min)
Control	3.61 (0.26)	6.09 (0.41)
Experimental	3.36 (0.23)	5.66 (0.39)

Note. Standard errors are shown in parentheses.

If performance group (as determined by recall) is included in the analysis, making it a 2 (Condition: experimental and control) x 2 (Performance group: low and high), condition still did not influence participants' decision to study more by total number of sessions ($F(1, 136) = .62$, $MSE = 3.75$, $p = .43$, $\eta^2_p < .01$) or by time ($F(1, 136) = .68$, $MSE = 9.73$, $p = .41$, $\eta^2_p < .01$). In contrast, high performers did choose to study more times ($F(1, 136) = 24.79$, $MSE = 3.75$, $p < .001$, $\eta^2_p = .15$) and for more total time ($F(1, 136) = 23.27$, $MSE = 9.73$, $p < .001$, $\eta^2_p = .15$) than low performers overall. However the interaction effects were non-significant, showing that low and high performers were similarly affected by condition for number of study sessions ($F(1, 136) = .12$, $MSE = 3.75$, $p = .73$, $\eta^2_p < .01$) and total time studying ($F(1, 136) = .36$, $MSE = 9.73$, $p = .55$, $\eta^2_p < .01$).

2.3.3 Summary

Nelson and Narens' (1990) model of metacognition contains two processes – monitoring and control processes. They argued that monitoring processes monitor ongoing cognitive activity and allow people to update the desired state of cognition. That is, if people want to recall 15 items in a list, they would monitor ongoing learning and compare that level of learning to the desired state. If people believe that they have learned 15 items, they would use control processes to discontinue study. Given the connection between monitoring and control processes, improving the fidelity of metacognitive monitoring should lead to an accompanying improvement in metacognitive control. In Experiment 3 though, improved metacognitive monitoring was

not accompanied by improved control processes. Possible reasons why there were no improvements will be considered in the general discussion.

3. GENERAL DISCUSSION AND CONCLUSIONS

The experiments in this dissertation were inspired by the classroom finding that low performers have lower subjective confidence in their predictions compared to high performers (Miller & Geraci, 2011c). The goal of the current studies was to test whether asking participants to focus on confidence before making predictions would improve their metacognitive monitoring and subsequent metacognitive control. In Experiment 1, participants studied Swahili-English paired associates and made a prediction about their future memory performance, one group also made a confidence judgment about the accuracy of the prediction. Then both groups make a second performance prediction about their future memory performance. In Experiment 2, participants made four consecutive performance predictions, with the experimental group making confidence judgments after each prediction. The results from Experiments 1 and 2 showed that indeed, metacognitive monitoring accuracy was improved when participants made confidence judgments between performance predictions but not in the control condition when participants only made performance predictions. In Experiment 1, participants' second performance prediction was more accurate than their original performance prediction and in Experiment 2 all adjusted predictions were more accurate than the original prediction. Moreover, in Experiment 2, the final prediction was more accurate than the second prediction indicating that even after 3 adjustments participants could still improve their metacognitive accuracy.

An important finding of the present study is the efficiency with which reporting subjective confidence improved participants' metacognitive monitoring accuracy. Other successful attempts to improve or train metacognitive monitoring in the laboratory are often time-intensive. Some regimens required as many as 23 hours of training distributed over multiple days (Lichtenstein & Fischoff, 1980), weekly or semi-regular training for 16 weeks (Miller & Geraci, 2011b; Nietfeld, Cao, & Osborne, 2006), or hour-long sessions distributed over five days (Adams & Adams, 1958; Kelemen et al., 2007). These training studies also included detailed feedback and descriptions of calibration to the participants. In contrast, the present study required neither explanations of calibration nor feedback and participants showed immediate improvement in their metacognitive accuracy.

Although the data unequivocally suggest that calibration improves following participants' subjective confidence reports, the current studies do not explain why calibration improved. One explanation is that participants were required to think more carefully about the state of their knowledge when they made confidence judgments and this more careful consideration led to more accurate predictions. This hypothesis is consistent with the process-oriented class of modifications that have been used to debias retrospective confidence judgments (Keren, 1990). Process-oriented modifications to improve calibration are preferable over response-oriented modifications. One example of a response-oriented modification is explaining to a participant that people are overconfident and when they make their prediction they should automatically lower the prediction. In this way, response-oriented modifications only changes participants'

metacognitive calibration because they have an abstract awareness of the overconfidence effect but they do not actually consider how the overconfidence effect applies to the state of their own knowledge or future performance. Thus, the effect of response-oriented modifications may be more transient than a process-oriented modification.

It is assumed that process-oriented modifications are longer lasting than response-oriented modifications, but this an empirical question. Future research might explore the lasting benefits of reporting subjective confidence and adjusting performance predictions. Answering this question could be accomplished multiple ways. For example, one could simply ask participants to report subjective confidence before one memory test but not before future memory tests. Alternatively, an experimenter might ask a participant to report subjective confidence between performance predictions for a series of memory tests and examine if first predictions become more calibrated over time. In this way, one could determine if participants learned something about their learning capabilities and whether or not having the experience thinking more deeply about their own learning and knowledge conferred benefits to calibration for future tests.

Related to this last point is the question of generalization to future contexts. In the present paradigm, participants studied foreign language-English word pairs (e.g., lulu-pearl) and then made predictions about their future performance on a memory test. While this sort of rote-memorization paradigm is easy to create in the lab, its generalization to other learning situations may be limited to language learning. The effect of reporting subjective confidence between predictions when more conceptual learning is required has yet to be determined. In addition, this would need to be tested in

the classroom setting where there are lots of other variables at play to determine if the effect generalizes to real-life testing situations.

An alternative hypothesis of the improved calibration effect seen in Experiments 1 and 2 is that the passage of time, from the first to subsequent predictions, was the cause for participants' recalibration, akin to a delayed-JOL effect. However, results from the control conditions, which show no significant calibration improvement, argue against this hypothesis. Because a similar amount of time elapsed between the predictions in the control and experimental conditions but only the experimental condition showed improvement suggesting that reporting subjective confidence between predictions led to the improvement in calibration.

The results also suggest that having participants make confidence ratings disproportionally benefitted calibration for high performing participants – a finding that has been referred as the "Matthew Effect." Interventions that disproportionally benefit higher performing participants are frequent. For example, Kelemen et al. (2007) showed that after 5 study sessions and performance predictions, higher performing participants' predictions became more calibrated whereas there was no change in calibration for lower performing participants. In another study, low performing participants actually became more overconfident in the intervention condition relative to control, but high performers recalibrated in the intervention condition (Ehrlinger et al., 2008). One study has shown, however, that low performing participants' metacognitive accuracy can be improved when they are given incentives and feedback (Miller & Geraci, 2011b). Perhaps because neither feedback nor incentives were provided in the current experiments, lower

performing participants could not recalibrate their predictions. Future research should seek interventions that benefit both groups of participants or even interventions that target low performing participants.

With the finding that providing confidence judgments can improve metacognitive monitoring established in Experiments 1 and 2, Experiment 3 examined whether providing confidence judgments could also affect metacognitive control. The important model of metacognition posits that there is a loop, with metacognitive monitoring processes informing metacognitive control processes (Nelson & Narens, 1990). This model suggests that improvements in monitoring should be accompanied by improvements in control. The results from Experiment 3 indicated that there was no effect of making confidence judgments on metacognitive control. Metacognitive control was operationalized as the number of times participants chose to restudy the to-be-remembered material when they were given a norm-of-study. Participants who had reported confidence about the accuracy of their predictions did not study more than control condition participants, who did not make confidence judgments. Another analysis of the data showed that barely 20% of the participant sample (n = 30) predicted they would remember 15 or more of the items and actually remembered 15 or more of the items. Perhaps, the 15 item norm-of-study was simply too high. Lowering the norm-of-study may reveal a beneficial effect of confidence judgments on metacognitive control in future research.

Both correlational and experimental research designs have shown the connection between improved monitoring and improved control and performance outcomes

(Everson & Tobias, 1998; Nelson et al., 1994; Nietfeld, Cao, & Osborne, 2006; Thiede, 1999; Thiede et al., 2003) so it is unlikely that the model of metacognition is incorrect. Another suggestion as to why metacognitive control was not influenced is that monitoring was not improved. I believe this suggestion can be disputed given that there were consistent monitoring improvements in Experiments 1 and 2. Rather there may implementation reasons why there was no connection between improved monitoring and improved control in Experiment 3. First, the lack of effect of making confidence judgments on study time could have been due to a lack of motivation among participants. For students to choose the option to keep studying the items they must be motivated to do so. From the instructions, each participant read that their goal, or norm-of-study, was to remember at least fifteen items. And if their memory performance was 15 items or more, they would receive a "prize." It is possible participants were not motivated to earn the prize and were content knowing they would not remember fifteen or more items. In fact, barely 25% of the sample had recall performance of 15 or more items and the percentage of participants with a final prediction of 15 or more items was only 59%. Motivating participants more than what was accomplished in the present experiments could reveal the monitoring and control relationship that was hypothesized. A second reason why improved monitoring did may not have lead to improved control in Experiment 3 is that the manipulation takes time for participants to trust that their subjective confidence reports about their predictions are better indicators of learning than their performance predictions. Perhaps more experience and feedback about the benefits of reporting subjective confidence would eventually lead to added benefits for

metacognitive control. Or the manipulation could affect study strategies on something more subtle than allocation of study time.

Based on the results of Experiments 1 and 2, there are at least two future research agendas that are most pressing. The first is to continue to explore the boundary of the improved calibration effect. Experiment 2 included one original prediction and three adjusted predictions. The results indicated that participants' calibration was improved at the third adjusted prediction relative to the first and second prediction. The obvious question then is – would participants continue to recalibrate if asked to make a fourth or fifth adjusted prediction? By definition, there is a limit to the accuracy of metacognitive monitoring, and so determining the number of predictions required to reach complete accuracy would be informative.

In addition, future research must determine if underconfident performance predictions can be recalibrated using the subjective confidence intervention. Or, does making confidence ratings only decrease performance predictions? Certainly, overconfident predictions are more problematic than underconfident judgments in a real-life testing situation because overconfidence would result in discontinuing study or in other settings, overconfidence means continuing to drive longer than one is able or using a gun without gun safety knowledge. Because they are more problematic, overconfident judgments were the focus in these dissertation studies. And the results indicated that overconfident judgments could be recalibrated. One might suggest limiting the analyses in the current dataset to only underconfident predictions to determine if they were

recalibrated. But because the paradigm was designed to elicit overconfidence there were very few underconfident predictions making such analyses imprudent.

One setting where the present findings are applicable is in educational practice. Students regularly make predictions about how much they know (or do not know) of the to-be-remembered material. Conceivably, decisions to keep studying or discontinue studying are tied to these predictions. And so decisions based on predictions have a lasting impact on students' lives by influencing exam scores, course grades, graduation, and eventual career choices. Methods to improve the accuracy of students' predictions, including reporting subjective confidence, could be used in school settings during study. Future research could explore how such practices could be implemented at all levels of education.

A core-finding in metacognitive research is that participants are very often overconfident about their future performance (Bol et al., 2005; Burson et al., 2006; Hacker et al., 2000; Kelemen et al., 2007; Kruger & Dunning, 1999; Krueger & Mueller, 2002; Miller & Geraci, 2011b). In the present experiments, participants were as much as 25% overconfident about their future memory performance. This is a striking dissociation between monitoring and actual performance. Are people hopelessly unaware of their own cognition? Perhaps not – in the classroom, previous research showed that although low performers' over predicted performance they were not confident about these predictions was low, indicating to us that there may be some awareness among low performers that their predictions were too high (Miller & Geraci, 2011c). The research presented here takes advantage of the seemingly privileged access

of confidence judgments to a person's real state of knowledge to improve prediction accuracy. The results suggest that requiring participants to focus on their confidence in their performance predictions offers one promising method for reducing overconfidence.

REFERENCES

Adams, P. A., & Adams, J. K. (1958). Training in confidence-judgments. *The American Journal of Psychology, 71,* 747-751.

Alicke, M. (1985). Global self-evaluation as determined by the desirability and controllability of trait adjectives. *Journal of Personality and Social Psychology, 49,* 1621-1630.

Bol, L., Hacker, D.J., O'Shea, P., & Allen, D. (2005). The influence of overt practice, achievement level, and explanatory style on calibration accuracy and performance. *The Journal of Experimental Education, 73,* 269-290.

Brown, A. (1987). Metacognition, executive control, self-regulation, and other more mysterious mechanisms. In F. E. Weinert & R. H. Kluwe (Eds.), *Metacognition, motivation, and understanding* (pp. 21-29). Hillsdale, NJ: Lawrence Erlbaum Associates.

Buehler, R., Griffin, D., & Ross, M. (1994). Exploring the "planning fallacy": Why people underestimate their task completion times. *Journal of Personality and Social Psychology*, *67*, 366–381.

Burson, K. A., Larrick, R. P., & Klayman, J. (2006). Skilled or unskilled, but still unaware of it: How perceptions of difficulty drive miscalibration in relative comparisons. *Journal of Personality and Social Psychology, 90,* 60-77.

Connor, L. T., Dunlosky, J., & Hertzog, C. (1997). Age-related differences in absolute but not relative accuracy. *Psychology and Aging, 12,* 50-71.

Dunlosky, J., Domoto, P. K., Wang, M. L., Ishikawa, T., Roberson, I., Nelson, T. O., & Ramsay, D. S. (1998). Inhalation of 30% nitrous oxide impairs people's learning without impairing people's judgments of what will be remembered. *Experimental and Clinical Psychopharmacology, 6,* 77-86.

Dunlosky, J., & Hertzog, C. (1998). Training programs to improve learning in later adulthood: Helping older adults educate themselves. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 249-275). Mahwah, NJ: Erlbaum.

Dunlosky, J., Kubat-Silman, A., & Hertzog, C. (2003). Training monitoring skills improves older adults' associative learning. *Psychology and Aging, 18,* 340-345.

Dunlosky, J., & Metcalfe, J. (2009). *Metacognition.* Thousand Oaks, CA: SAGE Publications.

Dunlosky, J., Rawson, K. A., & McDonald, S. L. (2002). Influence of practice test on the accuracy of predicting memory performance for paired associates, sentences, and text material. In T. J. Perfect & B. L. Schwartz (Eds.), Applied Metacognition (pp. 68-92). Cambridge, UK: Cambridge University Press.

Dunlosky, J., Serra, M. J., Matvey, G., & Rawson, K. A. (2005). Second-order judgments about judgments of learning. *The Journal of General Psychology, 132,* 335-346.

Dunning, D. (2005). Self-insight: Roadblocks and detours on the path to knowing thyself. New York: Taylor & Francis Books.

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implication for health, education, and the workplace. *Psychological Science in the Public Interest, 5,* 69-106.

Ehrlinger, J., Johnson, K., Banner, M., Dunning, D., & Kruger, J. (2008). Why the unskilled are unaware: Further explorations of (absent) self-insight among the incompetent. *Organizational Behavior and Human Decision Processes, 105*, 98-121.

Everson, H. T., & Tobias, S. (1998). The ability to estimate knowledge and performance in college: A metacognitive analysis. *Instructional Science, 26,* 65-79.

Flavell, J. H. (1979). Metacognition and cognitive monitoring. *American Psychologist, 34,* 906-911.

Flavell, J. H. (1987). Speculations about the nature and development of metacognition. In F. E. Weinert & R. H. Kluwe (Eds.), *Metacognition, motivation, and understanding* (pp. 21-29). Hillsdale, NJ: Lawrence Erlbaum Associates.

Flavell, J. H., Friedrichs, A. G., & Hoyt, J. D. (1970). Developmental changes in memorization processes. *Cognitive Psychology, 1*, 324-340.

Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J., Rees, G. (2010). Relating introspective accuracy to individual differences in brain structure. *Science, 329,* 1541-1543.

Fuchs, A. H., & Milar, K. J. (2003). Psychology as a science. In D. F. Freedheim (Ed.), *Handbook of psychology, Volume 1: The history of psychology* (pp 1-26). Hoboken, NJ: Wiley.

Gehring, W. J., & Fencsik, D. E. (2001). Functions of the medial frontal cortex in the processing of conflict and errors. *The Journal of Neuroscience, 21,* 9430-9437.

Gramzow, R. H., Willard, G., & Mendes, W. B. (2008). Big tales and cool heads: Academic exaggeration is related to cardiac vagal reactivity. *Emotion, 8,* 138-144.

Grimes, P. W. (2002). The overconfident principles of economics students: An examination of a metacognitive skill. *The Journal of Economic Education, 33,* 15-30.

Hacker, D. J., Bol, L., Horgan, D., & Rakow, E. A. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology, 92,* 160-170.

Hacker, D. J., Bol, L., & Bahbahani, K. (2008a). Explaining calibration accuracy in classroom contexts: The effects of incentives, reflection and explanatory style. *Metacognition and Learning, 3,* 101-121.

Hacker, D. J., Bol, L., & Keener, M. C. (2008b). Metacognition in education: A focus on calibration. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of metamemory and memory* (pp. 429-455). New York: Taylor & Francis Group.

Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology, 56,* 208-216.

Hertzog, C., & Dunlosky, J. (2011). Metacognition in later adulthood: Spared monitoring can benefit older adults' self-regulation. *Current Directions in Psychological, 20,* 167-173.

Hertzog, C., Kidder, D. P., Powell-Moman, A., & Dunlosky, J. (2002). Aging and monitoring associative learning: Is monitoring accuracy spared or impaired? *Psychology and Aging, 17,* 209-225.

Hertzog, C., Sinclair, S. M., & Dunlosky, J. (2010). Age differences in the monitoring of learning: Cross-sectional evidence of spared resolution across the adult life span. *Developmental Psychology, 46,* 939-948.

Huynh, H., & Feldt, L.S. (1976). Estimation of the Box correction for degrees of freedom from sample data in randomized block and split-plot designs. *Journal of the American Statistical Association, 65*, 1582-1589.

Johnson, M. K., Hashtroudi, S., & Lindsay D. S. (1993). Source monitoring. *Psychological Bulletin, 114,* 3-28.

Johnson, M. K., & Raye, C. L. (1981). Reality Monitoring. *Psychological Review, 88,* 67-85.

Kao, Y. C., Davis, E. S., & Gabrieli, J. D. E. (2005). Neural correlates of actual and predicted memory formation. *Nature Neuroscience, 8,* 1776-1783.

Karpicke, J. D. (2009). Metacognitive control and strategy selection: Deciding to practice retrieval during learning. *Journal of Experimental Psychology: General, 138,* 469-486.

Karpicke, J. D., Butler, A. C., & Roediger, H. L. (2009). Metacognitive strategies in student learning: Do students practise retrieval when they study on their own? *Memory, 17,* 471-479.

Karpicke, J. D. & Roediger, H. L. (2008). The critical importance of retrieval on learning. *Science, 319,* 966-968.

Kelemen, W. L., Winningham, R. G., & Weaver, C. A., III. (2007). Repeated testing sessions and scholastic aptitude in college students' metacognitive accuracy. *European Journal of Cognitive Psychology, 19,* 689-717.

Keren, G. (1990). Cognitive aids and debiasing methods: Can cognitive pills cure cognitive ills? In J. Caverni, J. Fabre, & M. Gonzales (Eds.), *Cognitive bias* (pp. 523-552). Oxford: North-Holland.

Knouse, L. E., Bagwell, C. L., Barkley, R. A., & Murphy, K. R. (2005). Accuracy of self-evaluation in adults with ADHD: Evidence from a driving study. *Journal of Attention Disorders, 8,* 221-234.

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review, 100,* 609-639.

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103,* 490-517.

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

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77,* 1121-1134.

Kruger, J., & Dunning, D. (2002). Unskilled and unaware – but why? A reply to Krueger and Mueller (2002). *Journal of Personality and Social Psychology, 82,* 189-192.

Krueger, J., & Mueller, R. A. (2002). Unskilled, unaware, or both? The better-than-average heuristic and statistical regression predict errors in estimates of own performance. *Journal of Personality and Social Psychology, 82,* 180-188.

Leonesio, R. J., & Nelson, T. O. (1990). Do different metamemory judgments tap the same underlying aspects of memory? *Journal of Experimental Psychology: Learning, Memory and Cognition, 16,* 464-470.

Lichtenstein, S., & Fischoff, B. (1980). Training for calibration. *Organizational Behavior and Human Performance, 26,* 149-171.

Lichtenstein, S., Fischoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, and A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306-344). New York: Cambridge University Press.

Luu, P., Flaisch, T., & Tucker, D. M. (2000). Medial frontal cortex in action monitoring. *The Journal of Neuroscience, 20,* 464-469.

Mendez, M. F., & Cummings, J. L. (2003). *Dementia: A clinical approach*. Philadelphia, PA: Butterworth Heinemann.

Metcalfe, J. (1999). Metamemory: Theory and data. In E. Tulving and F. I. M. Craik (Eds.), *Oxford handbook of memory,* (pp. 197-214). Oxford: Oxford University Press.

Metcalfe, J., & Kornell, N. (2005). A region of proximal learning model of study time allocation. *Journal of Memory and Language, 52,* 463-477.

Miller, T. M., & Geraci, L. (2011a). *Probability information influences participants' allocation of study time.* Manuscript submitted for publication.

Miller, T. M. & Geraci, L. (2011b). Training metacognition in the classroom: How incentives and feedback influence exam predictions. *Metacognition and Learning, 6,* 303-314.

Miller, T. M., & Geraci, L. (2011c). Unskilled but aware: Reinterpreting overconfidence in low-performing students. *Journal of Experimental Psychology: Learning, Memory & Cognition, 37,* 502-506.

Moulin, C. (2002). Sense and sensitivity: Metacognition in Alzheimer's Disease. In T. Perfect (Ed.), *Applied Metacognition,* (pp. 197-223). New York: Cambridge University Press.

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95,* 109-133.

Nelson, T. O., & Dunlosky, J (1991). When people's judgments of learning are extremely accurate at predicting subsequent recall: The "delayed-JOL effect". *Psychological Science, 2,* 267-270.

Nelson, T., & Dunlosky, J. (1994). Norms of paired-associated recall during multi-trial learning of Swahili-English translation equivalents. *Memory, 2,* 325-335.

Nelson, T. O., Dunlosky, J., Graf, A., & Narens, L. (1994). Utilization of metacognitive judgments in the allocation of study during multitrial learning. *Psychological Science, 5,* 207-213.

Nelson, T. O., Dunlosky, J., White, D. M., Steinberg, J., Townes, B. D., Anderson, D. (1990). Cognition and metacognition at extreme altitudes on Mount Everest. *Journal of Experimental Psychology: General, 119,* 367-374.

Nelson, T. O., McSpadden, M., Fromme, K., Marlatt, G. A. (1986). Effects of alcohol intoxication on metamemory and on retrieval from long-term memory. *Journal of Experimental Psychology: General, 3,* 247-254.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation*, (Vol. 26, pp. 125-173). New York: Academic Press.

Nietfeld, J. L., Cao, L., & Osborne, J. W. (2005). Metacognitive monitoring accuracy and student performance in the postsecondary classroom. *The Journal of Experimental Education, 74,* 7-28.

Nietfeld, J. L., Cao, L., & Osborne, J. W. (2006). The effect of distributed monitoring exercises and feedback on performance, monitoring accuracy, and self-efficacy. *Metacognition and Learning, 1,* 159-179.

Pannu, J. K., & Kaszniak, A. W. (2005). Metamemory experiments in neurological populations: A review. *Neuropsychology Review, 15,* 105-130.

Pierce, B. H. & Smith, S. M. (2001). The postdiction superiority effect in metacomprehension of text. *Memory & Cognition, 29,* 62-67.

Rhodes, M. G., & Tauber, S. K. (2011). The influence of delaying judgments of learning on metacognitive accuracy: A meta-analytic review. *Psychological Bulletin, 137,* 131-148.

Rounis, E., Maniscalco, B., Rothwell, J. C., Passingham, R. E., & Lau, H. (2010). Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness. *Cognitive Neuroscience, 1,* 165-175.

Salat, D. H., Kaye, J. H., & Janowsky, J. S. (2001). Selective preservation and degeneration in the prefrontal cortex in aging and Alzheimer disease. *Archives of Neurology, 58,* 1403-1408.

Salthouse, T. A. (2004). What and when of cognitive aging. *Current Directions in Psychological Science, 13,* 140-144.

Schraw, G. (2000). Assessing metacognition: Implications of the Buros Symposium. In G. Schraw and J. C. Impara (Eds.), *Issues in the measurement of metacognition.* Lincoln, NE: Buros Institute of Mental Measurements.

Schwartz, B. L. (1994). Sources of information in metamemory: Judgments of learning and feelings of knowing. *Psychonomic Bulletin & Review, 1,* 357-375.

Schwartz, B. L., & Bacon, E. (2008). Metacognitive neuroscience. In J. Dunlosky and R. A. Bjork (Eds.), *Handbook of metamemory and memory*. New York: Psychology Press.

Schwartz, B. L., & Metcalfe, J. (1992). Cue familiarity but not target retrievability enhances feeling-of-knowing judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 1074-1083).

Simons, R. F. (2010). The way of our errors: Theme and variations. *Psychophysiology, 47,* 1-14.

Singh-Manoux, A., Kivimaki, M., Glymour, M. M., Elbaz, A., Berr, C., Ebmeier, K. P., … Dugravot, A. (2011). Timing of onset of cognitive decline: Results from Whitehall II prospective cohort study. *British Medical Journal, 344,* doi:10.1136/bmj.d7622

Thiede, K. W. (1999). The importance of monitoring and self-regulation during multitrial learning. *Psychonomic Bulletin & Review, 6,* 662-667.

Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology, 95,* 66-73.

Underwood, B. J. (1966). Individual and group prediction of item difficult for free learning. *Journal of Experimental Psychology, 71,* 673-679.

Wilkinson, T. S., & Nelson, T. O. (1984). FACTRETRIEVAL2: A Pascal program for assessing someone's recall of general-information facts, confidence about recall correctness, feeling-of-knowing judgments for nonrecalled facts, and recognition of nonrecalled facts. *Behavior Research Methods, Instruments, and Computers, 16,* 486-488.

Wundt, W. M. (1873). Selected texts from writings of Wilhelm (S. Diamond, trans.). In R. W. Rieber (Ed.), *Wilhelm Wundt and the making of a scientific psychology* (pp. 155-177). New York: Plenum.

Zachary, R.A. (1986). *Shipley Institute of Living Scale, Revised Manual.* Los Angeles, CA: Western Psychological Services.

APPENDIX

Swahili	English
ankra	Invoice
bahasha	Envelope
chaza	Oyster
chimbo	Quarry
desturi	Custom
duara	Wheel
fahali	Bull
gharika	Flood
jibini	Cheese
kamba	Rope
kasuku	Parrot
ladha	Flavor
lawama	Blame
nafaka	Corn
nanga	Anchor
nira	Yoke
sahani	Plate
talaka	Divorce
ubini	Forgery
yamini	Oath

VITA

Tyler Michael Miller

Contact Information

Mailing Address:	Department of Psychology Texas A&M University 4235 TAMU College Station, TX 77843
E-mail:	milltyl@tamu.edu

Education

Ph.D. Experimental Psychology, Cognitive 2012
Texas A&M University, College Station, TX
Using Subjective Confidence to Improve Metacognitive Monitoring and Control
Advisor: Dr. Lisa Geraci

M.S. Experimental Psychology, Behavioral Neuroscience 2008
Emporia State University, Emporia, KS
Advisor: Dr. Cathy Grover

B.A. Psychology 2004
Minor: Art
Buena Vista University, Storm Lake, IA