Spring 2020

Does being unaware predict being unskilled? Analyzing the predictive ability of the Metacognition Awareness Inventory

Daniel A. Scheibe

DOES BEING UNAWARE PREDICT BEING UNSKILLED?

ANALYZING THE PREDICTIVE ABILITY OF THE METACOGNITION

AWARENESS INVENTORY

A Thesis

Presented To

Eastern Washington University

Cheney, Washington

In Partial Fulfillment of the Requirements

For the Degree

Master of Science in Experimental Psychology

By

Daniel A. Scheibe

Spring 2020

THESIS OF DANIEL A. SCHEIBE APPROVED BY

Electronically approved 6/10/2020

____ DATE ________

DR. DANIELLE SITZMAN, GRADUATE STUDY COMMITTEE

Electronically approved 6/10/2020

____ DATE ________

DR. PHILIP WATKINS, GRADUATE STUDY COMMITTEE

Electronically approved 6/10/2020

____ DATE ________

DR. JACQUELINE COOMES, GRADUATE STUDY COMMITTEE

ABSTRACT

DOES BEING UNAWARE PREDICT BEING UNSKILLED?

ANALYZING THE PREDICTIVE ABILITY OF THE METACOGNITION AWARENESS INVENTORY

by

Daniel A. Scheibe

Spring 2020

Students' perceptions of how well they know information ultimately impact the study choices they make. The more accurate these perceptions are, the more prepared students are to be academically successful. Thus, the current study aimed to find an efficient way to quickly identify students who struggle with this self-assessment, and ultimately classroom performance, using a tool designed to assess metacognitive ability. Participants first completed a metacognition scale designed to assess learning strategies – the Metacognitive Awareness Inventory (MAI). Next, the participants were administered a 29-minute lecture followed by a brief exam at the end of the lecture. There were three types of metacognitive judgements used in this study: a) A global predictive judgment, b) item-by-item confidence judgments, and c) a global postdictive judgment. The exam results were compared to participants' self-report on the MAI and participants' judgements. It was hypothesized that participants with higher MAI scores would do consistently better on the post-lecture exam than participants with lower MAI scores and be more accurate with their metacognitive judgments. Additionally, it was hypothesized that participants in the lowest metacognitive quadrant on the MAI would be overconfident in their predictions of exam performance. Neither hypothesis was supported: The MAI does not appear to be a reliable measure for predicting accuracy of metacognitive judgments or immediate academic performance.

ACKNOWLEDGMENTS

The author would like to thank several contributors for their roles in this project. Thanks goes to Dr. Danielle Sitzman for her many rigorous edits and comments on this manuscript, as well as being first chair on this project. Thanks go to Dr. Philip Watkins and Dr. Jacqueline Coomes for their role as committee members on this project – serving as second and third chair, respectively.

Additionally, thanks go to research assistants Shelby Madsen and Shaylynn Sapp who made themselves available to learn the procedure and run participants. This project would not have been possible if it were not for this team of competent, motivated individuals providing feedback and support.

TABLE OF CONTENTS

Does Being Unaware Predict Being Unskilled?

For over a century now in the U.S. it has been important to pursue academic success, but now more than ever it is imperative that American higher education finds a way to turn college freshmen into college graduates. One proposed method for preventing the spiral that leads to dropping out is to identify students that will struggle as early as possible and help said students overcome challenges with correctly calibrating their abilities. While some students may organize their skills and knowledge into structure, there appears to be little encouragement from instructors to students to develop this structure (Schraw & Moshman, 1995). For many instructors, helping students at the individual level with topics like study strategies and organizing students' skills and knowledge is not practical due to time constraint and large-scale issues. For that reason, self-report measures of student awareness and learning are both valuable and practical for large-scale use in academia (Schellings & Van Hout-Wolters, 2011). The purpose of the present study is to determine if the Metacognitive Awareness Inventory (MAI), a common metacognitive assessment used in education fields, will be predictive of student lecture retention ability and metacognitive awareness. It is of particular interest whether the MAI will be more predictive than other easily accessible student variables such as grade point average.

Metacognition

Metacognition as a construct has been integrated in the field of cognitive psychology for the past 40 years (Flavell, 1979). Perhaps most broadly put, metacognition is an umbrella term referring to cognition about other cognitions, or thinking about thinking (Dunlosky & Metcalfe, 2009). Historically, researchers distinguish between metacognitive monitoring and metacognitive control processes. Metacognitive monitoring involves what one knows about one's own cognition, and metacognitive control processes involve how one uses knowledge of one's own cognition to regulate cognitive processes and as a by-product, behaviors.

The concepts of metacognitive monitoring and metacognitive control processes appear daily in higher education. In fact, enhancing metacognition is part and parcel of becoming a better student (Rhodes et al., 2020). Students use cognitive knowledge when choosing how they will study for an exam, and metacognitive knowledge when they are deciding if they have studied enough for an exam (Serra & DeMarree, 2016). For example, if a student glances briefly through her notes and decides that she knows the material, she might decide that she has studied enough and move on to other endeavors. That in-the-moment metacognitive judgment of the student's own learning can be strongly influential on that student's test outcome. The more accurately a student can judge their knowledge, the better use they will make of their time and the more prepared they will be for exams. We are constantly making judgments about what we know and do not know and making decisions based on these judgments (Brown et al., 2014).

In order to adequately plan for academic success, students must select appropriate strategies and allocate appropriate resources that affect performance (Schraw & Graham, 1997). Metacognition has further been related to educational successes and failures through a relation to the academic locus of control, or the belief that one is able to perform a task (Arslan & Akin, 2014). Arslan and colleagues found that metacognition is positively associated with an internal academic locus of control. In other words, the more metacognitively in-tune students are, the more they tend to believe they control their own academic success. Thus, identifying students who struggle to assess their own cognition - and helping to improve that assessment - should lead to academic benefits.

One often over-looked influence on students' self-assessment is the students' instructor (Siedlecka et al., 2018). If an instructor can accurately help students understand any disparities between what the students know and what they perceive they know, then the only elements missing for success are motivation and the correct tools to close the gap. When instructors understand what students know – and what they think they know – towards the beginning of courses, then the instructors are better prepared to design instruction in an appropriate way that facilitates learning (Ambrose et al., 2010). The idea of metacognition, or "knowing what you know", is a powerful tool when considering education at all levels. Metacognition directly relates to how individuals monitor and regulate their cognitive abilities and performances (Schraw & Graham, 1997), and as such it is a crucial component to education. Because metacognition is a fundamental element of student success, it is important to consider reliable ways to measure it.

Assessing Metacognition

Metacognition can be measured in a predictive manner or a postdictive manner. Predictive judgments require people to estimate the likelihood of remembering information in a future point in time; for example, presenting a student with a topic they just learned and asking the student to make a judgment of how likely they are to remember that information on a later test. Postdictive judgments are done after taking an examination or assessment. These judgments come in various forms and involve the participant providing a confidence rating in the accuracy of an individual test question or retrospective judgment of an overall test score (Callender et al., 2014; Nelson et al., 2004). The most accurate judgments are the retrospective judgments because there is less variance in these judgments. Once participants have taken an exam, they have a more thorough understanding of the variables at play in their judgments, and the judgments become more accurate. These judgments appear to be more indicative of a general "monitoring" skill, or metacognitive ability (Schraw et al., 1995). There is a debate in the literature whether a general metacognitive skill exists of if it is domain-specific skill, but recent studies support the idea that a general metacognitive skill exists and is one component of development of higher level cognitive functions (Bellon et al., 2020; Geurten et al., 2018).

Going a step farther, predictive and postdictive judgements are high-level cognitive functions that can assess average performance (Global judgments) or performance on a single item (Relative judgments; Rhodes, 2019). Global (absolute) judgments assess how well someone can predict their overall performance. For example, a participant may be asked, "Out of 100 percent possible, what percent correct do you believe you scored on this test?" Relative (item-by-item) judgments, in contrast, examine whether the individuals are able to discriminate between the accuracy of specific items (Rhodes, 2019). For example, a relative judgment might ask, "On a scale of 0% to 100%, how confident are you that you answered this question correctly?" If participants can accurately assess their performance, they should provide higher judgments to items they answer correctly and lower judgments to items they answer incorrectly. Of most interest in the current study are predictive and postdictive global judgments of exam performance and item-by-item confidence judgments of specific items on the exam. It is important to measure both global and relative judgments because they might be measuring different metacognitive abilities. Research suggests that people differ greatly not only in metacognitive ability, but also conceptualization of metacognition (Schraw & Moshman, 1995). Global judgments force participants to provide broad, often challenging judgments that often lead to lower reliability (Krueger & Mueller, 2002), while item-by-item confidence judgments are much more specific and isolated. Item-by-item confidence judgments also deal with inherently fewer variables, so they tend to elicit smaller ranges and better accuracy (Hartwig & Dunlosky, 2013).

It is not surprising that confidence judgments have the interest of cognitive psychologists, because our cognitions often arise naturally and lead us to make either well-founded or rash decisions (Dunlosky & Metcalfe, 2009). While research has been fairly widespread on metacognitive judgments since the 1970s, we still do not fully understand the underlying processes in relationship to education. However, there are natural disparities between higher performing and lower performing students in academics and research in the past 20 years indicates that there is may be a robust link between metacognitive awareness and academic success (i.e., the Dunning-Kruger effect; Kruger & Dunning, 1999; Serra & DeMarree, 2016).

Dunning-Kruger Effect

The Dunning-Kruger Effect was born from one of the first studies that aimed to analyze individual differences in metacognitive ability (Kruger & Dunning, 1999). High performing individuals are often able to accurately distinguish between what they do and do not know, while low performing individuals struggle to make this distinction. Kruger and Dunning (1999) suggested a dual burden for the poorest performing participants.

Incompetent individuals tend to be unaware of their incompetence, thus robbing themselves of the theoretical opportunity to do something to better their situation.

In a flagship study that developed the Dunning-Kruger Effect, Kruger and Dunning (1999) conducted four independent studies that tested participants' metacognition in different ways. Across four separate studies, Kruger and Dunning (1999) studied metacognitive abilities and performance in multiple domains: Humor, logical reasoning, grammar, and finally, manipulating competence to see if that improved metacognitive skills. This series of studies revealed that there is consistently a robust finding that there is an unskilled and unaware group that is low-performing and exhibits low metacognitive skills, and paradoxically, the way to make incompetent individuals realize their own incompetence is to make them competent.

The Dunning-Kruger effect is also robust in a classroom setting. In a study directly using a classroom context, Hacker and colleagues (2000) addressed the questions of how accurately students can both predictively and postdictively assess their abilities. They also asked questions related to accuracy varying over multiple tests. They examined 99 undergraduate college students enrolled in two sections of an educational psychology course. A major emphasis of this course was on the very nature of self-assessment. Just prior to exams, students completed an assessment of what percentage of items they expected to get correct, and immediately following exams students estimated how they performed on the exam (as measured by total items correct). They were asked to reflect and analyze why their predictions and postdictions were or were not accurate and develop an action plan to prepare for the next exam. The lowest performing students showed strong overconfidence, which become increasingly exacerbated the lower their true score was. In other words, students who performed the worst may have been aware that they did not do extremely well on the exam, but they were still unable to accurately pinpoint how poor their performance was. While the high-performing students increased their predictive and postdictive accuracy throughout the course, the low-performing students continued to show little predictive accuracy: They remained unaware of their level of incompetence. This research suggests that the lowest performing students demonstrated the worst metacognitive predictions and postdictions of their knowledge: An applied finding supporting the Dunning-Kruger effect in the classroom (Hacker et al., 2000). This work has been replicated and expanded upon by other studies (Mattern et al., 2010; Serra & DeMarree, 2016), but to date the literature has not found a successful way to target this unskilled and unaware group in a quick, reliable way.

The Dunning-Kruger effect demonstrates a clear inability for the lowest performing students to correctly assess their inadequacies. The lowest performing students (the unskilled-unaware population) overestimate their performances, while the highest performing students tend to slightly underestimate their abilities (Mattern et al., 2010). Although an underestimation could act as a small hinderance to highly skilled students as they pursue their academic goals, the much more dramatic consequences are seen in the unskilled-unaware population. The dramatic overestimation of preparedness for this group serves as a temporary emotional benefit with unfortunate long-term consequences. In fact, research suggests that the unskilled are motivated to ignore their poor performance so they can maintain a more blissful emotional state and feel better about themselves (Kim et al., 2015). This temporary “bliss” is of course not helpful to these unskilled-unaware students in the long run, and the discrepancy in metacognitive ability between high-performing students and low-performing students leads to high-performing students exhibiting more effective test preparation practices, better test performances, and superior monitoring processes than their counterparts (Filho, 2009). Thus, being able to effectively identify these students early in a course, before major exams, may be an important factor in getting students the help and resources they need to be successful. Because research suggests that the unskilled and unaware are resistant to becoming accurate with their predictions (Hacker et al., 2000), it would be highly valuable for instructors to have a quick, easy way to identify these students early in a course.

Metacognitive Awareness Inventory

Often researchers look at the term “metacognition” with specific assessments of information (e.g., tests for a course), but some researchers have attempted to create more general inventories to assess a broader perspective of metacognition (Brown, 1987; Schraw & Moshman, 1995). The Metacognitive Awareness Inventory (MAI) is a metacognitive self-assessment tool specifically related to student learning and study strategies (Brown, 1987). The MAI was developed in educational psychology as a tool to assess students’ techniques, increase awareness of poor strategies, and facilitate appropriate study strategies. Because the MAI is the only inventory to date that deals exclusively with student study strategies (Harrison & Vallin, 2017), it is of particular interest whether the simple inventory can yield meaningful information about students’ metacognitive abilities as a predictor of performance. It is presently unclear whether metacognitive scales developed in education map onto metacognitive accuracy as defined in cognitive psychology, but if they do, this may be a good way for teachers to detect students who will struggle at a preliminary level. The present study uses a widely accepted metacognitive inventory designed specifically to monitor learning strategies.

The MAI was the first reliable measure for assessing metacognition through self-report methodology (Schraw & Dennison, 1994). This scale taps the traditional theories of metacognition that consider knowledge of cognition (or metacognitive monitoring) and regulation of cognition based on this knowledge (or metacognitive control). The MAI consists of 52 true or false questions. Each question is an "I" statement for the participant, and the participants must decide if that statement applies to them (true) or does not apply to them (false). There are no reverse-scored items on the MAI, meaning that a score of 52 would be a "perfect score" and the student would be self-reporting perfect study strategies and metacognition. If students are accurately self-reporting their study strategies and metacognition, then a higher score on the MAI equates to being more metacognitively aware. MAI scores get coded into two main categories: Knowledge about cognition (metacognitive monitoring), and regulation of cognition (metacognitive control). An example question about knowledge of cognition is, "I know what the teacher expects me to learn." An example question addressing regulation of cognition is, "I find myself pausing regularly to check my comprehension."

Knowledge of cognition further divides up into the three commonly considered subsets of metacognitive knowledge: Declarative knowledge, procedural knowledge, and conditional knowledge (Schraw & Graham, 1997; Schraw & Moshman, 1995). Declarative knowledge is the factual knowledge the learner needs before being able to use critical thinking. Procedural knowledge is the application of the declarative knowledge. Conditional knowledge is the application of declarative knowledge and procedural knowledge under different circumstances where specific processes or skills should transfer (Schraw & Dennison, 1994; Schraw & Moshman, 1995). Regulation of cognition divides into five categories: Planning, information management strategies, comprehension monitoring, debugging strategies, and evaluation. Planning is goal setting and allocating resources prior to learning. Information management strategies are skills and strategy sequences used to process information more efficiently. Comprehension monitoring is assessment of one's learning or strategy use. Debugging strategies are used to corrected comprehension issues and performance errors. Evaluation is the analysis of performance and strategy effectiveness after a learning episode (Schraw & Dennison, 1994; Schraw & Moshman, 1995).

It is yet uncertain whether this type of self-report scale and a general measure of academic metacognitive ability can predict performance in the classroom. Theoretically, if students score higher on the MAI (indicating higher levels of self-awareness), then they should be able to perform better on academic assessments and more accurately predict how well they will perform. The MAI is a simple tool and easy to administer: If it has predictive qualities related to student performance, then there are several implications for education. First, instructors have an easy tool to decipher which students may initially need more help in their classroom and can provide these students with the materials and resources to be successful. Second, the MAI would be an easy way for students to evaluate their own study strategies and provide students with a reality check to their perceived abilities compared to their actual abilities. Third, from only one class session, or even a pre-class survey, instructors already have a way to judge their students' ability. Understanding what students know coming into courses allows instructors to more efficiently design their coursework and be effective catalysts facilitating student learning (Ambrose et al., 2010).

The Present Study

Several studies show that there is a strong correlation between metacognitive ability and success in learning outcome studies (Callender et al., 2016; Filho, 2009; Rhodes et al, 2020; Schraw et al., 1995; Serra & DeMarree, 2016). Additionally, other studies have shown that metacognitive inventories (such as the MAI) are a reliable way to predict metacognitive abilities (Pedone et al., 2017; Schraw & Dennison, 1997; Semerari et al., 2012). The question of the current study is: Can the MAI reliably predict which individuals will be higher performing and lower performing in terms of both metacognitive ability and testing lecture retention? If the answer to this question is yes, then the applications for education are substantial. Instructors could quickly and easily identify students that are likely to struggle the most in the course in a matter of 20 minutes (Reid et al., 2017).

The current study examined how the MAI related to both objective and subjective (metacognitive) performance on an exam. After completing the MAI, participants watched a video lecture and took a test over the material in the lecture. Prior to the exam, students made broad predictions about how well they thought they would perform. On the exam, participants answered multiple choice questions and rated their confidence in the accuracy of their response. Finally, at the end of the exam, students were asked to provide a general postdictive assessment of their absolute score on the exam.

There are two hypotheses for this study. First, it is expected that individuals who score higher on the MAI will do better on the lecture test than individuals that score lower on the MAI. Second, it is predicted that individuals in the lowest quadrant of the MAI will be less metacognitively accurate as exhibited by being less accurate with (a) their overall predictive judgments, (b) their individual item-by-item confidence judgments and (c) their overall postdictive absolute judgment on the exam.

Method

Participants

A total of 49 students from Eastern Washington University, a mid-size public university in the Pacific Northwest, participated in the study. The age range of participants in this study was 18-27 (M = 20.63, SD = 3.76). Participants were primarily female (n = 41). All students participated for partial course credit in undergraduate psychology classes at EWU. This study adhered to all the ethical principles and was approved by the Institutional Review Board. No specifically identifying information was collected, and student responses were coded such that they were kept confidential and anonymous.

Materials

Metacognitive Inventory

The primary purpose of this study is to draw conclusions about the Metacognitive Awareness Inventory related to education. The MAI consists of 52 true/false “I” statements like, “I have a specific purpose for each strategy I use,” and, “I think of several ways to solve a problem and choose the best one” that load onto eight sub-categories (See Appendix 1). Each of the eight sub-categories has between 4 and 10 items for coding after the participant completes the 52-item inventory. This system of evaluating metacognitive abilities goes a step farther than just labeling participants with an umbrella "metacognitive score." Two data-check items were added into the MAI to make sure that participants were paying attention to the MAI questions.

Video Lecture

The lecture used in this study was a 29-minute video lecture on the topic of marine biology. Specifically, the lecture discussed the topic of Arctic and Antarctic environments, which was a singular lecture in a larger lecture series through *The Great Courses Plus.* (Todd, 2010). The lecture was used with consent from *The Great Courses Plus* for academic purposes. Prior to taking the lecture, participants were given a survey to assess a) their major, b) the extent – if any - of classes they have taken in biology, and c) on a 9-point Likert scale how much knowledge they have of specifically marine biology. This topic was chosen specifically because it is unlikely that the participants in this study have been exposed to much marine biology.

Post-Lecture Assessment

Participants were asked to evaluate their performance in three different ways. They were asked (a) a one-question global predictive assessment of how well they think they will do on the exam, (b) individual item-by-item confidence judgments for each question on the test, and (c) a one-question global postdictive assessment of how well they did on the exam.

The global predictive assessment stated, "You are about to be tested on the material you learned. The test will consist of only multiple-choice questions. On a scale of 0-100, with 0 meaning 0% of questions answered correctly and 100 being 100% of questions answered correctly, how well do you think you will do on that test?" The assessment consisted of 25 multiple-choice questions about material from the video lecture (See Appendix 2). Questions were across many different sub-topics and covered material addressed throughout the 29-minute lecture. Example questions are, “Which word is assigned to krill by the lecturer to denote krill’s significance to their ecosystem,” and “Approximately what percentage of earth’s ice is in Antarctica?” Each question had four possible answers. Each question had two components: a) The question itself and b) a confidence judgment which participants completed after they provided an answer to the question. The confidence prompt was, “On a scale of 25% (random chance) and 100% (absolutely certain), how confident are you in the accuracy of your answer to this question?” At the end of the exam, the global postdictive assessment stated, “Now that you have completed the multiple-choice exam, please re-visit the question you were asked before the exam. On a scale of 0-100, with 0 meaning 0% of questions answered correctly and 100 meaning 100% of questions answered correctly, how well do you think you did on the test?” Finally, there was a brief demographic survey at the end of the packet collecting information, class standing, high school grade point average, and college grade point average.

Procedure

This study was conducted in a small classroom with groups of 1-10 participants. Participants read through consent forms, and then completed the first packet which included the MAI and the marine biology knowledge 3-question survey. Once all the participants in the group completed the first packet, the research assistant provided participants with materials for which to take notes and informed the participants that although taking notes is optional, it is encouraged. Then the research assistant played the 29-minute video lecture. At the end of the video lecture, the participants’ notes were collected, and the second packet was handed out. The second packet included the one-question global predictive question, the 25-item test (paired with item-by-item confidence judgments), the global postdictive question, and the demographic questionnaire. Participants were instructed to complete each section of the packet before proceeding to the next section. Upon completion of the second packet, participants were thanked for their time and free to leave. The study lasted for an average of approximately 55-60 minutes.

Results

The present student looked at the MAI as a predictor of exam scores and accuracy of metacognitive judgments. The first section analyzes basic performance on exam scores, metacognitive judgments, and MAI scores. Following this, the two main hypotheses for the study are analyzed and discussed .

Basic Performance

Exam Scores

The average exam score was 59.67% (*SE* = 2.01%) and scores ranged from 36% to 92%. The exam was 25 multiple choice questions: Each question had four possible answers. Therefore, random chance would be a score of 25.00% on the exam. A one-sample t-test was run which showed that participants performed significantly better than chance on the exam, $t(48) = 17.25$, $p < .001$. There were no questions that 100% of participants answered correctly (91.8% of participants correctly answered the easiest question) and there were no questions that 0% of participants answered correctly (16.3% of participants correctly answered the hardest question).

Global Judgments

Prior to taking the exam, participants were prompted to predict their performance on the exam on a scale of 0% to 100%. The average predicted score was 56.74% (*SE* = 2.61%) with a wide range: Participants predicted they would score from 0% to 90% on the exam. Two participants incorrectly provided a range prediction instead of a specific score. Thus, the average of that range was used for their prediction (i.e. a prediction of 40%-60% would be scored as a prediction of 50%). Three participants did not provide predictive confidence judgments. After the exam, participants responded to the same question, but now as a postdictive judgment with the added benefit of fully knowing what they were being asked to evaluate. The average postdictive judgment was 59.18 (*SE* = 3.28%), with nearly the same, wide range as the predictive judgments: 10% to 90%. A paired sample t-test was run which revealed that there were not significant differences between predictive and postdictive judgments at the participant level, $t(44) = -1.14$, $p = .259$.

Metacognitive Accuracy

The accuracy of metacognitive judgements was explored by examining both the precision of global predictive/postdictive judgments and item-by-item confidence judgments. In accordance with previous literature, Goodman-Kruskal gamma correlations (γ) were used to correlate judgments with performance. This non-parametric correlation assumes judgments are ordinal as opposed to having scale properties (Rhodes, 2019). Similar to regular Pearson "*r*" correlation values, gammas range between -1.00 and 1.00. A higher gamma correlation suggests a stronger correlation between the two variables (i.e. a γ of .75 suggests a much stronger relationship between two variables than a γ of .20). For the sake of the present study, higher gamma values – gammas closer to 1.00 –

represent more accurate judgments (i.e., higher judgments for items answered correctly and lower judgments to items answered incorrectly, or more accurate global judgments).

Relative Accuracy

Calculating a gamma correlation – between item-by-item confidence judgments and accuracy of a response - can control for participants using judgment scales in different ways. For example, one participant might provide a 75% confidence judgment and another participant might provide a 95% confidence judgment, when both participants are trying to relay the same level of confidence. Gamma correlations control for the differences in the subjectivity of these judgments.

A gamma correlation between item-by-item confidence judgements and accuracy of a response (scored as 0 for incorrect and 1 for correct) was calculated for each participant, and then averaged across all participants. Overall, there was a significant correlation (γ =.49, SE = .04), $t(48) = 17.25$, $p < .001$, indicating that, on average, participants fairly accurately discriminated between correct and incorrect answers on the exam. However, there was also a wide range in correlations (-.36 to .94) suggesting that some participants were better able to predict their performance than others (to be further discussed in the Dunning-Kruger effect section).

Absolute Accuracy

In addition to item-by-item judgments, participants also made global predictive and postdictive judgments. The accuracy of these judgments was measured by comparing the differences between predictive judgments and exam scores. For example, if a participant predicted they would score 50% , but they actually scored 80%, then the difference would be 30%. A positive difference score (i.e. +30%) represents under-confidence; a negative difference score (i.e. -30%) represents over-confidence. As seen in Figure 1, participants' predictive judgments before the exam correlated strongly with their postdictive judgments after the exam, γ =.69, $p < .001$. This suggests a trend of participants having a fairly stable evaluation of their test performance, both before and after the exam. A gamma correlation was run between participants' predicted exam scores and true exam scores, which did not reveal a significant correlation, $\gamma = .21$, $p = .140$. However, there was a significant positive correlation between participants' postdicted exam scores and true exam scores , $\gamma = .39$, $p < .001$. Thus, after completing the exam, participants were better able to predict their performance than they were prior to the exam.

The Dunning-Kruger Effect

In order to replicate the Dunning-Kruger effect in a way consistent with the literature, a quartile split was done by exam scores (Hacker, 2000; Kruger & Dunning, 1999; Serra & DeMarree, 2016). The highest performing participants on the exam (the top quartile) should have more accurate global predictions compared to the lowest performing participants on the exam (the bottom quartile). Additionally, to replicate the Dunning-Kruger effect, the top quartile of participants should underestimate their exam performance (positive difference scores), while the bottom quartile of participants should overestimate their exam performance (negative difference scores).

The difference scores between predictive judgments and actual exam scores were tested against zero for both participants in the top quartile and participants in the bottom quartile. Overall, participants in the bottom quartile significantly overestimated their performance (M = -16.60, SE = 4.40), $t(9) = -3.77$, $p = .004$, while participants in the top quartile significantly underestimated their exam performance (*M* =15.53, *SE* = 5.38), *t*(12)= 2.89, *p* = .014. These scores were significantly different from one another, *t*(21)= -4.43, *p* < .001.

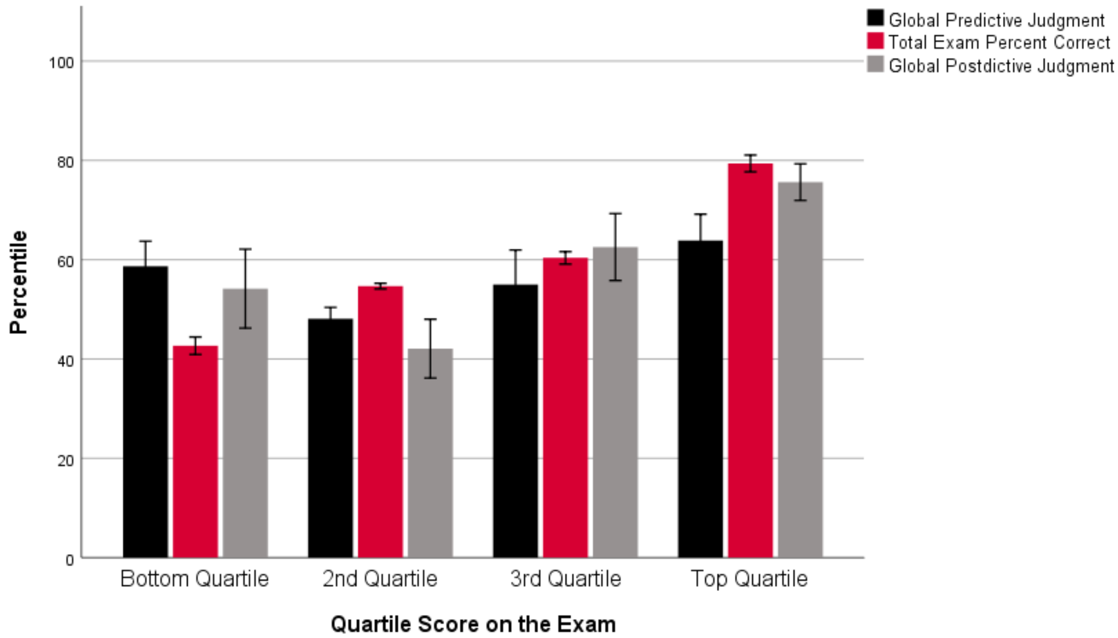
The same set of analyses were run on difference scores between postdictive judgments and exam scores. Although participants in the bottom quartile continued to significantly overestimate their exam performance (*M* = -11.71, *SE* = 5.26), *t*(11)= -2.23, *p* = .048, participants in the top quartile accurately predicted their performance and were not significantly underconfident (*M* = 3.77, *SE* = 3.81), *t*(12)= .99, *p* = .343. These scores were significantly different from one another, *t*(23)= -2.41, *p* = .024.

As seen in Figure 1, the data from the present study replicated the Dunning-Krueger effect in that the bottom quartile (on exam scores) exhibited overconfidence of their scores for both predictions and postdictions while the top quartile (on exam scores) exhibited initial under-confidence, but after taking the exam these participants were able to accurately judge their performance.

In addition to global predictive judgment and global postdictive judgments, the item-by-item confidence by accuracy relationships were examined to see if there were differences between the top quartile and bottom quartile (of exam scores) An independent samples was approaching significance to suggest that perhaps the gamma correlations – between item accuracy and item confidence judgments – were more accurate for the top quartile (*M* = .57, *SE* = .06) compared to the bottom quartile (*M* = .34, *SE* = .09), *t*(24)= -2.03, *p* = .054. Participants who performed best on the exam may be better able to accurately distinguish between correct and incorrect responses on the final exam than low performing participants.

Figure 1

Quartile-split exam scores compared to predictive and postdictive judgments



Note. Error bars show standard errors.

MAI Results

The primary goal of the present study centers on questions related to the MAI and its potential predictive abilities. A higher numerical score on the MAI equates to a higher self-reported metacognitive awareness, with the highest possible score of 52. Participants self-reported an average score of 36.76 (scores ranged from 24-49, *SE* = .95). This study examined whether the MAI is predictive of a) exam performance and b) metacognitive ability. Quartile splits were done based on MAI scores to attempt to reveal any predictive qualities the MAI might have when comparing the highest and lowest MAI self-reported scores (See Table 1).

Quartile splits have the potential to bring to light effects that can be lost in standard correlation and regression models. The bottom quartile of the MAI (*n* = 12) averaged a score of 28.50 (*SE* = .68), the second quartile of the MAI (*n* =12) averaged a score of 33.92 (*SE* = .43), the third quartile of the MAI (*n* = 13) averaged a score of 38.62 (*SE* = .33), and the top quartile of the MAI (*n* =12) averaged a score of 45.83 (*SE* = .64). For a full breakdown of how the different self-reported quartiles performed on exam scores, global predictions, item-by-item accuracy of confidence judgments, and global postdictions, see Table 1.

Table 1

Comparing MAI quartiles for metacognitive predictions, accuracy, and exam scores.

	Sample Size	Exam Score	Pre-Exam Prediction	Item Accuracy by Confidence Gamma	Post-Exam Prediction
Bottom Quartile	12	58.67% (4.16%)	53.75% (6.63%)	.63 (.06)	53.58% (6.77%)
2nd Quartile	12	66.00% (3.50%)	56.25% (4.85%)	.49 (.08)	55.67% (6.93%)
3rd Quartile	13	56.62% (3.32%)	57.55% (4.22%)	.38 (.08)	63.92% (6.22%)
Top Quartile	12	57.67% (4.93%)	59.73% (5.19%)	.49 (.09)	63.54% (6.56%)

Note. Standard errors of the mean are provided in parentheses.

Hypothesis 1: Exam Scores and the MAI

It was hypothesized that participants who self-reported higher scores on the MAI would also score higher on the exam. To support the first hypothesis, the results would have to show a predictive ability of the MAI regarding exam scores. If this hypothesis is accurate, there should be a clear relationship between MAI scores and exam scores, such that in general higher MAI scores would predict higher exam scores, and lower MAI scores would predict lower exam scores. Quartile splits were done to specifically analyze the group of participants which were hypothesized to be the unskilled and unaware (the bottom quartile on the MAI).

A regression analysis indicated that scores on the MAI did not significantly predict scores on the post-lecture exam, $\beta = -.10$, $t(47) = -.64$, $p = .525$. Additionally, the highest self-reported scores on the MAI ($n = 12$) were compared to the lowest self-reported scores on the MAI ($n = 12$). An independent two-sample t-test revealed no difference in exam scores – as seen in Table 1 - between the top MAI quartile and the bottom MAI quartile, $t(22) = .72$, $p = .878$. Holistically, these results suggest that participants who self-reported higher scores on the MAI did not perform better on the exam than participants who self-reported lower scores on the MAI.

Hypothesis 2: Metacognitive accuracy and the MAI

The second hypothesis was that participants in the highest self-reported quadrant of the MAI would be significantly more accurate on their metacognitive judgments compared to the participants in the lowest self-reported quadrant of the MAI. To support the second hypothesis, the results would have to show a predictive ability of the MAI regarding accuracy of metacognitive judgments. This hypothesis was tested with global predictive and postdictive judgments, and item-by-item relationships between item accuracy and confidence. Quartile splits were done again to try and illuminate any differences between the highest and lowest MAI scores.

Global Predictive Judgments

Global predictive judgments were provided after watching the lecture but prior to taking the exam. For this facet of the second hypothesis to be true, higher scores on the MAI should predict more accurate global predictive judgments. It was hypothesized that these differences would be lower for participants with higher self-reported MAI scores, and that that the participants in the highest quadrant of MAI scores would show under-confidence while the participants in the lowest quadrant of MAI scores would show over-confidence.

A simple regression was run to test the MAI's predictive ability for the accuracy of global predictive judgments. Overall scores on the MAI did not predict the accuracy of global predictive judgments compared to true exam scores, $\beta = .15$, $t(45) = .97$, $p = .336$. A two-sample t-test was run to compare the top and bottom quartiles on the MAI: This test found no difference between the top quartile on the MAI ($M = -.82$, $SE = 6.98$) and the bottom quartile on the MAI ($M = 4.92$, $SE = 6.43$) on the average differences between predictive judgments and true exam scores, $t(21)= .61$, $p = .551$. These results suggest no relationship between MAI scores and accuracy of global predictive judgments.

Global Postdictive Judgments

Another simple regression was run to evaluate if MAI scores predicted the accuracy of global postdictive judgments. These judgments were provided after both watching the lecture and completing the post-lecture exam. For this facet of the second hypothesis to be true, higher scores on the MAI should predict more accurate global postdictive judgments. Similar to the global predictive accuracy, the global postdictive accuracy was measured by comparing the absolute value of the differences between postdictive judgments and exam scores.

A simple regression was run to test the MAI's predictive ability for the accuracy of global postdictive judgments. Overall scores on the MAI did not predict the accuracy of global postdictive judgments compared to true exam scores, $\beta = .21$, $t(47) = 1.44$, $p =$

.156. A two-sample t-test was run to compare the top and bottom quartiles on the MAI: This test found no difference between the top quartile on the MAI (*M* = -5.88, *SE* = 6.08) and the bottom quartile on the MAI (*M* = 5.08, *SE* = 5.08) on the average differences between postdictive judgments and true exam scores, $t(22)= 1.38$, $p = .180$. These results suggest no relationship between MAI scores and accuracy of global postdictive judgments.

Item-By-Item Accuracy

A simple regression was run to assess if MAI scores predict gamma correlations between item accuracy and confidence judgments. It was hypothesized that participants with higher MAI scores would demonstrate more metacognitive accuracy (as indicated by stronger gamma correlations). Overall scores on the MAI did not significantly predict gamma correlations between item accuracy and item confidence, $\beta = -.21$, $t(47) = -1.41$, $p =.164$.

An independent-sample t-test was run comparing the bottom quartile on the MAI and the top quartile on the MAI for average gamma correlations between item-by-item confidence and item accuracy. This test did not indicate a significant difference between the top and bottom quartiles on the MAI, but the difference was trending towards significance, $t(22)=1.20$, $p = .089$. However – in contrast to what this study proposed - the bottom quartile on the MAI outperformed the top quartile on the MAI in terms of accuracy (see Table 1 for a full comparison by quartiles). In summary, none of the hypotheses related to the MAI were confirmed: There were no statistically significant findings about significant relationships between the MAI and global predictive judgments, item-by-item gamma correlations, or global postdictive judgments.

Discussion

The present study aimed to assess whether the MAI – a metacognitive self-assessment measure commonly used in education contexts – can predict performance and metacognitive accuracy on an educationally-relevant task. It was hypothesized that higher scores on the MAI would predict a) higher post-lecture exam scores, b) more accurate pre-lecture confidence judgements, c) more accurate post-lecture confidence judgments, and d) more accurate item-by-item confidence judgements. Data did not support these hypotheses, which suggests that the MAI – in its original 52-question true or false form – may not be predictive of student learning or the accuracy of student confidence judgments. If the MAI is supposed to be a measure of metacognitive awareness, and metacognitive awareness is an essential component of being a good student (Dunlosky & Metcalfe, 2009; Rhodes, 2020; Soderstrom et al., 2016), then why was there no relationship between any of the aforementioned variables?

There are several possible reasons which range from the nature of self-report to the MAI itself. First, self-report measures often have threats to validity (Boekaerts & Corno, 2005; Cromley & Azevedo, 2006). It is challenging for many participants to assess their strengths and weaknesses; there are often extraneous factors – like mood and recent events – that influence participants' responses (Nelson, 1984; Winne & Perry, 2005). Using a self-report measure on the topic of metacognition is inherently even more challenging: Participants are required not just to articulate how well they know something (i.e. how well they know facts about United States history) but they must perform a higher-level cognitive process evaluating their own cognitions (i.e. they must report whether they know how well they did on an exam). This higher-level self-assessment could be inherently challenging and perceived differently across participants. Specifically, lower-performing students might especially struggle with self-evaluations, further compromising a self-report study designed to target this population (Krueger & Dunning, 1999).

Even though self-report has challenges, according to a review done by Dinsmore et al. (2008) , self-report is still the second-most-used method to collect information about metacognition. Because the literature widely reinforces the use of self-report when studying metacognition, self-report alone is likely not the sole factor leading to rejecting the hypotheses of the present study. Another possible explanation for the lack of relationship between MAI scores and other variables is due to the method of response employed in the MAI. True or false options alone can be interpreted differently by different students (Winne & Perry, 2005) and these answer options also often do not encompass a wide enough range of specific options for participants (Tourangeau et al., 2000). For example, participants may not believe that a statement is fully true for them, but when only presented with polarizing options – true or false – they are forced to make a choice that may not accurately represent their true opinion.

Because of this response range issue, the MAI has been re-structured in some studies to try and draw out relationships (Coutinho, 2007; Young & Fry, 2008). One method that provides a wider range is to use a scale ranging from one through five paired with phrasing about how applicable the concept or habit is to the participants, (Harrison & Vallin, 2017). For example, the first question on the MAI states, “I ask myself periodically if I am meeting my goals” (Schraw & Dennison, 1994). Instead of participants evaluating an all-or-nothing true or false evaluation, they would instead have the option of choosing five options ranging from, “Always true for me,” to “Always false for me.” These expansions on the MAI take the original inventory a step farther to try and pull useful information, but to date there is still little evidence that the MAI is an effective empirical tool for drawing conclusions and making predictions.

Because this study replicated the robust finding of the Dunning-Kruger effect – even though this study was highly underpowered – this study expands on previous literature to provide more support for the Dunning-Kruger effect. The Dunning-Kruger effect was still present in this study despite both a) too small of a sample size and b) a restricted range. The postdictive judgments are perhaps the best test of the Dunning-Kruger effect (Kruger & Dunning, 1999; Hacker, 2000; Hartwig & Dunlosky, 2014), so the ability of the top quartile of participants to calibrate their postdictions after the exam while the participants in the bottom quartile still remained overconfident in their postdictions replicates the Dunning-Krueger effect.

Other researchers argued that when examining confidence judgments, unskilled-unaware participants exhibit lower confidence than their peers (even though they are still overconfident of their own performance), so naturally it appears that they are somewhat aware of their incompetence (Handel & Fritzsche, 2015). Still others have argued that all individuals, whether skilled or unskilled, are likely to struggle to correctly calibrate their metacognitive judgments due to a fairly universal lack of insight (Burson et al., 2006). However, after further research and statistical control for the above concerns with the Dunning-Kruger effect, some argue that the dramatic miscalibration by the lowest-performing students is still present (Handel & Dresel, 2018; Kruger & Dunning, 2002; Sheldon et al., 2015). This effect holds across a variety of arenas; the three different arenas in the original study (Kruger & Dunning, 1999) and emotional intelligence (Sheldon, et al., 2015) just to name a few. For that reason, the replication of the Dunning-Kruger effect in this study seems consistent with the literature.

Limitations

There were several limitations to the present study, both within the parameters of the study and also due to unforeseen challenges affecting many institutions of higher education. The most glaring limitation to the study is the small sample size (*n*=49). Due to the nature of the study and the aim to compare participants within a quartile split, the ideal number of participants for this study would have been a minimum of 120. However, the study was run just pre-dating the outbreak of COVID-19. It was not possible to continue collecting participants due to the international pandemic, and research participation had to be halted at 49. Because of the small sample size, questions linger about whether there might be different findings with a larger sample. Given the fact that there were some trends towards significance (i.e. the relationship between item-by-item gamma correlations and MAI scores), it is feasible that more significant relationships might have appeared with a larger sample. COVID-19 also made the analysis and presentation of these data more challenging. With universities going online and in-person contact being mostly if not wholly prohibited, interpreting the data from the present study had to be done primarily in isolation: A process not typical of data in the social sciences.

Another limitation of this study is that the original 52-item true or false version of the MAI was chosen as the primary measure of interest. There have been other studies that may have more practical variations of the MAI (Magno, 2008, 2010; Young & Fry,

2008) – including different scoring options and reverse-scoring questions - which may have been more appropriate for the present study as well. Most recently, Harrison and Vallin (2017) proposed a 19-item version of the MAI through confirmatory factor analysis. Other proposals for scoring the original 52-item version have also been proposed (Hughs, 2015; Youngs & Fry, 2008). Building on one of these previous studies – for example, trying the 19-question version of the MAI proposed by Harrison and Vallin (2017) - may have elicited trends that provide valuable information.

One final limitation is that the MAI could potentially be tapping a more global metacognitive ability that does not necessarily apply to specific situations very well. Participants in the present study were not tracked across multiple classes or even multiple sessions; they only experience a one-hour session for the entire study. Given the small scope of the study, it is possible that there may be different findings in a multi-session study or a classroom study. Studies operating under these parameters elicited strong Dunning-Kruger effects (Hacker et al, 2000; Serra & DeMarree, 2016) and had strong metacognitive components. Comparing some of these more longitudinal elements with the MAI may have been a better way to assess what (if anything) the MAI can predict.

Future Directions

Although there are clear limitations with the present study, there are a few interesting takeaways as well. Primarily, these data build on previous data in suggesting that the original 52-question true or false version of the MAI likely does not have predictive abilities for student retention or accuracy of confidence judgments (Harrison & Vallin, 2017). However, there still may be a potential for tweaking current educational metacognitive inventories - like the MAI, the Motivated Strategies for Learning

Questionnaire or MSLQ (Pintrich & de Groot, 1990), and the Learning and Study Strategies Inventory or LASSI (Weinstein et al., 1987) – or developing a new inventory to act as an educational tool helping students monitor learning and instructors shape their courses. Future research should investigate various options of these inventories to see if there are valid, reliable forms of these assessments. It is currently an open question whether there is a metacognitive skill that can be tapped in education or metacognition is a domain-specific construct. Further studies should aim to continue to investigate this question.

Conclusion

This study provided an opportunity for a simple, easy to administer tool from educational psychology to be tested for its merits and implications as a tool to be used in post-secondary education. The results of the present study were largely underwhelming due to a myriad of factors, but these data suggest that the MAI is likely not a reliable tool for instructors. While it is important for instructors to gauge their students' learning (Ambrose et al., 2010), these data suggest that the MAI is probably not the best way to quantify student's study strategies and metacognitive abilities. There is simply too much variability in MAI scores, perception of true and false questions, and ability of students to accurately assess their ability to self-assess that the MAI leads to too much variance to draw conclusions or predictions. Until there is a simpler, reliable, valid version of the MAI (or a similar measure), instructors in post-secondary education should turn to other methods to gauge their students' abilities.

References

Ambrose, S.A., Bridges, M.W., DiPietro, M., Lovett, M.C., & Norman, M.K. (2010). *How learning works. 7 research-based principles for smart teaching.* Jossey-Bass.

Arslan, S., & Akin, A. (2014). Metacognition: As a predictor of one's academic locus of control. *Kuram Ve Uygulamada Egitim Bilimleri, 14*, 33-39.

Bellon, E., Wim, F., & De Smedt, B. (2020). Metacognition across domains: is the association between arithmetic and metacognitive monitoring domain-specific? *Plos One*. https://doi.org/10.1371/journal.pone.0229932

Boekaerts, M., & Corno, L. (2005). Self-regulation in the classroom: A perspective on assessment and intervention. *Applied Psychology, 54,* 199-231. Doi: 10.1111/j.1464-0597.2005.00205.x

Brown, A. L. (1987). Metacognition, executive control, self-regulation, and other more mysterious mechanisms. In F. Weinert & R. Kluwe (Eds.), *Metacognition, motivation, and understanding* (pp. 65-116). Mahwah, NJ: Erlbaum.

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. doi:10.1037/0022-3514.90.1.60

Coutinho, S. A. (2007). The relationship between goals, metacognition, and academic success. *Educate~: The Journal of Doctoral Research in Education, 7*(1), 39-47.

Cromley, J. G., & Azevedo, R. (2006). Self-report of reading comprehension strategies: What are we measuring? *Metacognition and Learning, 1*, 229-247. Doi: 10.1007/s11409-006-9002-5

Dinsmore, D. L., Alexander, P.A., & Loughlin, S. M. (2008). Focusing the conceptual lens on metacognition, self-regulation, and self-regulated learning. *Educational Psychology Review, 20,* 391-409. doi: 10.1007/s10648-008-9083-6

Dunlosky, J., & Metcalfe, J. (2009). *Metacognition*. Sage Publications, Inc.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist, 34,* 906-911. doi:10.1037/0003-066X.34.10.906

Geurten, M., Meulemans, T., & Lemaire, P. (2018). From domain-specific to domain-general? The developmental path of metacognition for strategy selection. *Cognitive Development, 48,* 62-81. https://doi-org.ezproxy.library.ewu.edu/10.1016/j.cogdev.2018.08.002

Händel, M., & Dresel, M. (2018). Confidence in performance judgment accuracy: The unskilled and unaware effect revisited. *Metacognition and Learning,* doi:10.1007/s11409-018-9185-6

Händel, M., & Fritzsche, E. S. (2016). Unskilled but subjectively aware: Metacognitive monitoring ability and respective awareness in low-performing students. *Memory & Cognition, 44*(2), 229-241. doi:10.3758/s13421-015-0552-0

Hartwig, M. K., & Dunlosky, J. (2014). The contribution of judgment scale to the unskilled-and-unaware phenomenon: How evaluating others can exaggerate over- (and under-) confidence. *Memory & Cognition, 42*(1), 164-173. doi:10.3758/s13421-013-0351-4

Harrison, G.M, & Vallin, L.M. (2017). Evaluating the Metacognitive Awareness Inventory using empirical factor-structure evidence. *Metacognition and Learning*, 13(1), 15–38. https://doi-org.ezproxy.library.ewu.edu/10.1007/s11409-017-9176-z

Hughs, J. A. (2015). *Impact of online self-regulated professional development on technology and engineering educators metacognitive awareness.* (Doctoral dissertation). ProQuest Dissertations & Theses Global database.

Kim, Y., Chiu, C., & Bregant, J. (2015). Unskilled and don't want to be aware of it: The effect of self-relevance on the unskilled and unaware phenomenon. *Plos One, 10*(6)

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*(2), 180-188. doi:10.1037/0022-3514.82.2.180

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121-1134. doi:10.1037/0022-3514.77.6.1121

Kruger, J., & Dunning, D. (2002). Unskilled and unaware--but why? A reply to krueger and mueller (2002). *Journal of Personality and Social Psychology, 82*(2), 189-192. doi:10.1037/0022-3514.82.2.189

Magno, C. (2008). Reading strategy, amount of writing, metacognition, metamemory, and apprehension as predictors of English written proficiency. *Asian EFL Journal, 29*, 15-48.

Magno, C. (2010). The role of metacognitive skills in developing critical thinking. *Metacognition and Learning, 5,* 137-156. doi: 10.1007/s11409-010-9054-4

Mattern, K. D., Burrus, J., & Shaw, E. (2010). When both the skilled and unskilled are unaware: Consequences for academic performance. *Self and Identity, 9*(2), 129-141. doi:10.1080/15298860802618963

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. Psychological Bulletin, 95, 109-133. doi: 10.1037/0033-2909.95.1.109

Nelson, T. O., Narens, L., & Dunlosky, J. (2004). A revised methodology for research on metamemory: Pre-judgment recall and monitoring (PRAM). *Psychological Methods, 9*(1), 53-69. doi:10.1037/1082-989X.9.1.53

Pedone, R., Semerari, A., Riccardi, I., Procacci, M., Nicolò, G., & Carcione, A. (2017). Development of a self-report measure of metacognition: The metacognition self-assessment scale (MSAS) instrument description and factor structure. *Clinical Neuropsychiatry: Journal of Treatment Evaluation, 14*(3), 185-194.

Pintrich, P. R., & de Groot, E. V. (1990). Motivated Strategies for Learning Questionnaire. Retrieved from PsycTESTS. Doi: 10.1037/t09161-000.

Reid, A. J., Morrison, G. R., & Bol, L. (2017). Knowing what you know: Improving metacomprehension and calibration accuracy in digital text. *Educational Technology Research and Development, 65*(1), 29-45. doi:10.1007/s11423-016-9454-5

Rhodes, M. G. (2019). Metacognition. *Teaching of Psychology, 46*(2), 168–175. https://doi-org.ezproxy.library.ewu.edu/10.1177/0098628319834381

Rhodes, M.G., Cleary, A.M., & DeLosh, E.L. (2020). *A guide to effective studying and learning. Practical strategies from the science of learning.* New York, NY: Oxford University Press.

Schellings, G., & Van Hout-Wolters, B. (2011). Measuring strategy use with self-report instruments: Theoretical and empirical considerations. *Metacognition and Learning, 6*, 83-90. doi: 10.1348/000709906X106147

Schraw, G. (1997). The effect of generalized metacognitive knowledge on test performance and confidence judgments. *Journal of Experimental Education, 65*(2), 135-146. doi:10.1080/00220973.1997.9943788

Schraw, G. & Dennison, R.S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology*, 19, 460-475.

Schraw, G. Dunkle M.E., Bendixen, L.D., & Roedel, T.D. (1995). Does a general monitoring skill exist? *Journal of Educational Psychology, 87*(3), 433-444. https://doi-org.ezproxy.library.ewu.edu/10.1037/0022-0663.87.3.433

Schraw, G., & Graham, T. (1997). Helping gifted students develop metacognitive awareness. *Roeper Review: A Journal on Gifted Education, 20*(1), 4-8. doi:10.1080/02783199709553842

Schraw, G., & Moshman, D. (1995). Metacognitive theories. *Educational Psychology Review, 7*(4), 351-371. doi:10.1007/BF02212307

Semerari, A., Cucchi, M., Dimaggio, G., Cavadini, D., Carcione, A., Battelli, V., & Smeraldi, E. (2012). The development of the metacognition assessment interview: Instrument description, factor structure and reliability in a non-clinical sample. *Psychiatry Research, 200*(2-3), 890-895. doi:10.1016/j.psychres.2012.07.015

Serra, M. J., & DeMarree, K. G. (2016). Unskilled and unaware in the classroom: College students' desired grades predict their biased grade predictions. *Memory & Cognition, 44*(7), 1127-1137. doi:10.3758/s13421-016-0624-9

Sheldon, O. J., Dunning, D., & Ames, D. R. (2014). Emotionally unskilled, unaware, and uninterested in learning more: Reactions to feedback about deficits in emotional intelligence. *Journal of Applied Psychology*, *99*(1), 125-137. doi:10.1037/a0034138

Siedlecka, M., Skóra, Z., Paulewicz, B., Fijalkowska, S., Timmermans, B., & Wierzchon, M. (2018). Responses improve the accuracy of confidence judgements in memory tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* doi:10.1037/xlm0000608

Soderstrom, N. C., Yue, C. L., & Bjork, E. L. (2016). Metamemory and education. In J. Dunlosky, & S. K. Tauber (Eds.), (pp. 197-215). New York, NY: Oxford University Press.

Todd, S. (2010). *Life in the world's oceans* [Film]. The Great Courses Plus.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response.* New York: Cambridge University Press.

Weinstein, C., E., Schulte, A., & Palmer, D., R. (1987). *The learning and study strategies inventory* Clearwater, FL: H&H Publishing.

Winne, P. H., & Perry, N. E. (2005). Measuring self-regulated learning. In M. Boekaerts, P. R. Pintrich & M. Zeidner (Eds.), *Handbook of self-regulation* (pp 531-566). Burlington, MA: Elsevier Academic Press.

Young, A., & Fry, J.D. (2008). Metacognitive awareness and academic achievement in college students. *Journal of the Scholarship of Teaching and Learning, 8*(2), 1-10.

Appendixes

Appendix 1

The Metacognitive Awareness Inventory (Modified for this study)

MAI

The following questions address your habits related to your learning strategies and processes. Specially, please address these questions in an educational framework. Please answer true or false as appropriate. Please try and answer as truthfully as possible.

		True	False
1	I ask myself periodically if I am meeting my goals.		
2	I consider several alternatives to a problem before I answer.		
3	I try to use strategies that have worked in the past.		
4	I pace myself while learning in order to have enough time.		
5	I understand my intellectual strengths and weaknesses.		
6	I think about what I really need to learn before I begin a task.		
7	I know how well I did once I finish a test.		
8	I set specific goals before I begin a task.		
9	I slow down when I encounter important information.		
10	I know what kind of information is most important to learn.		
11	I ask myself if I have considered all options when solving a problem.		
12	I am good at organizing information.		
13	I consciously focus my attention on important information.		
14	I have a specific purpose for each strategy I use.		
15	I learn best when I know something about the topic.		
16	I know what the teacher expects me to learn.		
17	I am paying attention and am answering these questions accurately.		
18	I am good at remembering information.		
19	I use different learning strategies depending on the situation.		
20	I ask myself if there was an easier way to do things after I finish a task.		
21	I have control over how well I learn.		
22	I periodically review to help me understand important relationships.		
23	I ask myself questions about the material before I begin.		
24	I think of several ways to solve a problem and choose the best one.		
25	I summarize what I've learned after I finish.		
26	I ask others for help when I don't understand something.		
27	I can motivate myself to learn when I need to.		
28	I am aware of what strategies I use when I study.		
29	I find myself analyzing the usefulness of strategies while I study.		
30	I use my intellectual strengths to compensate for my weaknesses.		
31	I focus on the meaning and significance of new information.		

32	I create my own examples to make information more meaningful.		
33	I am a good judge of how well I understand something.		
34	I find myself using helpful learning strategies automatically.		
35	I find myself pausing regularly to check my comprehension.		
36	I know when each strategy I use will be most effective.		
37	I ask myself how well I accomplish my goals once I'm finished.		
38	I draw pictures or diagrams to help me understand while learning.		
39	I ask myself if I have considered all options after I solve a problem.		
40	I try to translate new information into my own words.		
41	I have stopped paying attention and no longer am answering accurately.		
42	I change strategies when I fail to understand.		
43	I use the organizational structure of the text to help me learn.		
44	I read instructions carefully before I begin a task.		
45	I ask myself if what I'm reading is related to what I already know.		
46	I re-evaluate my assumptions when I get confused.		
47	I organize my time to best accomplish my goals.		
48	I learn more when I am interested in the topic.		
49	I try to break studying down into smaller steps.		
50	I focus on overall meaning rather than specifics.		
51	I ask myself questions about how well I am doing while I am learning something new.		
52	I ask myself if I learned as much as I could have once I finish a task.		
53	I stop and go back over new information that is not clear.		
54	I stop and reread when I get confused.		

Appendix 2

The 25-Question Exam

Post-Lecture Exam

After taking the test on the lecture material, please answer the following question to the best of your ability.

1. What term relates to photosynthetic and chemosynthetic activity of primary producers?
    a. Productivity
    b. Luciferescence
    c. Selectivity
    d. Chlorophyll A

    On a scale of 25% (random chance) and 100% (absolutely certain), how confident are you in the accuracy of your answer to this question?

    ____

2. What is another name for "The Polar Front?"
    a. The Antarctic Convergence
    b. The Artic Convergence
    c. The Sub-Antarctic Surface
    d. The Arctic Surface

    On a scale of 25% (random chance) and 100% (absolutely certain), how confident are you in the accuracy of your answer to this question?

    ____

3. Which word is assigned to krill by the lecturer to denote krill's significance to their ecosystem?
    a. Keystone species
    b. Apex species
    c. Vital species
    d. Fundamental species

    On a scale of 25% (random chance) and 100% (absolutely certain), how confident are you in the accuracy of your answer to this question?

    ____

4. What is the term referring to the process of an organism using their own light to navigate and function?
    a. Luciferescence
    b. Photosynthesis
    c. Bioluminescence
    d. Chemosynthesis

On a scale of 25% (random chance) and 100% (absolutely certain), how confident are you in the accuracy of your answer to this question?

____

5. What is the length of an average krill?
   a. 1-2 cm long
   b. 3-4 cm long
   c. 5-6 cm long
   d. 7-8 cm long

   On a scale of 25% (random chance) and 100% (absolutely certain), how confident are you in the accuracy of your answer to this question?

   ____

6. What is the name of the bluff which was named for two ruined expeditions re-uniting which the lecturer talks about in the first part of the lecture?
   a. Cape Hope
   b. Cape Larson
   c. Cape Reunion
   d. Cape Well-Met

   On a scale of 25% (random chance) and 100% (absolutely certain), how confident are you in the accuracy of your answer to this question?

   ____

7. Most of our current fisheries are _______ and ________.
   a. Open water; temperate
   b. Open water; extreme
   c. Shelf-based; temperate
   d. Shelf-based; extreme

   On a scale of 25% (random chance) and 100% (absolutely certain), how confident are you in the accuracy of your answer to this question?

   ____

8. Which of the following is true of krill?
   a. They are herbivores
   b. They are carnivores
   c. They are omnivores
   d. Their diet is mostly unknown to researchers

   On a scale of 25% (random chance) and 100% (absolutely certain), how confident are you in the accuracy of your answer to this question?

____

9. In addition to leopard seals, what other species does the lecturer refer to as the natural predators of the Antarctic region?
    a. Emperor Penguins
    b. Great White Sharks
    c. Blue Whales
    d. Killer Whales

    On a scale of 25% (random chance) and 100% (absolutely certain), how confident are you in the accuracy of your answer to this question?

    ____

10. Chemosynthesis is a process that occurs entirely in the domain of:
    a. Air
    b. Light
    c. Water
    d. Bacteria

    On a scale of 25% (random chance) and 100% (absolutely certain), how confident are you in the accuracy of your answer to this question?

    ____

11. How is productivity measured?
    a. Grams of Hydrogen per meters squared per year
    b. Grams of Carbon per meters squared per year
    c. Grams of Mercury per meters squared per year
    d. Grams of Oxygen per meters squared per year

    On a scale of 25% (random chance) and 100% (absolutely certain), how confident are you in the accuracy of your answer to this question?

    ____

12. Which ecosystem that accounts for roughly 50% of the earth's surface does the lecturer state we know very little about?
    a. Abyssal Plane
    b. Antarctic Shelf
    c. Sub-Antarctic Trench
    d. Abyssal Trench

    On a scale of 25% (random chance) and 100% (absolutely certain), how confident are you in the accuracy of your answer to this question?

    ____

13. What is the name of the lecturer from the video you watched?
    a. Dr. Sean Todd
    b. Dr. Carl Larson
    c. Dr. Todd Larson
    d. Dr. Gary Cousins

    On a scale of 25% (random chance) and 100% (absolutely certain), how confident are you in the accuracy of your answer to this question?

    ____

14. What is the name of the group of predators that luminate their hunting light to a color that is invisible to their own predators?
    a. Glowfish
    b. Patagonian Toothfish
    c. Lightfish
    d. Dragonfish

    On a scale of 25% (random chance) and 100% (absolutely certain), how confident are you in the accuracy of your answer to this question?

    ____

15. Approximately what percentage of earth's ice is in Antarctica?
    a. 70%
    b. 75%
    c. 90%
    d. 95%

    On a scale of 25% (random chance) and 100% (absolutely certain), how confident are you in the accuracy of your answer to this question?

    ____

16. What important international treaty was signed in 1946 according to the lecture?
    a. The International Convention for Antarctic Wildlife
    b. The International Convention for the Regulation of Antarctic Wildlife
    c. The International Convention for the Regulation of Whaling
    d. The International Convention of Whaling Regulations

    On a scale of 25% (random chance) and 100% (absolutely certain), how confident are you in the accuracy of your answer to this question?

    ____

17. What is the reason cited in lecture that there is generally more productivity in shallow areas than deeper areas?
    a. Shallow areas have a more diverse animal population
    b. Shallow areas have a less diverse animal population
    c. Photosynthesis is linked to light availability, which is best in spring and summer
    d. Mixing processes in shallow areas allows nutrients to come to the surface

    On a scale of 25% (random chance) and 100% (absolutely certain), how confident are you in the accuracy of your answer to this question?

    ____

18. What does the lecturer refer to as "the last frontier on our planet?"
    a. Space
    b. The oceans
    c. Antarctica
    d. Knowledge about Antarctic ecosystems

    On a scale of 25% (random chance) and 100% (absolutely certain), how confident are you in the accuracy of your answer to this question?

    ____

19. What is the approximate surface area of Antarctica?
    a. 13 million kilometers squared
    b. 23 million kilometers squared
    c. 33 million kilometers squared
    d. 53 million kilometers squared

    On a scale of 25% (random chance) and 100% (absolutely certain), how confident are you in the accuracy of your answer to this question?

    ____

20. Which of the following types of animals was <u>not discussed</u> in lecture as being a predator of krill?
    a. Seabirds
    b. Whales
    c. Seals
    d. Sharks

    On a scale of 25% (random chance) and 100% (absolutely certain), how confident are you in the accuracy of your answer to this question?

    ____

21. How is most of Antarctica as a geographic region classified?
    a. A glacier
    b. An ice sheet
    c. A wasteland
    d. A desert

    On a scale of 25% (random chance) and 100% (absolutely certain), how confident are you in the accuracy of your answer to this question?

    ____

22. What is the best way to tell if you have crossed the polar front?
    a. The water changes from green to blue
    b. The water changes from blue to green
    c. There are markers in the water that denote the edges of the front
    d. The temperature of the water changes

    On a scale of 25% (random chance) and 100% (absolutely certain), how confident are you in the accuracy of your answer to this question?

    ____

23. How is productivity affected as one moves away from the equator - outside 40 degrees north and south of the equator?
    a. Productivity dramatically increases
    b. Productivity slightly decreases
    c. Productivity dramatically decreases
    d. There is no noticeable change in productivity at these cutoffs

    On a scale of 25% (random chance) and 100% (absolutely certain), how confident are you in the accuracy of your answer to this question?

    ____

24. What is the market name of the Patagonian Toothfish?
    a. Antarctic Sea Bass
    b. Chilean Sea Bass
    c. Patagonian Sea Tuna
    d. Southwestern Sea Tuna

    On a scale of 25% (random chance) and 100% (absolutely certain), how confident are you in the accuracy of your answer to this question?

    ____

25. Because of the absence of light, productivity in the abyssal plane must be:
    a. Chemosynthetic
    b. Photosynthetic
    c. Both chemosynthetic and photosynthetic
    d. Neither chemosynthetic nor photosynthetic

    On a scale of 25% (random chance) and 100% (absolutely certain), how confident are you in the accuracy of your answer to this question?

VITA

Author: Daniel A. Scheibe

Place of Birth: Richland, Washington

Undergraduate School Attended: Whitworth University, 2011-2015

Graduate School Attended: Eastern Washington University, 2018-2020

Degrees Awarded:	Bachelor of Arts in Psychology (Minor in Theology), 2015, Whitworth University
Honors and Awards:	Graduate Student Assistantship, 2018-2019, Eastern Washington University
	Academic All-American (Baseball), 2015, Whitworth University
	Dennis Spurlock Award for Leadership, 2015, Whitworth University
	Graduated Summa Cum Laude, 2015 Whitworth University
	Most Outstanding Major, 2015, Whitworth University
	Weyerhaeuser Younger Scholar Award, 2014, Whitworth University
	Provost’s Honor Roll, 2011-2015, Whitworth University
Professional Experience:	Cognitive Psychology Lab Manager, 2019-2020, Eastern Washington University
	President, Psi Chi and Psychology Club, 2014-2015, Whitworth University
	Vice President, Psi Chi and Psychology Club, 2014-2015, Whitworth University
	Vice President, Psychology Club, 2014-2015, Whitworth University
	Research Assistant 2018-2020, Eastern Washington University
	Research Assistant, 2012-2015, Whitworth University