JUDGEMENTS OF LEARNING IN THE CONTEXT OF WORKED MATHEMATICS QUESTIONS IN COLLEGE STUDENTS

by

Nicholas Anstead, B.S.

A thesis submitted to the Graduate Council of
Texas State University in partial fulfillment
of the requirements for the degree of
Master of Arts
with a Major in Psychological Research
August 2022

Committee Members:

William Kelemen, Chair

Rebecca Deason

Logan Trujillo

FAIR USE AND AUTHOR'S PERMISSION STATEMENT

Fair Use

Duplication Permission

ACKNOWLEDGEMENTS

First, I would like to thank Dr. Kelemen for introducing me to the psychological field of metacognition and helping me take these first few steps into what I see now as a fascinating field of psychology. I am grateful too for Dr. Deason and Dr. Trujillo, who used their expertise, advice and many, many revision suggestions to broaden the scope of this paper. My gratitude goes again to Dr. Trujillo, as well as to Dr. Ceballos and Dr. Lasser for writing the letters of recommendation that got me into this mess in the first place. It would be amiss to not thank my family for their support throughout my seven years at Texas State, from being a doe-eyed, giddy undergraduate student in 2015, to a now bleary-eyed, prostrated graduate student in 2022 — I could not have done it without you all. To Ryan, Ezra, and Maria, thank you for being the friends I could count on any day and any time to keep me company, to motivate me to study for even the most mundane courses, and to make me laugh myself to tears. To my fellow members of the GLG, thank you for the camaraderie we had on those late-night-turned-early-morning study sessions, last minute paper revisions, and making these last two years of grad school some of the best years of grad school I've ever had. Finally, Dr. Anstead, if you're reading this, get back to writing your book!

TABLE OF CONTENTS

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

LIST OF ABBREVIATIONS

Abbreviation	**Description**
EOL	Ease of Learning
FOK	Feeling of Knowing
G	Goodman-Kruskal Gamma Correlation
JOL	Judgement of Learning
MARS	Math Anxiety Rating Scale
M	Mean
SD	Standard Deviation
η_p^2	Partial Eta Squared

I. INTRODUCTION

Metacognition involves the self-assessment and control of one's own cognitive abilities. Buratti and Allwood (2002) wrote that metacognitive and cognitive processes do not necessarily operate separately from one another, but instead they have a constant give-and-take relationship wherein cognition feeds into metacognition, which provides feedback and control for the cognitive domain. Cognition does not require conscious awareness for tasks to be executed, and information can be passively gathered from the surrounding environment. Metacognition, in contrast, can involve both actively analyzing one's own cognition, requiring a significant amount of mental effort, as well as cognitive processes that are not consciously performed (Koriat & Levy-Sadot, 2000) By self-monitoring one can determine when a change in approach is needed or when a cognitive goal has been achieved.

Within the overarching term of metacognition, there is a subset of tasks involving judgments of one's own mental capabilities executed before the metacognitive processes flow from the metacognitive level to the cognitive level (Dunlosky & Metcalfe, 2009). Expanding our understanding of metacognition is essential to the process of learning, so to ask why it is important to research metacognition is to ask why it is important to learn at all. This research endeavors to expand existing research on metacognition from the verbal domain, where it has been studied extensively, into the area of mathematics, where only one published article exists.

II. ELEMENTS OF METACOGNITION

Dunlosky and Metcalfe (2009) proposed there are three stages of learning (acquisition, retention, and retrieval), with five metacognitive processes that can be associated across these stages. In the initial acquisition stage, an individual will begin with an Ease-of-Learning Judgement and a Judgement of Learning. Throughout the acquisition, retention and retrieval state, Feeling-of-Knowing Judgements are performed. In the final retrieval state, Source-Monitoring Judgements will be made and then a final judgement on the confidence that the retrieved information is correct.

Ease-of-Learning (EOL) judgements entail an individual evaluating how difficult learning about a given item will be in relation to other items. EOLs are made prior to actually studying the items. EOLs allow an individual to understand what content should be prioritized prior to studying the content. If an individual feels they have a weaker understanding of topic “A” than topic “B” then they may choose to study topic “A” more thoroughly to gain a better understanding of it. If a student were tasked to study for a physics test and an English test on the same day, the student will likely opt to give more study time to the class they believe they will have more difficulty studying, assigning the subjectively more difficult task a lower EOL judgement compared to the test in the class they believe to be easier to learn the material of.

A judgement of learning (JOL) is made during or after studying, where an individual evaluates how well they have learned a given set of words, a paragraph, a problem-solving method, etc. These JOLs can be performed at two time: An immediate JOL, wherein immediately after studying a given item, they will be asked for their JOL, or a delayed JOL, where a longer period of time passes after studying before the JOL is elicited (Dunlosky & Metcalfe, 2009). Inaccurate JOLs can lead to poor learning outcomes, whereas if a JOL was performed correctly then it can aid in targeting what needs to be learned. Additionally, a JOL can aid in deciding what specific manner of practice is more effective in the learning of a given task. Some methods of learning are better than others, and measurements of these JOLs in a research environment can aid in discovering the most effective way of learning a given task. For example, an individual may be practicing their competency on solving algebraic equations. When taking a moment to evaluate what areas are still in need of practice, a JOL is performed, estimating how well each portion of what they are studying has been learned. Once this JOL has been performed, the contents being studied can be prioritized for an optimal learning experience.

Feelings of knowing (FOKs) are metacognitive judgments that, while the individual cannot immediately recall the proper term, the individual will successfully recognize a given item if it is presented to them. FOKs can lead to the tip-of-the-tongue phenomenon, wherein a word, name, or phrase is perceived as being on the brink of being recalled but simply cannot pass the threshold from a cognitive state into an expression of the object or scenario in mind. Individuals may be able to accurately predict future performance on a currently inaccessible memory. If an individual were to be asked what sea is South of Italy and North of Libya, while they may not know immediately what the correct answer to the question is, but offers a high FOK, it can be predictive of the individual correctly identifying the Mediterranean Sea as the correct answer when presented with a list of bodies of water.

A source monitoring judgement is a metacognitive task wherein an individual mentally evaluates the source of a memory. Memories are rarely perfect recollections of a moment but are instead a conglomerate of many separate factors being combined to create an estimation of past events or information. Emotions play a significant role in piecing together these memories, with an individual's present mood acting as a lens through which to view the scenario. What may once have been a happy memory may be viewed with disdain depending on the mood an individual finds themselves in later. In the end, memories are not perfectly recalled, but they are reconstructed from both the fragments that are present in the mind and the physical cues around the subject at the present moment. For example, an individual may find themselves performing a source monitored judgement when asked to discuss the efficacy of using lavender oil to treat depression. The individual must first ask themselves if this treatment option was reported in a scientific journal, or if it was posted on an internet forum with no data to back the claim.

Finally, confidence in retrieval is an element of metacognition wherein an individual gives a judgement on the perceived accuracy of a recalled memory. This judgement can also be referred to as retrospective confidence. As with other types of judgments, retrospective confidence can be helpful or harmful depending on the accuracy of the judgments. Dunlosky and Metcalfe (2009) described a court case in which a key witness had a distorted memory of the incident, however the individual felt a high level of retrospective confidence. This scenario of a distorted memory of the incident coupled with a high level of retrospective confidence led the witness to be convinced that what they described was true, though later on their testimony was found to be false. Additionally, individuals can feel overconfident about what is considered common knowledge, which leads to being incorrect in what the individual was convinced was the correct answer.

III. BASIC METHODS OF JUDGEMENTS OF LEARNING

A commonly utilized method to investigate an individual's JOLs involves the cued recall of word pairs. In this task, a participant is given numerous pairs of words and asked to provide a JOL on how well they have learned each item. The pairs can be related, wherein the word pairs are commonly associated with the other (e.g., Sand-Beach) or unrelated, where there is no common factor between the two (e.g., Bread-Shoe) (Townsend & Heit, 2011). In word pair tasks, recall has been found to be significantly improved when one of the two words in a pair is given as a cue for the recall of the second word. This is further improved when both words in the pair are related to the other (Rhodes, 2016). The relation of words in a list is an example of an intrinsic cue for JOLs (Koriat, 1997). JOLs can also be influenced by extrinsic cues (e.g., type of presentation or speed of presentation) as well as mnemonic cues (e.g., ease of access). The present study focused on intrinsic cues because these have been found to be a powerful predictor of JOLs in single-list learning (Koriat, 1997).

The accuracy of JOLs can be assessed in two main ways, either through measures of calibration (absolute accuracy) or resolution (relative accuracy). The absolute accuracy of a JOL can be measured on a percentage scale, which allows a participant to give their JOLs an interval-level percent score rather than an ordinal-level self-assessment measure such as the traditional Likert scale. This percentage scoring system allows for the ability to compute the difference between predicted and actual recall. Scoring is calculated by subtracting the recall score from the participant's given JOL. For example, if a participant gives a mean JOL of 40% but earns a mean recall score of 50% then a difference score of -10% can be calculated. Naturally, we cannot expect a participant to give a perfect judgment of how well they have learned, and so we see this variation in the participant's JOL accuracy. For participants there can be either overconfidence, wherein a participant gives a mean JOL score higher than what they earned, (i.e. a mean JOL of 60% but a mean recall score of 40% for a net value of +20%) or underconfidence, where a participant scores higher than they anticipated (i.e. a mean JOL of 40% but a mean recall score of 60% for a net value of -20%) (Rhodes, 2016). Net positive scores represent overconfidence, whereas net negative scores indicate underconfidence. This signed difference between predicted and actual recall is known as a bias score.

Relative accuracy of a JOL can be measured with the Goodman-Kruskal Gamma, which is utilized when calculating correlations from ordinal data or higher. The Gamma correlation value ranges from -1, indicating an exact negative correlation, to +1, indicating an exact positive correlation; a value of 0 indicates that there is no correlation present (Ruiz & Arroyo, 2016). The Gamma correlation can be calculated by dividing the net number of correct pairs (i.e., when a lower JOL item is not recalled and a higher JOL item is recalled) minus the number of incorrect pairs divided by the total number of pairs. For example, a set of 86 pairs may have 40 correct JOL-recall pairs and 46 incorrect JOL-recall pairs, resulting in a net value of -6, divided by the total 86, resulting in a Gamma correlation of approximately -.07.

IV. VARIABLES IMPACTING JUDGMENTS OF LEARNING

There is a plethora of variables that can affect JOLs, both in the structuring of the JOL task and materials, and in regard to the physiological state of the participant. Within the structuring of the JOL task and materials, word pairs can vary in their relation to each other (related or unrelated, Myers, Rhodes & Hausman, 2020; Matvey, Dunlosky & Schwartz, 2006) timing of the JOL, (immediate or delayed, e.g., Matvey et. al 2006) and word-image pairing (e.g., Carpenter & Geller, 2020). As will be discussed below, there are also influences that are beyond the formatting of the JOL task itself and instead are based on the participant's physical status, including physical exercise (e.g., Zuniga, Mueller, Santana, & Kelemen, 2019) and pharmacological influences (e.g., Izaute, & Bacon, 2005).

Regarding word pair recall tasks that vary in their relation to one another, (e.g., sand-beach compared with apple-car,) it has been found that word pairs that are related to each other are reliably better remembered than those that are not. Additionally, JOLs were higher and more accurate for related pairs (Matvey et al. 2006). A second study using similar factors as in Matvey et al. also utilized a group that had cued-recall of word pairs and a second which had free-recall of word pairs. Each of these groups also had subsets of participants who either did or did not perform a JOL task. It was found that those in the group that performed a JOL with cued recall had a significantly higher percentage of word pairs recalled compared to the cued recall group that did not perform a JOL (Myers et al., 2020). Additionally, participants in Myers et al., (2020) demonstrated that giving JOLs on related item pairs resulted in significantly improved cued recall and item recognition than participants who did not offer JOLs. For unrelated items, only item recognition was significantly influenced when a participant offered a JOL.

Timing of the JOL can also produce a significant difference in the magnitude and accuracy of the JOL. Again, in Matvey et al., (2006) it can be seen that with the inclusion of delayed JOLs, the difference between related and unrelated item sets becomes significantly higher in magnitude compared to the immediate JOL group. In addition, both the related and unrelated item sets in the delayed JOL group had a greater JOL magnitude than the immediate JOL (Matvey et al., 2006). This finding has been called the Delayed-JOL effect (Nelson & Dunlosky, 1991).

Carpenter and Geller (2020) addressed JOLs in relation to image - Swahili pairings. In an experiment utilizing a delayed JOL, a set of cards displaying either an English to Swahili translation or an image and Swahili text, it was found that the JOLs for the image - Swahili pairings were significantly higher than the English - Swahili pairs. Additionally, it was found that these increased JOL ratings were in tandem with the test scores, with those that had the image - Swahili pairing scoring higher than the English - Swahili pairing. In a second experiment with immediate JOLs and identical image - Swahili and English - Swahili pairings, it was again demonstrated that JOLs were higher with the image - Swahili cards rather than the English - Swahili cards. Once more the JOLs were accurate, with those using image - Swahili cards scoring significantly better compared to the English - Swahili group (Carpenter & Geller, 2020).

Moving from manipulation of the JOL task to manipulation of participant-related variables, it can be found that there are state-based manipulations that can affect memory and JOLs differently. It has been found that the addition of mild-to-moderate exercise has an influence on an individual's memory recall (Zuniga et al., 2019). In that study, participants either remained sedentary or engaged in light or moderate exercise. Those who were placed in either exercise condition had significantly higher recall scores, though there was not a significant difference in memory between the intensities. Findings indicate that individuals across all conditions showed significant overconfidence regarding their ability to recall items presented to them in an item-by-item JOL task. Additionally, all three groups displayed moderate but significant metacognitive accuracy. However, the primary difference between the sedentary versus either exercise group was in recall, not JOL magnitude or accuracy (Zuniga et al., 2019).

Similar results were found in Salas et al., (2011) where participants engaged in mild cardio exercise. The study consisted of two phases where participants would either engage in mild cardio exercise or remain sitting. The combinations in these phases result in conditions where the participant may engage in exercise in both phases, exercise in the first but remain sitting in the second phase, sit in the first phase but exercise in the second, or remain sitting throughout the entire study session. All participants were presented with 30 nouns and asked to give an immediate JOL in between the first and second phase of the study. After the second task assigned to the participants, an untimed free recall test was given. Results of the study showed that participants who exercised prior to encoding demonstrated greater recall when compared to the sedentary condition. JOLs were not significantly impacted by exercise.

Examples of pharmacological influences in relation to JOLs can be found in Izaute and Bacon (2005). Participants were given either a small dose of the benzodiazepine lorazepam or a placebo. These participants were then given a set of nonassociated word pairs in a paired association task for encoding, followed by a 4-minute retention phase. Participants gave delayed JOLs on their completion of a word pair when cued by the first word in the pair. Participants were then subject to a free-recall test and asked about their confidence level. Results indicated that while those in the experimental group had impaired recall performance and made more commission errors, there was no significant difference between the control and experimental groups in terms of the accuracy of their JOLs. It was also found that the experimental group demonstrated overconfidence compared to accurate judgements by the control group.

Consumption of caffeine has also been tested in regard to memory and JOLs. In one study, participants took a moderate level of caffeine or a placebo before being given a word pair JOL task. After approximately 24 hours participants were either given the same level of caffeine as the day prior, or switched to either a placebo or a dose of caffeine. The participants were then asked to recall the word pairs shown to them the previous day. Participants who had the same caffeine dosage on both days had a higher level of recall than the group who had altered levels of caffeine consumption, however those who had altered levels of caffeine showed greater accuracy in their JOLs (Kelemen & Creeley, 2003). A similar two-day study also had participants take a moderate level of caffeine or a placebo and then complete free recall, cued recall and recognition memory tasks, followed by a JOL. As was the case in the previous caffeine study, participants were either given the same or different amount of caffeine on the second day of testing. Results indicated that the group that received caffeine on both days showed higher level of free recall compared to the other three conditions, and all four conditions had results similar to one another on the cued recall and recognition tasks (Kelemen & Creeley, 2001).

The usage of nicotine may have some effect on metacognitive processes as well, as can be seen in Kelemen & Fulton (2008). Participants with a history of smoking cigarettes were recruited for the study and were assigned to 1 of 4 conditions: A group having abstained from smoking for eight hours and were given nicotine-free gum, a group having abstained from smoking for eight hours and given gum containing nicotine, a group allowed to smoke as desired and given nicotine-free gum, and a group allowed to smoke as desired and given gum containing nicotine. Participants were asked to study a set of English nouns and provide a JOL for each word shown. Results showed that refraining from the usage of nicotine in the eight hours prior to the test had a significant negative impact on free recall as well as a reduction in the magnitude of the participant's JOLs. The usage of nicotine gum resulted in significantly greater sustained attention scores in both the restrained and unrestrained nicotine intake groups. Conversely, in Kelemen & Kaighobadi (2007) when participants were given cigarettes that either contained nicotine or were nicotine free, there were no significant differences in memory performance, memory recall or prospective remembering tasks.

V. JUDGEMENTS OF LEARNING FOR NON-VERBAL TASKS

Many of the examples so far have used verbal stimuli, and little research has examined JOLs for nonverbal stimuli. In the last decade there have been several papers, using a nonverbal problem-solving task (Baars, van Gog, de Bruin, & Paas, 2017a; Baars, van Gog, Paas, & de Bruin, 2017b; Baars, van Gog, de Bruin, & Paas 2018; Erickson & Heit, 2015). The studies by Baars et al. (2017a; 2017b; 2018) included participants at the elementary (Baars et al., 2017b; Baars et al., 2018) and high school (Baars et al., 2017a) education levels and focused on the optimal placement order of study materials, JOL-task order, and JOL duration.

Erickson & Heit (2015) utilized samples of both high school and college students to offer judgements on their performance when taking exams across multiple subjects. They conducted two studies to evaluate the absolute and relative metacognitive accuracy of samples of high school and college students regarding their accuracy in predicting their scores on tests across the subjects of biology, literature and mathematics. Participants in both studies were asked to predict their scores on each test prior to taking the tests, and to approximate their score after they had taken the tests. These predictions were then compared to the participants' actual scores.

In the high school participants, it was shown that participants displayed significant overconfidence in their predicted and post-exam scores versus their actual scores, with the greatest level of overconfidence in predictions being shown in the mathematics exam. When examining relative accuracy, there was a significant correlation between literature and mathematics prediction of scores and estimates of scores after the exams. Across all domains, there was a significant difference between predicted scores and actual scores, and post-test scores versus actual scores. These significant differences suggest that there was an increase in relative accuracy between the predicted and post-test estimations. Overall, there were greater levels of overconfidence prior to the examinations compared to afterwards, and mathematics produced the greatest level of overconfidence.

In the study using college students, questions across the same three domains (biology, literature and mathematics) were derived from the Scholastic Aptitude Test. In addition, participants were asked to complete the Math Anxiety Rating Scale (MARS). As with the high school students, there was overconfidence in the participants' scores, with the greatest level again being seen in the mathematics questions. It was shown there was a significant difference in the correlation between predicted and actual scores in biology and mathematics, and there were no significant differences in the school subjects in the scores post-test. Additionally, there were significant differences in predicted and post-test judgements compared to the actual correlations in literature and biology. There was generally overconfidence in math regardless of actual math scores. Participants reported an average score of 73 out of 115 on the MARS, indicating the college students had moderate levels of math anxiety during the study.

In a series of three studies, elementary and high school students were asked to study and give immediate or delayed JOLs on sets of logic-based problem-solving tasks involving changing water volumes in containers, Punnett squares, and mathematics questions (Baars, van Gog, de Bruin, & Paas, 2017a; Baars, van Gog, Paas, & de Bruin, 2017b; Baars, van Gog, de Bruin, & Paas 2018). In Baars et al. (2017b), elementary school children had five sections comprised of a mixture of a worked example, immediate JOLs, delayed JOLs, and an immediate problem. Test participants were placed into five different experimental groups: (a) a worked example, followed by an immediate JOL, (b) a worked example, followed by a delayed JOL, (c) a worked example, followed by an immediate problem and then an immediate JOL, (d) a worked example, followed by an immediate problem and then a delayed JOL and, (e) a worked example, followed by a delayed problem and then a JOL. It was hypothesized that the immediate JOLs would be more accurate than delayed JOLs. Additionally, it was hypothesized that delayed problem solving will increase the accuracy of JOLs more than immediate problem solving, basing this off of evidence from delayed generation strategies from learning expository text.

The example question included an instructional guide on how to do a logic-based task about altering the volume of water in several containers to have a predetermined end value in the final container. Results indicated that the first hypothesis was not supported, and did not find biases or deviation between participants who made immediate or delayed JOL after a worked example. The second hypothesis was also not supported as there were no significant differences in accuracy of JOLs made after delayed or immediate problem solving. The children also showed more accurate JOLs on less complicated problems but tended to overestimate performance on future problem-solving tasks.

Similarly, in a study including high school students, (Baars et al., 2017a) participants were taught how to use Punnett squares to find the probabilities of an organism’s offspring. A Punnett square is an equiangular quadrilateral diagram used to illustrate the mathematical probability of offspring having a given set of distinct traits passed on by either parent organism. The study looked at the timing of a JOL in relation to its accuracy after problem solving, hypothesizing that immediate JOLs will be more accurate than delayed JOLs in regard to problem solving and after studying a worked example. The same five conditions were used, (a) a worked example, followed by an immediate JOL, a worked example, followed by a delayed JOL, (b) a worked example, followed by an immediate problem and then a JOL; (c) a worked example, followed by an immediate problem and then a delayed JOL, or (d) a worked example, followed by a delayed problem and then a JOL. It was acknowledged that secondary school students would likely have better ability to self-monitor learning than primary school students, which may lead to more accurate JOLs. The main hypothesis was that solving a problem after a guided example would be an effective generation strategy. Additionally, it was hypothesized that delayed summary generation would lead to more accurate JOLs compared to immediate JOLs. Results indicated there was not a significant difference between immediate and delayed JOLs in both relative and absolute accuracy after practice. Regulation accuracy was high in those who had practice problems. Timing of practice problems was unexpectedly higher for immediate practice than delayed practice despite relying on long term memory to solve delayed practice. JOLs were found to be most accurate for the most complex tasks.

Some research has examined JOLs on problems directly describing math questions. Baars et al. (2018) tested a group of elementary school aged children (8 to 10 years old) to elicit JOLs after being shown how to solve similar math problems. Participants in this study were put into groups of either immediate JOLs or delayed JOLs. Participants in both groups were given a practice problem as an example of the questions that will be presented to the participants during the problem-solving phase of the experiment. The questions that were asked to be solved during the experiment entailed addition and subtraction problems, which the students' teacher confirmed were tasks that were similar, but not identical, to tasks the students had already learned. The mathematical questions were categorized as 1) "Addition without carrying" (e.g. 414 + 135 + 250 = ?), where the total of any value in the placements (ones, tens, or hundreds place) would not surpass 9 and thus carry over to the next placement, 2) "Addition with carrying" (e.g. 119 + 313 + 238 = ?) where a total value in the placement would surpass 9 and thus carry over into the next base-10 position, 3) "Subtraction with borrowing tens" (e.g. 676 - 139 = ?) where a value in the tens place is "borrowed" by the value in the ones place such that the value in the ones place being subtracted from is greater than the subtracting value[1], and "Subtracting with borrowing tens and hundreds" (e.g. 634 - 497 = ?, a process which is similar to the previous subtraction process, but now also including items in the hundreds place. Participants were asked to complete four questions of increasing complexity in each of the four categories and report their perceived difficulty of the question and approximate how much mental effort they put into each question using a 5-point Likert scale. After completing the mathematical questions where JOLs were offered, participants engaged in a posttest where they were asked to complete four questions with identical formatting but different values, each one pertaining to one of each of the four question formats. The participants did not have to offer JOLs on the questions; these questions were utilized to calculate JOL accuracy. JOLs were offered on

[1] In the example given, the 6 in the ones place must be greater than the 9 in the subtracting value. In order to resolve this, the 6 "borrows" a 1 value from the tens place to make the problem appear as 66(16) - 139, now allowing for 9 to be subtracted from 16 to then result in a value of 7.

a 5-point scale of 1 to 5 for each of the types of questions asked. Participants in the immediate JOL group gave their JOLs after each question, while the delayed JOL participants gave their JOLs after completing all four questions in a given category. Results for relative monitoring accuracy indicated that the mean Gamma correlation in the immediate JOL group had a significant difference from zero, while those giving a delayed JOL did not have a significant difference from zero. The absolute accuracy of JOLs did not differ significantly between the immediate and delayed JOL groups.

VI. PRESENT RESEARCH

The present research uses algebraic equations as the material to be learned instead of the traditional word-pairing or other verbal tasks that have been nearly ubiquitous in JOL tasks. Past research has shown that cues intrinsic to the items themselves are powerful predictors of JOLs in single-trial experiments. The present research examined two intrinsic cues, number of terms in the problem and number of steps in the solution. If these cues do not impact JOL magnitude, this finding would be a major deviation from past research using words or texts and could represent a boundary condition for Koriat's (1997) cue utilization approach.

The primary purpose of this study is to examine if college students can monitor their future performance of math problems successfully. If so, then Gamma correlations should be significantly non-zero and bias scores should not be significantly different from zero. Hypothesis 1a is that mean Gamma correlations will be moderately positive and will be significantly higher than zero, which would reflect significant relative metacognitive accuracy for these novel materials. Hypothesis 1b is that mean Bias scores will not be significantly different from zero, which would indicate accurate absolute metacognitive accuracy. Thus, Hypotheses 1a and 1b will extend work on JOLs using verbal materials in colleges into a new area using mathematical equations.

A second purpose of this study is to explore if JOLs are sensitive to two factors that can vary in solved mathematical equations: the number of terms in an equation and the number of steps in the solution. If so, then there should be significant differences in mean JOLs between conditions. Hypothesis 2 is that JOLs will be higher for equations with 2 terms compared with equations with 4 terms and that JOLs will be higher for 3-

step solutions compared with 4-step solutions. This finding would be the first to identify factors that can influence metacognition for mathematical equations. Finally, the accuracy of Gamma correlations and Bias scores will be compared across conditions to explore if the number of terms or length of solutions also has an impact on metacognitive accuracy. No specific predictions are made for these exploratory analyses.

VII. METHODS

Participants

A total of 36 participants provided usable data. Data from an additional 18 participants were excluded from all analyses: (a) 11 participants discontinued or were disconnected prior to completing the procedures and provided no correct answers on the test; (b) 2 participants did not comply with instructions and provided no correct responses on the test; and (c) 5 participants did not comply with instructions and provided a mean score of 1.2 out of 24 on the test. Thus, all analyses are based on a sample size of 36 participants, except where noted.

Participants were 18 years of age or over (M = 20.08, SD = 3.95). Participants were recruited from the Human Subjects Pool in the Psychology Department and were incentivized to participate by offering course credit in return for taking part in the study. The participants identified primarily as White/Caucasian at 75% of the study population, 8.3% identified as Black/African American, 5.6% identified as Asian, and 11.1% identified with other ethnicities. In regard to Hispanic identity, 44.4% participants identified as Hispanic, and 55.6% did not identify as Hispanic. For gender identity, 33.3% of participants identified as male, 63.9% identified as female, and 2.8% did not respond.

The math background of students varied slightly, with a majority having already taken a college level algebra course: 77.8% of participants had taken college algebra or greater (i.e., Sophomore, Junior or Senior level math courses), 16.7% of participants had taken a mathematics course that acted as a prerequisite to the College Algebra course but had not taken college algebra, and 5.6% of participants had not yet taken any math courses at the university. The average math anxiety level on the MARS was moderate (*M* = 2.74, *SD* = .83), where 1 indicates “Not at all” to 5, indicating “Very much.”

In order to establish the academic area of interest the participants had, the participants were asked for their major, which was then categorized by the specific college within Texas State University the students were a part of. This distribution can be seen on Table 1.

Table 1.
Number of Participants in Each College Within Texas State University

College	Number and Percent of Participants
College of Applied Arts	13(36.1%)
College of Business Administration	3(8.3%)
College of Education	3(8.3%)
College of Fine Arts and Communication	1(2.8%)
College of Health Professions	4(11.1%)
College of Liberal Arts	7(19.4%)
College of Science and Engineering	5(13.9%_

Design

This experiment used a 2 X 2 (number of terms X number of steps) within-subjects factorial design. Number of terms in the mathematical equation was a within-subjects independent variable: all participants received equations with either two terms (e.g., “2(x - 3)” being composed of “2” and “x - 3”) or four terms (e.g., “3(t + 2) + 2(t + 2)” being composed of “3”, “2”, and two “t + 2” sections) in the equation. Number of steps in the solution provided also was a within-subjects independent variable: half of the equations had three steps in the solution and the other half of the problems had four steps in the solution.

The main dependent variables were the magnitude of participants' JOLs for each equation, the accuracy in solving the equations, and the relationship between JOL magnitude and accuracy (i.e., metacognitive accuracy). Metacognitive accuracy was assessed with bias scores (absolute metacognition accuracy) and Gamma correlations (relative metacognitive accuracy).

Materials and Procedure

Participants were asked to complete an abbreviated version of the MARS to evaluate their levels of math test anxiety, numerical task anxiety, and math course anxiety (Alexander & Martray, 1989). The set of mathematical problems to be used in this study were adapted from the questions in Star et al. (2015), which were used to evaluate learning strategies. Participants will be presented with a set of these completed mathematical problems one at a time, asked to review the questions for 6 seconds and then give a JOL. Each participant will be shown 24 total equations. In order to divert the participants' attention away from mathematical problem-solving tasks and into a verbal problem-solving task, the participants were then be given a linguistic-based distraction task for 8 minutes, which asks participants to type a chain of as many names starting with the last letter of the prior name together (for example, Natalie-Elliot-Thomas). Finally, the same set of mathematical questions was presented (without the solution shown) for participants to solve. The test was untimed. The test was administered using a computer or mobile device with an internet connection. Given the study was being given online, the exact location of the participants at the time of the study varied.

VIII. RESULTS

Mean levels of JOLs, accuracy, bias scores, and *G*s are shown in Table 2. For *G*s, a large number of scores were indeterminant, which occurred when there was no variation in the 6 JOLs or accuracy scores in a particular condition. For example, *G* would be indeterminant if a participant used the same JOL for all 6 items in a condition, or recalled all 6 items in a condition correctly (or incorrectly). Overall, *G*s were indeterminant in 111 of 144 (77%) cases. Out of the 36 participants used in this study, only 15 (42%) provided determinate *G*s in at least one of the conditions.

Table 2.
Mean (and Standard Deviation) of Levels of JOLs, Accuracy, Bias Scores and Gs Across Conditions

Condition	JOLs	Accuracy	Bias	Gamma
2 Terms and 3 Steps	.882 (.186)	.926 (.141)	-.044 (.152)	.214 (.827)
2 Terms and 4 Steps	.924 (.134)	.898 (.170)	.026 (.213)	-.157 (.837)
4 Terms and 3 Steps	.858 (.192)	.843 (.192)	.016 (.232)	.076 (.857)
4 Terms and 4 Steps	.879 (.183)	.889 (.207)	-.01 (.224)	-.425 (.721)

The first set of hypotheses concerned whether or not participants could reliably predict future performance on the items. If so, then *G* correlations should be significantly non-zero whereas bias scores should not be significantly different from zero. To test these predictions, mean *G*s and bias scores were compared to 0 using one-sample *t*-tests.

Hypothesis 1a was that mean *G*s would be moderately positive and will be significantly higher than zero, which would reflect significant relative metacognitive accuracy for these novel materials. Conducting a set of one-sample *t*-tests across conditions, it can be seen that no significant results were obtained. In the 2x3 condition $t(7) = .71$, p =.502; in the 2x4 condition $t(6) = -.50$, $p = .635$; in the 4x3 condition $t(10) = .29$, $p = .775$; lastly, in the 4x4 condition $t(6) = -1.56$, $p = .170$. These results should be interpreted with the utmost caution, however, because of the very low numbers of valid *G*s that were obtained. Spearman's rho was not used as, similar to the *G*s, a large number of participants were correct on all questions across conditions

Hypothesis 1b was that mean Bias scores would not be significantly different from zero, which would indicate accurate absolute metacognitive accuracy. Conducting a set of one-sample *t*-tests across conditions, it can be seen that there were no significant differences from 0. In the 2x3 condition $t(35) = -1.72$, $p = .094$; in the 2x4 condition $t(35) = .73$, $p = .470$; in the 4x3 condition $t(35) = .40$, $p = .690$; in the 4x4 condition t(35) = -.27, $p = .786$. Although it is difficult to interpret null results, the mean values were close to 0 in all four conditions as predicted, suggesting that participants' bias was minimal overall.

The second purpose of this study was to explore if the number of terms and steps in the solution impacted mean JOLs, accuracy, and metacognitive accuracy between conditions. Hypothesis 2 was that JOLs would be higher for equations with 2 terms compared with equations with 4 terms and that JOLs would be higher for 3-step solutions compared with 4-step solutions.

A series of 2x2 within-subjects ANOVAs were conducted across mean JOLs, mean accuracy, and mean bias values (it was impossible to analyze *G*s across condition because no participants had valid *G*s in all four conditions). Several results in this study proved to be statistically significant. As predicted, it was found that mean JOL values were higher for 2-term equations (*M* = .90, *SD* = .16) compared with 4-term equations (*M* = .87, S*D* = .19), *F*(1,35) = 10.92, *p* = .002, η_p^2 = .238. There were also significant differences found according to the number of steps in the given algebra solution, *F*(1,35) = 6.06, *p* = .019, η_p^2 = .15. Contrary to expectations, however, JOLs were higher for 4-step solutions (*M* = .90, S*D* = .16) compared with 3-step solutions (*M* = .87, S*D* = .19). The interaction between the number of terms and the number of steps was not significant, *F*(1,35) = 1.07, *p* = .309, η_p^2 = .03.

Regarding the ANOVA results for mean accuracy, the number of terms was found to be non-significant, *F*(1,35) = 3.86, *p* = .058, η_p^2 = .10. The number of steps done to solve the equation was also non-significant, *F*(1,35) = .30, *p* = .586, η_p^2 = .01. The interaction between the number of terms and the number of steps in the equation was significant, *F*(1,35) = 4.38, *p* = .044, η_p^2 = .11. When conducting a post-hoc paired sample *t*-test for accuracy there was a significant difference between the 2-term, 3-step accuracy value and the 4-term, 3-step accuracy value, *t*(35) = 2.77, *p* = .009. When participants were solving equations with 4 terms it can be seen that participants were less accurate in both the 3-step and 4-step conditions, see Figure 1.

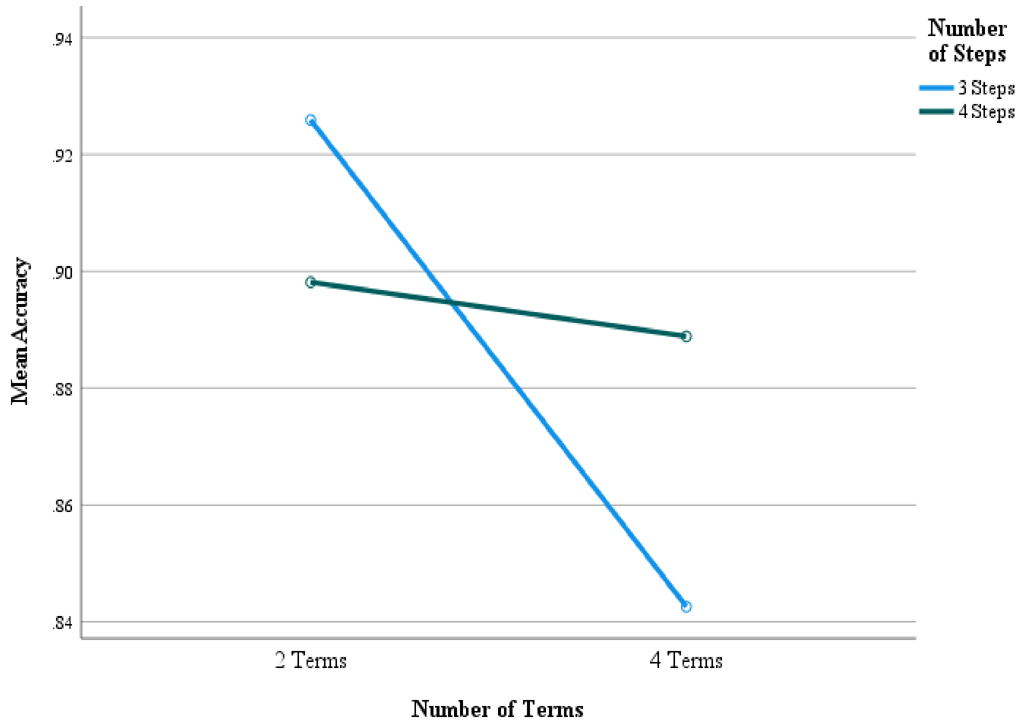

Figure 1. Mean Accuracy of participants' JOLs demonstrates a significant interaction between the number of steps and number of terms.

Similarly, when conducting the ANOVA for mean Bias, it was again found the number of terms was non-significant, $F(1,35) = .24$, $p = .624$, $\eta_p^2 = .01$. Additionally, the number of steps was non-significant, $F(1,35) = 1.17$, $p = .287$, $\eta_p^2 = .03$. There was a significant interaction between the number of terms and the number of steps, $F(1,35) = 5.18$, p = .029, $\eta_p^2 = .13$. Upon conducting a post-hoc paired samples *t*-test it can be seen that there was a significant difference between the 2-term, 3-step Bias value and the 2-term, 4-step Bias value, $t(35) = -2.43$, $p = .021$. The participants' mean Bias values increased between the 2-term and 4-term in the 3-step condition. Conversely, the mean Bias values in the 4-step condition decreased between the 2-term and 4-term equations, see Figure 2.

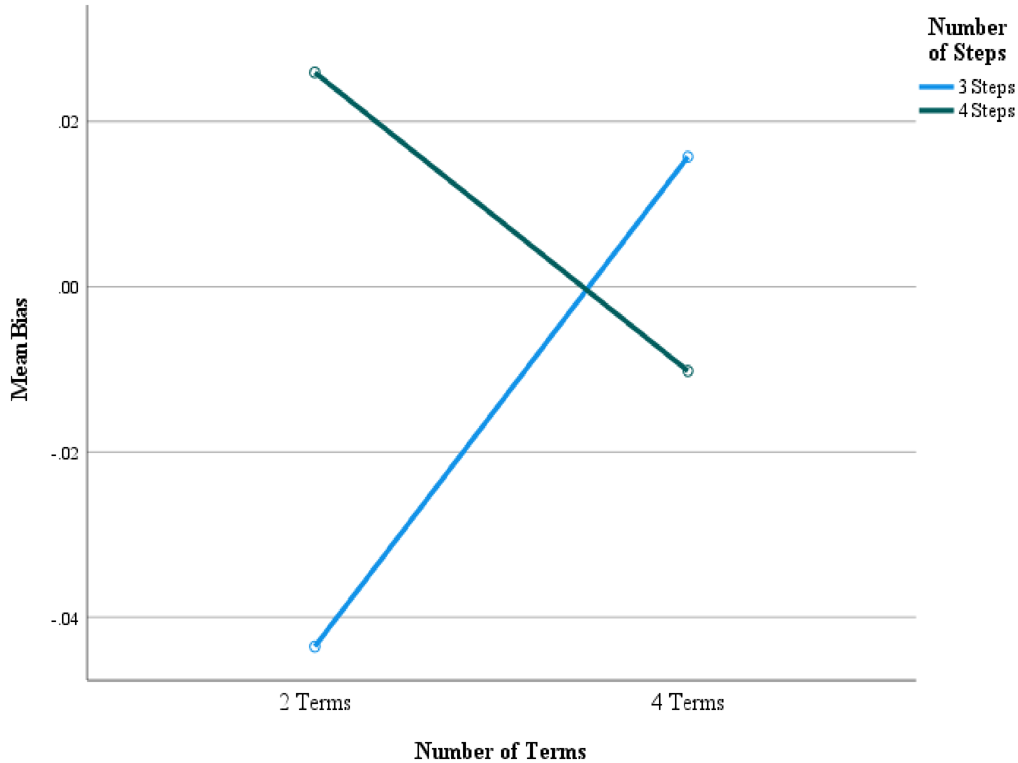

Figure 2. Mean Bias of participants' JOLs demonstrates a significant interaction between the number of steps and number of terms.

Exploratory analyses also were conducted to examine the relationship between MARS scores and the dependent variables. When conducting Pearson correlations between the participant's score on the MARS, some significant results can be seen. When looking at the participants' mean JOLs it can be seen that the 2x3 (r = -.41, p = .018), 2x4 (r = -.49, p = .001), 4x3 (r = -.48, p = .004) and 4x4 (r = -.50, p = .003) conditions were all statistically significant with moderate correlation in all conditions. Pearson correlations conducted between the participants' MARS scores and Accuracy did not have significant results in the 2x3 (r = -.21, p = .253), 2x4 (r = -.07, p = .695) and 4x4 (r = -.21, p = .238) conditions. The 4x3 condition proved to be significant, r = -.37, p = .033. There were no significant Pearson correlations when looking at Bias scores (2x3 r = -.29, p = .106; 2x4 r = -.25, p = .158; 3x4 r = -.01, p = .940; 4x4 r = -.20, p = .259).

IX. DISCUSSION

The purpose of this study was to examine if college students can monitor their future performance of math problems successfully, and if so, what factors impact their performance. The results indicated that participants were able to monitor their performance accurately. In addition, both the number of terms in the algebra equation and the number of steps in the solution acted as statistically significant predictors in JOLs. Additionally, it can be seen that the interaction between the number of terms in an algebra equation and the number of steps needed to solve it were significant in both accuracy and bias.

Hypothesis 1 concerned the overall accuracy of participants' JOLs using two measures of metacognitive accuracy. Hypothesis 1a was that Gs would be modest and significantly non-zero across conditions. Unfortunately, a definite conclusion regarding the study's results for hypothesis 1a cannot be reached as there was a very low number of *G*s that were calculated in relation to what was expected; only 23% of all cases allowed it to be possible for *G*s to be calculated. In fact, indeterminant *G*s were so common that none of the participants had valid *G*s across all four conditions. This is a result of many of the study's participants tendency to answer all 6 questions in any of the given equation formats correctly. Hypothesis 1b was that the participants' bias scores would not be significantly different from 0. This hypothesis was supported, with no significant values different from 0 using one-sample *t*-tests across conditions. Although it is difficult to interpret null results, the lack of significance in this suggests accurate absolute accuracy. Participants were neither significantly over- or underconfident. In Erickson and Heit (2015) it was shown that there was significant overconfident bias in regard to their score on the mathematics portion of the questions being asked, occurring in both the high school and college participants. While overconfidence was observed in that context of mathematics-based JOLs, in our study there was no significant bias in both the number of terms nor in the number of steps in the algebraic problems. In addition to differences in the type of questions, another possible explanation was that the present focus was on these two elements of terms and steps in each of the four sub-sets of algebraic problems, not on an overall scoring of all the problems.

Regarding hypothesis 2, when analyzing JOLs with ANOVAs it was found that the JOLs were significantly higher in algebraic equations that were composed of 2 terms versus 4 terms. This finding supported the prediction that equations with fewer terms would be judged as easier to solve. Contrary to predictions, however, JOLs were significantly greater in algebraic equations showing 4-step solutions as opposed to 3-step solutions. The cause of this surprising finding is unclear. It may indicate that students were more familiar with the type of solution used in the 4 step solutions compared with the 3-step solutions. It is possible that when students were being taught how to solve similar algebra problems that the instructors had shown students to distribute the coefficient to the values within the parentheses, a method that acted as a "catch-all" for solving algebra problems of this structure. In this study what was shown to participants were equations that were in a specific format that allowed for these shortcuts in solving the equation. In all the 4-term algebraic questions the values within the parentheses were identical for both of the terms (for example, "3(x +3) + 5(x + 3)"), and in all the 2-term algebra questions there was a common factor that both sides could be divided on (for example, in "3(x – 5) = 15," both sides could be divided into a whole number). The familiarity with the 4-step method of solving these equations may have allowed for the participants to more easily recognize the method provided to solve the equation which led to the participants giving a greater JOL for the method that they were more familiar with.

The exploratory analyses involving the MARS and JOL Pearson correlations showed that in all four conditions there was a moderate, negative correlation. This is to say that as an individual's level of math anxiety increases, their JOLs significantly decreased. Thus, as individuals who were more anxious about mathematics perceived themselves as being less able to solve the mathematics problems, which is consistent with previous findings (e.g., Erickson & Heit, 2015).

Looking back at the studies conducted by Baars et al., there are some parallels to be seen. For example, in Baars et al. (2017b), participants were more accurate when questions were less complicated, which mirrors this study where equations that were composed of 2 terms had greater accuracy than 4-term questions. In Baars et al. (2017a) JOLs were seen to be more accurate for problems which were more complex, contrasting both the findings from Baars et al. 2017b and our own findings in the 2 versus 4 terms conditions. However, this finding does have a similarity to the current study regarding hypothesis 2, where we found that the more complicated 4x4 condition was seen to have greater accuracy than the 4x3 condition. Baars et al. (2018) showed that when using mathematical questions, participants in the immediate JOL group showed mean Gamma correlations significantly different from zero. Unfortunately, in this study a majority of Gammas were indeterminate in any of the four conditions, thus making it difficult to make a definite conclusion. All four of the conditions used in this study proved to have non-significant results using one-sample *t*-tests. Our findings in regard to bias showed that there was no significant bias among the four conditions within the study. While this study focused on the bias between conditions of algebraic problems, the relation between mathematics and bias is contrary to what was previously observed by Erickson and Heit (2015), where participants showed significant changes in bias before and after test conditions.

Applications

This study is unique compared with other JOL research because very few studies have examined performance in mathematics as the outcome variable. While this study utilizes relatively simple algebraic equations, the same principle of what has been studied here can be applied to other fields of mathematics in both pure (i.e., topology, calculus, etc.) and applied mathematics (physics, statistics, etc.). In a practical sense, this work can aid in identifying what methods of solving mathematical questions students judge as more easily understood, which could help in mathematics instruction, particular in college-level algebra courses.

Limitations

This study did have some limitations in terms of the materials, online administration, and sample. First, overall performance was high on the math problems, and only 6 problems of each type were included across conditions. These factors led to a lack of variation in JOLs and actual performance within conditions that made analyses of *G*s untenable. Using a completely within-subjects design increased the sensitivity of our statistical tests but led to numerous cases in which there was no variation within one or more dependent variable for a condition.

A second concern is related to the online administration of the procedures.

Because participants were able to take part in this study unsupervised, with any resources available to solve the algebraic equations, it is uncertain what method they used to solve for the variable. Participants may have used the solution that was provided, a novel method, or outside resources (e.g., a friend, computer program or website). Generally, the online format for presenting JOL stimuli is adequate for use when looking for participants' correct pairing for word or image pairs, which cannot be looked up with the ease that an algebraic equation may allow.

A third concern involves the sample. Due to the nature of how the participants were recruited (i.e., through the Texas State University Psychology Department Human Subject Participant Pool), only a very specific sample could be used, which did not allow for the examination of those with less experience in algebraic problem solving. Given the education level of the participants, having a more varied sample with different levels of knowledge on how to solve algebraic problems may be beneficial, as it would allow for those that are less educated in algebraic problem solving to be evaluated. Using a student sample with greater variability in their mathematics background may have produced greater variation in responses (Alternatively, it may be the performance on these types of math problems is inherently bimodal). In any case, the ability to generalize to other students is limited by our use of convenience sampling.

Finally, there was a limitation regarding the question formatting. Due to practical constraints regarding the minimum number of steps to solve a given 2-term or 4-term equation (that is, equations had an absolute minimum number of 3 steps needed to solve for the variable), only 3- and 4-step solutions were displayed. This resulted in a fairly homogenous set of items which produced a lack of variability in performance.

Future Directions

While there has been tremendous growth in the study of JOLs over the past 30 years, nearly all of that work has been conducted using words (e.g., recall of single words, cued recall, or text comprehension). This study is among the first to examine whether basic JOL findings for words would extend to worked mathematics problems. The number of terms in the problem and steps in the solution represent what Koriat (1997) called "intrinsic" metamemory cues, which have been shown to impact JOLs using words and texts. This study is the first to demonstrate the impact of intrinsic cues on JOLs in the area of mathematics. We did not examine the other two types of cues he proposed, namely extrinsic cues (e.g., during of stimulus presentation, repetition of items, etc.) or mnemonic cues (e.g., retrieval fluency). Future researchers may wish to examine the relative impacts of those cues as well. In addition, it would be useful to determine whether or not the shift Koriat noted from a reliance on intrinsic cues to mnemonic cues with practice in word repetition extends to math problems.

The present study also has applied importance. This study utilized relatively simple algebraic equations from Star et al. (2015), which required only a basic understanding of algebra, such as understanding the principle of distributing a multiplier to values within parentheses. Given that approximately 78% of the study's participants had taken college algebra or more advanced mathematical courses, further studies should utilize increasingly complicated problem-solving methods in mathematics to evaluate the differences between JOLs and correct answers in what may be more challenging facets of mathematics. Future researchers may wish to include a combination of the present stimuli (which were relatively easy) with more complex problems. In addition, including more than six problems per condition would be advisable, to avoid indeterminant correlations due to lack of variation. Finally, it would be interesting to have mathematics instructors complete a similar version of this task, to determine if their perceptions of item difficulty match those of their students. Hopefully, this work will inspire future efforts to examine JOLs in mathematics to improve our understanding of the JOL theory as well as the practical processes involved in mathematics instruction overall.

REFERENCES

Alexander, L., & Martray, C. (1989). The Development of an Abbreviated Version of the Mathematics Anxiety Rating-Scale. *Measurement and Evaluation in Counseling and Development*, 22(3), 143-150

Baars, M., van Gog, T., de Bruin, A., & Paas, F. (2017). Effects of problem solving after worked example study on secondary school children's monitoring accuracy. *Educational Psychology*, *37*(7), 810–834. https://doi.org/10.1080/01443410.2016.1150419

Baars, M., van Gog, T., Paas, F, & de Bruin, A. (2017). Effects of problem solving after worked example study on primary school children's monitoring accuracy. *Applied Cognitive Psychology*, *28*(3), 382–391. https://doi-org.libproxy.txstate.edu/10.1002/acp.3008

Baars, M., van Gog, T., de Bruin, A., & Paas, F. (2018). Accuracy of primary school children's immediate and delayed judgments of learning about problem-solving tasks. *Studies in Educational Evaluation*, *58*, 51–59. https://doi.org/10.1016/j.stueduc.2018.05.010

Buratti & Allwood: Regulating Metacognitive Processes. In: Peña-Ayala, A. (2015). *Metacognition : fundaments, applications, and trends : a profile of the current state-of-the art*. Springer.

Carpenter, S. K., & Geller, J. (2020). Is a picture really worth a thousand words? Evaluating contributions of fluency and analytic processing in metacognitive judgements for pictures in foreign language vocabulary learning. *Quarterly Journal of Experimental Psychology*, *73*(2), 211–224. https://doi.org/10.1177/1747021819879416

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

Dunlosky, J., & Nelson, T. O. (1992). Importance of the kind of cue for judgements of learning (JOL) and the delayed-JOL effect. *Memory & Cognition*, *20*(4), 374–380

Erickson, S. & Heit, E. (2015). Metacognition and Confidence: Comparing Math to Other Academic Subjects. *Frontiers in Psychology, 6*. https://doi-org.libproxy.txstate.edu/10.3389/fpsyg.2015.00742

Izaute, M., & Bacon, E. (2005). Specific Effects of an Amnesic Drug: Effect of Lorazepam on Study Time Allocation and on judgment of Learning. *Neuropsychopharmacology*, *30*(1), 196–204.

Kelemen, W., & Creeley, C. (2001). Caffeine (4 mg/kg) influences sustained attention and delayed free recall but not memory predictions. *Human Psychopharmacology-Clinical and Experimental*, *16*(4), 309–319.

Kelemen, W. L., & Creeley, C. E. (2003). State-dependent memory effects using caffeine and placebo do not extend to metamemory. *The Journal of General Psychology, 130*(1), 70–86. https://doi.org/10.1080/00221300309601276

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*, 349-370. https://doi.org/10.1037/0096-3445.126.4.349

Matvey, G., Dunlosky, J., & Schwartz, B. (2006). The effects of categorical relatedness on judgements of learning (JOLs). *Memory*, *14*(2), 253–261. https://doi.org/10.1080/09658210500216844

Myers, S. J., Rhodes, M. G., & Hausman, H. E. (2020). Judgments of learning (JOLs) selectively improve memory depending on the type of test. *Memory & Cognition*, *48*(5), 745–758. https://doi.org/10.3758/s13421-020-01025-5

Nelson, T. O., & Dunlosky, J. (1991). When People's Judgments of Learning (JOLs) are Extremely Accurate at Predicting Subsequent Recall: The "delayed-JOL effect." *Psychological Science*, *2*(4), 267–270. https://doi-org.libproxy.txstate.edu/10.1111/j.1467-9280.1991.tb00147.x

Rhodes, M. G. (2016). Judgments of learning: Methods, data, and theory. In J. Dunlosky, S. K. Tauber, J. Dunlosky (Ed), & S. K. Tauber (Ed) (Eds.), *The Oxford handbook of metamemory.* (pp. 65–80). Oxford University Press.

Ruiz, M., & Arroyo, C. (2016). JOLer: A Java standalone application for simulating the Weaver & Kelemen's judgment of learning (JOL) model. *Anales de Psicología*, *32*(3), 893–898.

Salas, C. R. (1), Minakata, K. (2), & Kelemen, W. L. (2). 2011). Walking before study enhances free recall but not judgement-of-learning magnitude. *Journal of Cognitive Psychology*, *23*(4), 507–513. https://doi.org/10.1080/20445911.2011.532207

Son, L. K., & Metcalfe, J. (2005). Judgments of learning : Evidence for a two-stage process. *Memory & Cognition*, *33*(6), 1116–1129.

Star, J. R., Pollack, C., Durkin, K., Rittle-Johnson, B., Lynch, K., Newton, K., & Gogolen, C. (2015). Learning from comparison in algebra. *Contemporary Educational Psychology*, *40*, 41–54.

Townsend, C. L., & Heit, E. (2011). Judgments of learning and improvement. *Memory & Cognition*, *39*(2), 204–216.

Zuniga, K. E., Mueller, M., Santana, A. R., & Kelemen, W. L. (2019). Acute aerobic exercise improves memory across intensity and fitness levels. *Memory*, *27*(5), 628–636.