



Metacognitive Monitoring in Test-taking Situations: A Cross-cultural Comparison of College Students

Mariana V. C. Coutinho

Assis. Prof., Zayed University, Abu Dhabi, United Arab Emirates,
mariana.coutinho@zu.ac.ae

Elena Papanastasiou

Assoc. Prof., University of Nicosia, Nicosia, Cyprus, papanastasiou.e@unic.ac.cy

Stylianou Agni

Asst. Prof., University of Nicosia, Nicosia, Cyprus, stylianou.a@unic.ac.cy

John M. Vasko

PhD student, University of Maryland, College Park, United States, jvasko@umd.edu

Justin J. Couchman

Assoc. Prof., Albright College, Reading, United States, jcouchman@albright.edu

In this study, we examined monitoring accuracy during in class-exams for Emirati, American and Cypriot college students. In experiment 1, 120 students made local confidence-ratings for each multiple-choice question in a psychology exam and also estimated their performance at the end of the exam. In experiment 2, to investigate the effect of practice in monitoring accuracy, 69 students made confidence judgments in two consecutive exams. In Experiment 3, to evaluate whether rating one's confidence aids exam performance, 172 students completed the exam either with or without providing confidence judgments. In all three cultures, confidence judgments were accurate indicators of performance, but there were significant cross-cultural differences that probably arose from regional educational practices. Additionally, global judgments were subjected to bias. Whereas American students overestimated their performance, Emiratis underestimated it. Practice alone with confidence judgments did little to improve monitoring accuracy. But there was a significant benefit for students who rated their confidence compared to those that did not. The results are discussed in light of the literature on metacognitive accuracy.

Keywords: metacognition, monitoring accuracy, local and global judgments of confidence, cross-cultural study, testing

Citation: Coutinho, M. V. C., Papanastasiou, E., Agni, S., Vasko, J. M., & Couchman, J. J. (2020). Metacognitive Monitoring in Test-taking Situations: A Cross-cultural Comparison of College Students. *International Journal of Instruction*, 13(1), 407-424. <https://doi.org/10.29333/iji.2020.13127a>

INTRODUCTION

Students typically strive to perform well on exams. Reaching this goal depends to some extent on how accurately they evaluate their learning while preparing for exams. Students who can discriminate well-learned from less well-learned material are likely to regulate their learning more efficiently by for example allocating more study time to items they have not yet mastered, which therefore increases their chances of earning a better grade on an exam. Although students' assessments of their cognitions are relatively accurate, they are far from perfect (Dunlosky & Lipko, 2007). Maki (1998) reviewed 25 studies conducted in her laboratory and found that on average the correlation between students' judgments and performance was .27. A theoretical account for this low correlation is provided by the cue-utilization approach that proposes that metacognitive judgments are inferential in nature (Schwartz, Benjamin, & Bjork, 1997; Koriat, 2008). When individuals predict the likelihood of remembering a particular item, they do not directly monitor the strength of the memory trace but instead base their judgments on cues, that is, on any variable believed to be associated with learning, knowing, or feelings of uncertainty.

Although there has been extensive research on monitoring accuracy—that is, how closely judgments about ones' performance are to actual performance (Dunlosky & Metcalfe 2009, Koriat, 1997; 2008, 2011, 2012), only a few of these studies examined the accuracy of metacognitive judgments in estimating performance in testing-taking situations, and the reported findings are mixed (e.g., Couchman, Miller, Zmuda, Feather, & Schwartzmeyer, 2016; Nietfeld, Cao, & Osborne, 2005). Furthermore, what is currently known in the literature about monitoring accuracy comes mainly from studies conducted on students from Western countries, where overconfidence bias is relatively common among students and culture is traditionally conceptualized as individualistic (Bol & Hacker, 2012; Dunlosky & Rawson, 2012; Foster, Was, Dunlosky, & Isaacson, 2016; Rawson & Dunlosky, 2007). Hence, to advance the theoretical understanding of metacognition, more research is needed in Non-Western countries where cultural values, social norms, and history differ from those of Western countries. To expand on the current literature, the present research examined monitoring accuracy during in-class exams among college students who were of a traditionally collectivist culture (the United Arab Emirates), an individualistic culture (the United States), and of a hybrid culture whose social norms include aspects from Middle Eastern and Western cultures (Cyprus). Also, the present research evaluated the efficacy of introducing metacognitive judgments to exams as an educational tool.

Over the past decade, there has been an increased interest in how accurate people's judgments of their own learning are (e.g., Dunlosky & Lipko, 2007; Dunlosky & Rawson, 2012; Koriat & Goldsmith, 1996; Maki, 1998). One reason for this is that metacognitive judgments are not epiphenomenal. They influence behavior. What is more, they can either hinder or improve performance. When these judgments are accurate indicators of how well students have mastered a subject, they promote learning by directing students' attention towards material not yet learned. However, when these

judgments are biased, they hinder performance by giving students a false sense of having mastered the material.

Researchers have investigated monitoring accuracy for local and global metacognitive judgments in a classroom setting (e.g., Couchman et al., 2016; Nietfeld et al., 2005) as well as in the laboratory (e.g., Dunlosky, Rawson, & Middleton, 2005). Local metacognitive judgments are those made for each test item separately, and global metacognitive judgments are an overall estimate of how well one performed on the entire test. When comparing local and global judgments in terms of monitoring accuracy, prior studies reported mixed findings regarding which of the two judgments are better indicators of performance. For instance, in a laboratory study, Dunlosky et al. (2005) reported that when students were asked to make local and global metacomprehension judgments of text material, monitoring accuracy was higher for local judgments than global judgments. Similarly, in a recent classroom study whereby undergraduate students were asked to rate their level of confidence for every question in an exam and upon completion of the exam make a global judgment of what grade they were likely to earn, local judgments yielded higher monitoring accuracy than global judgments (Couchman et al., 2016). However, in another classroom study with relatively similar methodology as Couchman and colleagues, global judgments were reported to be more accurate indicators of performance than local judgments (Nietfeld et al., 2005). The reason the results between these studies differed is unclear but may be related to the nature of global judgment. Couchman et al. asked students to estimate the grade they would receive on an exam whereas Nietfeld and colleagues assessed their overall confidence in their performance.

One possible explanation for reported differences in monitoring accuracy between local and global judgments may be the result of the differential influence of various experience-based and analytical-based cues on these judgments. A consensus in the literature exists that local confidence ratings are largely mediated by retrieval fluency, an online experience about the easiness by which information is recalled from memory (Chua, Hannula & Ranganath, 2012; Koriat, 2012; Nelson & Narens, 1990). For example, when participants are asked to monitor their confidence, high confidence ratings are associated with answers that took less time to retrieve, regardless of whether or not the answer was correct (Koriat, Ma'ayan & Nussinson, 2006; Robinson, Johnson & Herndon, 1997). However, because correct answers are typically retrieved faster, reliance on this cue does not reduce accuracy (Koriat & Ackerman, 2010). Conversely, global judgments are less likely to be mediated by retrieval fluency because they are an aggregate evaluation of performance. That is, they encompass all of the questions answered, are displaced in time, and may thus be more subject to memory biases.

Furthermore, local confidence ratings in contrast to global judgments may be less likely to be influenced by wishful thinking, self-serving and memory biases, and beliefs unrelated to the task at hand. One reason for this is that real-time metacognitive judgments are not perceived as a direct reflection of the self or of one's self-efficacy, the ability to perform a task.

In fact, the influence of beliefs about the self and attentional bias on global judgments has been demonstrated. Ehrlinger, Mitchum, and Dweck (2016) reported that individuals who hold fixed beliefs regarding intelligence are more likely to overestimate their performance than those with incremental views of intelligence. In addition, they showed that the link between fixed beliefs of intelligence and overconfidence is mediated by an attention allocation bias toward easy items. Students who view intelligence as static tend to devote more attention to easy items than difficult items while taking a test. Because of this bias, easy items are more salient which in turn influences students' predictions of performance. In summary, students display overconfidence in part because as they assess their performance, they recall more easy items than difficult ones, creating an illusion that they performed better than they actually did.

It is possible that the difference between the two is driven by reliance on these two cues differently, with local, confidence ratings being more strongly influenced by experience-based cues than beliefs, while the opposite may hold for global judgments that are highly reflective. Considering that the influence of experience-based cues is often not accessible to conscious thought and is quite pervasive (Castel, McCabe, & Roediger, 2007; Rhodes & Castel, 2008, 2009) and that belief-based cues can be changed (more vulnerable to change), it is expected that the influence of culture in metacognition would be more likely to be observed for global judgments than local confidence ratings. However, to date, no study has investigated differences in monitoring accuracy between local and global judgments across cultures. This is one of the main goals of the present research.

Culture can be defined in many ways, with a common view being that culture is "the set of values and beliefs people have about how the world (both nature and society) works as well as the norms of behavior derived from that set of values" (Gorodnichenko & Roland, 2012) and on that basis it can be classified as collectivist or individualistic. Differences in culture have shown to affect cognition (Nisbett, Peng, Choi, & Norenzayan, 2001; Norenzayan, Smith, Kim, & Nisbett, 2002). For instance, research on reasoning reported differences in strategy preferences among individuals from collectivist versus individualistic cultures (Norenzayan et al., 2002), where individuals from Eastern Asian countries such as Korea are more likely to solve problems with intuitive strategies when analytical and intuitive solutions conflict whereas Western students prefer analytical strategies. However, little is known about the influences of culture in metacognition, especially regarding whether and how it influences monitoring accuracy for local and global judgments. The present research aims to complement the current literature on metacognition by evaluating monitoring accuracy between local and global judgments during an in-class examination among college students of a collectivist culture (UAE), individualistic culture (US), and a hybrid culture (Cyprus). Study 1 compared monitoring accuracy between confidence ratings and global judgments of performance across cultures. It was hypothesized that cultural differences would be found in global judgments but not local judgments, with American students displaying a higher level of overconfidence bias than Emirati and Cypriot students. Study 2 explored the effect of practice on improving the accuracy of metacognitive judgments. Study 3

examined whether adding local and global metacognitive judgment questions to exams boost performance in the exam.

Study 1

Study 1 examined the accuracy of local confidence judgments and global post-exam judgments in predicting performance during an in-class exam. We were particularly interested in determining (1) whether differences in monitoring accuracy exist between local judgments and global judgments, (2) whether monitoring accuracy varies as a function of culture (UAE, US, and Cyprus), (3) whether metacognitive ability predicts exam performance.

METHOD

Subjects

A total of 120 undergraduate students volunteered to participate in this study of which 51 attended college in the UAE, 28 in Cyprus, and 41 in the US. Ten UAE students who did not rate their confidence or for whom a gamma correlation (see below) could not be calculated were eliminated. The UAE students were all female and UAE nationals with a mean age of 22.53 (SD = 1.43). In the Cypriot sample, 79.31% of the students were female, 62.07% were from Cyprus, and 37.93% were from Greece with a mean age of 22.86 (SD = 1.90). The US students were 85.36% female, approximately 72.19% were white/Caucasian, 11.08% were black/African American, 11.11% were Hispanic/Latino, and 5.62% were Asian with a mean age of 20.54 (SD = 1.55).

Design and Procedure

Participation was voluntary, and the research ethics committee from the institutions where the data was collected approved the study. Undergraduate students enrolled in introductory psychology courses in the UAE, Cyprus, and the US were administered their normally scheduled exam, which was in English, included only multiple-choice questions, and consisted of a Scantron form to record their to-be-graded responses (not analyzed). They were instructed to circle their choice (A, B, C, or D) for each question and rate their confidence in their answer from 1 (very likely incorrect) to 5 (very likely correct) immediately after answering. They were also asked after taking the exam to estimate their grade on the exam on a scale from 0 to 10.

Data Analysis

Confidence ratings were analyzed in addition to the overall correct proportion. All of the exams were scored such that, for each question, the subjects had a set of responses (ratings and outcomes). Goodman-Kruskal gamma correlations are often used in metacognition paradigms to measure the relationship between actual performance and confidence judgments made by subjects (Dunlosky & Metcalfe, 2009). Gamma is calculated by comparing the responses to every question (ratings and outcomes) to every other question to determine whether the same logic was followed (that is, whether pairs are concordant, discordant, or tied). Gamma ignores ties, and thus the final score, ranging from -1.0 to +1.0, indicates the trend in confidence ratings relative to the trend

in actual performance outcomes. It is interpreted similarly to a correlation except that a t-test can determine whether the relationship is significantly different from zero. The data were analyzed using SPSS software.

FINDINGS

Exam Difficulty

The difficulty of the exam for each group of students (Emirati, Cypriot, and US students) was computed by measuring the number of questions they answered correctly. The average test scores out of 10 points for the Emirati, Cypriot, and US students were 8.5 (SD = 1.0), 5.4 (SD = 1.1), and 6.6 (SD = 1.5), respectively. Although test difficulty was not deliberately manipulated, it varied across groups. The Cypriot students had the most difficult test, followed by the US and Emirati students.

Monitoring accuracy for local judgments

To examine monitoring accuracy for local judgments, Goodman-Kruskal gamma correlations between the confidence ratings and the responses to each exam question were assessed for the Emirati, Cypriot and American students. The gamma correlations were .55 (SE = 0.09), $t(40) = 6.3$, and $p < .001$; .17 (SE = 0.06), $t(27) = 2.2$, and $p < .05$; and .5 (SE = 0.04), $t(40) = 12.7$, and $p < .001$ for the UAE, Cyprus, and US students, respectively. Gammas were significantly different from zero, indicating that the students in all three countries accurately differentiated questions that they knew well from those that elicited uncertainty (see Figure 1). This suggests that metacognition is an etic phenomenon shared universally by humans (and some animals; see Couchman, Coutinho, Beran, & Smith, 2010).

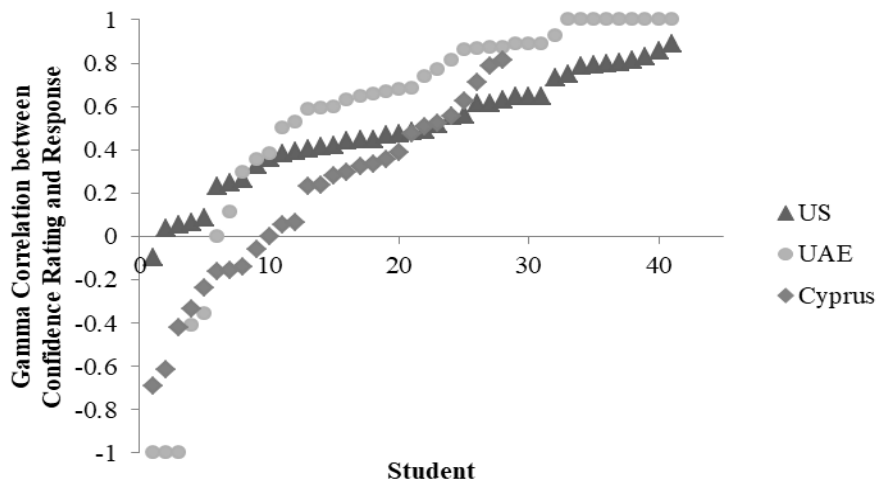


Figure 1
Gamma Correlation between Confidence Ratings and Response Per Student in the UAE, Cyprus and the USA, Ordered from Lowest to Highest Gamma for Each Group

To examine differences in monitoring accuracy based on culture, a one-way ANOVA was conducted with culture as the independent variable and monitoring accuracy measured by gamma correlations as the dependent variable. The analysis yielded a main effect of culture, $F(2,107) = 7.3$ and $p < .001$. Post hoc tests ($p < .05$) showed that this was due to lower gamma values for the Cyprus students compared to both the UAE and US students. This was likely because the Cypriot students had less experience with multiple-choice exams and their poor performance on the exam.

Monitoring accuracy for global judgments

Figure 2 shows the average correct proportion for the predictions and performance for the students in the US and UAE. Data from the students in Cyprus were not included in the analysis due to unavailability (having less experience with this format, many of the Cypriot students did not estimate their grades at the end of the exam). A significant difference was found between the grade estimation and the actual performance of the Emirati students, $t(40) = 4.5$ and $p < .003$, and for the US students, $t(40) = 5.2$ and $p < .001$. The Emirati students underestimated their performance on the exam by .83 (on a 10-point scale, $SE = .18$), while the US students overestimated their performance on the exam by .87 ($SE = .16$). Furthermore, the discrepancy between the predictions and the actual performance between students in the UAE and US was significantly different, $t(80) = 4.1$ and $p < .001$. While both cultures had accurate in-the-moment confidence ratings at the item level, they were inaccurate when asked to estimate their performance on the exam. Several studies have suggested that American students' educational overconfidence is due to generational increases in assertiveness, narcissistic traits, and high expectations (Tweng, 2009). It is notable that this effect was not found in the UAE students and was, in fact, significant in the opposite direction. This could be due to the language or cultural effects previously described.

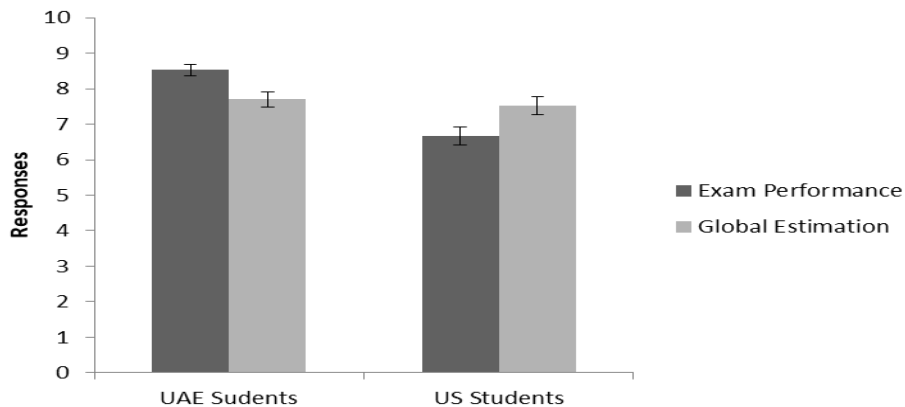


Figure 2
Post-exam Global Estimation of Performance Compared to Actual Performance for UAE and US Students, Converted to a 10-Point Scale. Error Bars Indicate Standard Error

Because of demographic differences between the US and Emirati sample, with the UAE sample having only females and the US sample having a few males in addition to females, an additional analysis assessing the differences between the grade predictions and the actual performance with only females was performed. The results did not change. The female US students overestimated their performance on the exam by .86 (SE = .17), and the difference between the predictions and the actual performance between them and the UAE students was significantly different as before, $t(74) = 6.4$ and $p < .001$.

Exam difficulty and monitoring accuracy for global judgments

To evaluate whether differences in global judgments originated from differences in test difficulty among the exams, two additional analyses were conducted in which the students were separated into groups based on their overall performance, and global estimations were separately compared to the performance of the participants in each group. Students who earned a grade of 7 or higher on the exam were placed in one group (higher performers) and those with a grade lower than seven were placed in a separate group (lower performers). We were interested in assessing whether overestimation bias was present across the high and low performers for the US students. The analysis revealed that this was indeed the case (see Figure 3).

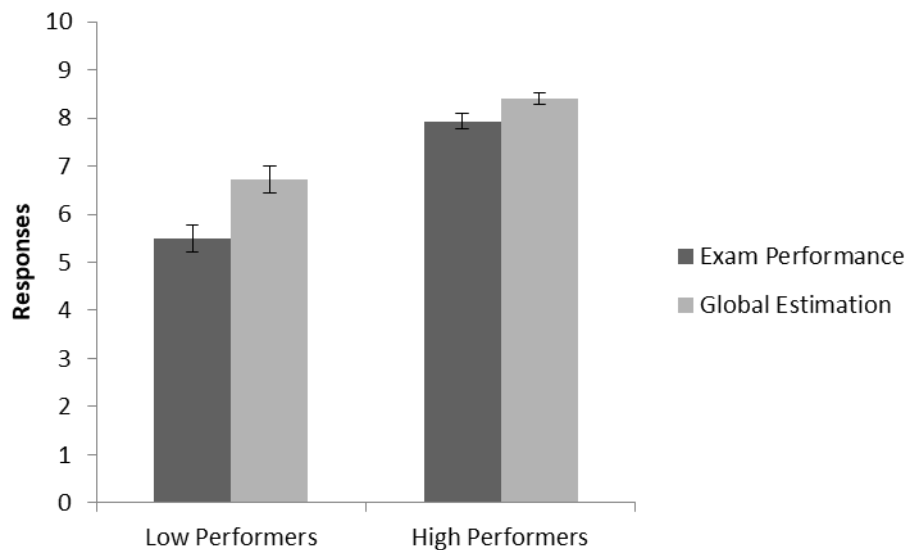


Figure 3

Post-exam Global Estimation of Performance Compared to Actual Performance for Low and High Performers in the US Sample

Grade estimation was significantly higher than actual performance for the high performers and low performers, $t(19) = 2.9$ and $p = .009$, and $t(17) = 4.4$ and $p < .001$, respectively.

Monitoring accuracy as a predictor of exam performance

This study also examined whether the students who were more accurate at metacognitive monitoring also performed better. The participants were separated into two groups based on their gamma correlations. Those with a gamma correlation above the median (.51) were characterized as high self-regulators and those with a gamma correlation below the median were characterized as low self-regulators. There were 55 participants in each group (which combined cultures). The mean gamma for the high self-regulators and low self-regulators was .77 and .10, respectively. The average performance (out of 10) for the two groups was 7.6 (SD = 1.68) and 6.5 (SD = 1.64), respectively. The high self-regulators performed better, $t(108) = 3.84$ and $p < .001$. Increased metacognitive ability thus predicted better overall performance.

This is notable because laboratory studies have shown that high monitoring accuracy in initial trials correlates with higher performance on later trials (Thiede, 1999), and major theories of self-regulated learning propose that students who monitor their mental processes more accurately will regulate their knowledge more efficiently, which will consequently improve performance (Hartwig, Was, Isaacson, & Dunlosky, 2012). The current study's results suggest that this link extends to real-world high-stakes settings.

Study 2

Study 2 evaluated whether the accuracy of the students' local confidence judgments improved with experience. Emirati and Cypriot students completed two exams at the designated time in the course using the same methods described in Study 1. The first exam was administered in the middle of the semester and the second exam at the end of the semester.

Method

Subjects

A total of 69 students participated; 41 were female UAE nationals with a mean age of 22.60 (SD = 1.46), 28 were from the Cyprus sample with 79.31% female, and 62.07% were from Cyprus (37.93% were from Greece) with a mean age of 22.86 (SD = 1.90). Nine students were excluded using the criteria described in Study 1. This study included only participants from the UAE and Cyprus because data from American students were unavailable due to curriculum reasons. Metacognitive questions could not be added to the final exam.

Design and procedure

It is similar to Study 1 with the exception that students completed two exams.

Data analysis

The data were analyzed similarly to Study 1. A small number of missing data points (4) were replaced with the average value for the group. This procedure preserved equal sample sizes per rating and did not change the general pattern of the results.

Results

Goodman-Kruskal gamma correlations between the confidence ratings and the responses of students for exam 1 and exam 2 for the UAE students were .54 (SE = .10), $t(33) = 5.3$, and $p < .001$, and .5 (SE = .07), $t(33) = 7.0$, and $p < .001$, respectively, and for the Cyprus students were .19 (SE = .08), $t(25) = 2.5$, and $p < .05$, and .13 (SE = .06), $t(25) = 2.2$, and $p < .05$, respectively. All of the gammas were significantly different from zero, showing that the students were quite good at rating their performance concurrently in real time. No improvement in monitoring accuracy was observed. The gamma correlations did not significantly differ from exam 1 to exam 2 for the students in the UAE, $t(33) = .203$ and $p = .84$, or Cyprus, $t(25) = .582$ and $p = .57$.

The average grade for exam 1 and exam 2 was 8.5 (SD = 1.0) and 8.2 (SD = 1.0) for the students in the UAE and 5.4 (SD = 1.1) and 5.7 (SD = 0.8) for the students in Cyprus. Similar to the gamma values, exam performance did not significantly change with practice for either the UAE, $t(33) = 1.6$ and $p = 0.115$, or Cyprus students, $t(25) = 1.11$ and $p = 0.275$.

These results suggest that simply using this exam system does not directly improve metacognition any more than simply taking a test improves test performance. However, because Study 1 showed such a strong link between metacognition and performance, we tested whether using this metacognitive monitoring system improved performance compared to a control test with no confidence ratings.

Study 3

Is there any performance gain associated with confidence ratings? We hypothesized that there would be because ratings serve as a memory cue that can prompt revisions to low-confidence questions. Also, confidence ratings can generally increase awareness over one's thoughts, which leads to better decision-making. These have traditionally been difficult hypotheses to test because performance differences are often small and individual classes can be highly variable in terms of student performance. Combining multiple cultures and courses into a single group reduced errors due to individual differences and course/instructor/college/culture effects. In this study, Emirati and American students either completed an exam with confidence ratings or without.

Method

Subjects

A total of 172 students participated in this study. Eighty were female UAE nationals with a mean age of 21.90 (SD = 1.30). Ninety-two US students were approximately 59.78% female, 85.87% white/Caucasian, 3.26% black/African American, 6.52% Hispanic/Latino, 1.09% Asian, and 3.26% other with a mean age of 19.35 (SD = 1.17). Half of the participants in each group (half of the UAE and US students) were assigned to the no-confidence rating condition and the other half to the confidence rating condition. Cypriot students were not included in this study because the data was unavailable due to methodological constraints. This study requires the administration of the same test to a large sample of students (40 or more) because of its two-condition

design. Unfortunately, a class with such sample size was not available at the time of testing.

Design and procedure

The procedures during the exam were identical to Study 2, except that one group did not provide confidence ratings.

Data analysis

The data were analyzed similarly to Study 2.

Results

The average grade earned by the students who had taken the exam with the confidence ratings was 8.04 (SD = 1.1) and without the confidence ratings was 7.65 (SD = 1.4). Students who kept track of their confidence performed significantly better than those who did not, $t(170) = 2.06$, $p < .05$, Cohen's $d = .31$.

DISCUSSION

The three studies herein investigated the relationships between metacognitive monitoring judgments and performance during in-class exams among students in the UAE, the US and Cyprus. Study 1 focused on the accuracy of local confidence ratings and global judgments and whether monitoring accuracy predicted exam performance. Students across all three cultures were capable of discriminating the questions they knew well from those that they were less certain, but Emirati and American students were significantly better than the Cypriot students at this task. Cypriot students' low but significant gamma correlation was likely due to their lack of familiarity with multiple-choice exams. Performing a task that was somewhat new most likely increased the demands of the task to a degree that fewer resources were available for monitoring their knowledge, which in turn could have led to a reduction in the judgment accuracy. It should be noted that the gamma correlations between the confidence judgments and the response among the US and Emirati students were higher than previously reported in other research studies (Dunlosky & Lipko, 2007; Maki, 1998). This could have been a result of the increase in motivation on the part of the sample since their performance counted toward their final grade in the course.

As predicted, differences in the global judgments were found between Emirati and American students. Emirati students underestimated their performance, whereas the American students overestimated it. The finding that American students displayed overconfidence is not new. This was in line with prior research conducted in the classroom (Couchman et al., 2016; Foster, Was, Dunlosky, & Isaacson, 2016) and in the laboratory (Dunlosky & Rawson, 2012) and was observed in domains other than academic performance such as safe driving (Marttoli & Richardson, 1998), sports (Middleton, Harris, & Surmom, 1996), occupational abilities (Haun, Zeringue, Leach, & Foley, 2000), and social skills (Swann & Gill, 1997). However, the reason for why Emirati students did not show similar optimistic bias toward their performance is less clear. Perhaps, Emirati students are less susceptible to the influence of factors that

contribute to overconfidence such as wishful thinking and fixed beliefs of intelligence (Ehrlinger et al., 2016; Foster et al., 2016).

Wishful thinking in this context refers to when students base their predictions on how they wish to have performed on the exam and not on how well they actually believe they did. This type of bias is associated with the maintenance of positive feelings of self-worth. Although individuals from collectivist cultures also have the need to maintain self-esteem, they engage in self-serving bias to a lesser extent than those from individualistic cultures (Chang & Asakawa 2003; Heine, Takata, & Lehman. 2000; Heine & Hamamura, 2007, Norasakkunkit & Kalick, 2002). Moreover, regarding beliefs of intelligence, prior studies have found that they are related to one's culture, with individuals from collectivist cultures more likely to perceive intelligence as malleable—it can change over time as the result of one's effort (Rattan, Savani, Naidu, & Dweck, 2012; Sternberg, 2004; Stevenson & Stigler, 1994). Such beliefs are not associated with an attentional bias toward easy items, which has been shown to support overconfidence in Western students (Heine et al., 2001).

The reasons listed above explain why the Emirati students did not overestimate their performance, but why did they display underconfidence? It may have something to do with the demands of learning and taking tests in English, their second language. Or it may be related to conformity and humility, which are highly valued in their culture. Conformity may encourage individuals to consider additional information regarding members of their group when predicting their performance, such as how their peers performed on the exam. Then this information may be weighted in their grade estimations. Furthermore, not wanting to stand out from the group may have led the Emirati students to reduce their estimations to a level that was close to an average grade.

Interestingly, when exam performance was compared between students with low and high monitoring accuracy, confidence rating accuracy predicted performance. The students who scored higher in monitoring accuracy performed better on the exam than those who scored lower. This finding is in line with Hartwig et al. (2012), who showed that the accuracy of confidence ratings in one task correlates with performance on another unrelated task such as an exam. Why is this the case? Students who score higher in monitoring accuracy are more likely to effectively regulate their behavior during learning and tests, which in turn improves their performance. When students are aware of what they know and do not know, they can take better control of their learning, by for example allocating additional study time to items that they have not learned well or reviewing exam questions that they were uncertain about the correct answer. However, students who have less insight into their knowledge will not know how to effectively allocate their study time and may quit studying earlier than needed or not revise questions that elicited doubt.

In line with the results of Study 1, prior research has demonstrated that metacognitive judgments are not epiphenomenal. They influence the regulation of behavior and performance (Couchman et al., 2016; Dunlosky & Rawson, 2012; Hartwig et al., 2012; Koriat, 2007; Koriat & Goldsmith 1996). For example, Dunlosky and Rawson (2012) had students study the definition of concepts, rate the likelihood that they would

correctly remember the definition, and subsequently tested their knowledge using a cue-recall test. The results were clear and straightforward: the accuracy of the students' judgments correlated with study strategies and performance. Those who were overconfident spent significantly less time studying the concepts to be remembered and performed significantly worse than those who displayed little overconfidence.

Considering the role that metacognitive monitoring plays in learning and performance, it is important to identify factors that have the potential to enhance monitoring accuracy. Study 2 investigated whether practice with rating one's confidence during exams could boost monitoring accuracy. Coutinho et al. (2015) showed that experience with monitoring one's uncertainty in a perceptual task reduces the cognitive load. That is, the more experience people have with monitoring their level of certainty in a task, the less they rely on working memory resources to assess whether or not they know the answer to a question. Since experience reduces working memory dependency, we reasoned that it could also enhance monitoring accuracy. However, the results did not support this hypothesis. There were no changes in how thoroughly the participants discriminated well-known from not well-known exam questions for both Emirati and Cypriot students. One explanation is that the amount of practice the students had was very limited. A recent study showed that overestimation of performance prevails even after students completed 10 exams (Foster et al., 2016). What seems to be missing from the present study and that of Foster et al. (2016) is feedback on the accuracy of the students' judgments. Prior research showed that when participants receive feedback during testing, monitoring accuracy increases (Perfect, Hullins, & Hunt, 2000). Future research might focus on this kind of feedback to improve metacognition, and the current study suggests that this should reach across cultures.

Study 3 compared the students' performance on an exam with and without the inclusion of local confidence judgments. We were interested in evaluating whether monitoring one's confidence during test-taking situation improve exam performance. The findings showed that it does. Students who rated their confidence earned slightly higher grades than those who did not. Why did this metacognitive component boost the students' performance? First, it prompted the students to engage in an analysis of knowledge that was more refined than what they typically do on their own. They evaluated their knowledge on a 5-degree criterion scale that was recorded right next to their response and could always be seen. This forced the implicit feeling of uncertainty to be expressed explicitly, which could have led to a more accurate assessment of the students' knowledge and improved self-regulation. Also, because each rating was recorded on the paper and available for them to see at any point during the exam, it could have served as a memory cue for which questions should (or should not) be reviewed again. On normal exams, each new feeling of uncertainty retroactively interfered with previous feelings. With past metacognitive ratings easily available, retroactive interference was less likely to prevent the students from forgetting the questions that caused the most uncertainty.

In general, the present findings demonstrated that local confidence judgments are more accurate estimators of performance than global judgments (Study 1), its inclusion in an exam is associated with an increase in exam grade (Study 3) but its accuracy is not

improved with practice (Study 2). However, as with any cross-cultural research, there were some limitations inherent to the data collection. The students attended separate colleges in three different countries, with different professors, course content, exams, and curricula. Although the metacognitive procedures were maintained as identical as possible, each exam analyzed did not contain the exact, same number of questions and varied in terms of subjective difficulty. Second, due to differences in culture or understanding instructions, Cypriot students did not make global judgments. Future research could help reduce this by first providing students with a training paradigm, although some non-compliance was probably inevitable. Third, the demographics of the sample regarding gender were not balanced across the countries. The Emirati sample consisted of only females whereas although the samples from the US and Cyprus were mixed, they did have more females than males. Overall, to the best of our knowledge, this study is the largest cross-cultural exploration of monitoring accuracy during in class exams.

The results of this study raise several questions. The first is how culture contributes to the development of metacognitive skills. Prior studies within the domain of reasoning found cultural differences in the usage of two types of reasoning strategies, with Koreans being more likely to choose intuitive strategies versus analytic ones and Americans displaying the opposite pattern (Norenzayan et al., 2002). Although none of these studies were conducted in the UAE, we expect similar results because like Korea, UAE's culture is collectivistic. Similarly, do differences in how intuitive and analytic processes mediate metacognitive judgments vary between individualistic and collectivist cultures? We predict that individuals from a culture that places strong weight on intuition may be more likely to base their judgments on how they feel and not on the analysis of explicit knowledge retrieved from long-term memory.

A second question concerns how to cultivate and enhance metacognition skills in the classroom. The present study showed the benefits associated with adding metacognitive judgments to an exam. We believe this procedure can be easily adopted by any faculty given its simplicity. In addition to having students rate their confidence, it is recommended that feedback on the accuracy of the students' judgments be provided to them. Rawson and Dunlosky (2007) observed an increase in performance when students were presented with the correct response after predicting their performance. Educating students about psychological bias and fallacies (for example, the first-instinct fallacy and heuristic availability, among others) also has the potential to improve metacognitive skills. Knowing that ease of access (or easy processing or familiarity) is not always indicative of the right answer could help students identify situations in which they might be making a mistake. Awareness of these cognitive phenomena could help build metacognitive skills and thus improve self-regulation and memory performance. Similarly, researchers should investigate whether test structure, for example, having particularly difficult questions in the beginning, middle, or end, might bias global judgments or change performance.

As previously noted, the current study's paradigm can be easily implemented in classrooms and is not limited to multiple-choice exams. It can be adapted to almost any

format. Importantly, for classroom application and future research, it can be utilized as a diagnostic tool for metacognitive skills. Teachers could add confidence rating questions to practice tests or other formative types of assessments to measure the accuracy of students' judgments. Doing so will help educational professionals identify students' needs and weaknesses. This could be used to foster metacognitive skills in individuals with intellectual or linguistic deficits. Students from less advantageous backgrounds, such as those living in poverty, typically suffer a decrease in IQ as they reach adolescence that is triggered by multiple environmental factors, including dropping out of school or a weak educational system (Nisbett, 2008). Improving metacognitive skills in this population could help them overcome some of these drawbacks.

Metacognition is a universal psychological phenomenon that plays an important role in school performance but is not limited to it. It affects many aspects of life, including decision-making and even general well-being (Hayes, 1987; Wells, 1995). Similar to any other skill, metacognition can be practiced and improved. We believe that the methods reported herein when applied to tasks familiar to students could be used to boost monitoring accuracy and performance across all backgrounds and cultures.

REFERENCES

- Alsheikh, N. (2014). The perceived and actual use of metacognitive reading strategies by the UAE high school students. *J. of ELT and Applied Linguistics*, 2(1), 140-153.
- Bol, L., & Hacker, D. J. (2012). Calibration research: where do we go from here? *Frontiers in Psychology*, 3, 229. doi:10.3389/fpsyg.2012.00229.
- Castel, A. D., McCabe, D. P., & Roediger, H. L. III, (2007). Illusions of competence and overestimation of associative memory for identical items: Evidence from judgments of learning. *Psychonomic Bulletin & Review*, 14, 107-111.
- Chang, E. C., & Asakawa, K. (2003). Cultural variations on optimistic and pessimistic bias for self versus a sibling: Is there evidence for self-enhancement in the West and for self-criticism in the East when the referent group is specified? *Journal of Personality and Social Psychology*, 84, 569-581.
- Chua, E. F., Hannula, D. E., & Ranganath, C. (2012). Distinguishing highly confident accurate and inaccurate memory: insights about relevant and irrelevant influences on memory confidence. *Memory*, 20, 48-62.
- Couchman, J. J., Coutinho, M. V. C., Beran, M. J., & Smith, J. D. (2010). Beyond stimulus cues and reinforcement history: A new approach to animal metacognition. *Journal of Comparative Psychology*, 124, 356-368.
- Couchman, J. J., Miller, N., Zmuda, S. J., Feather, K., & Schwartzmeyer, T. (2016). The instinct fallacy: The metacognition of answering and revising during college exams. *Metacognition & Learning*, 11(2), 171-185.
- Coutinho, M. V., Redford, J. S., Church, B. A., Zakrzewski, A. C., Couchman, J. J., & Smith, J. D. (2015). The interplay between uncertainty monitoring and working memory: Can metacognition become automatic? *Memory and Cognition*, 43(7), 990-1006.

- Dinsmore, D. L., & Parkinson, M. M. (2013). What are confidence judgments made of? Students' explanations for their confidence ratings and what that means for calibration. *Learning and Instruction, 24*, 4–14.
- Dunlosky, J., & Lipko, A. (2007). Metacomprehension: A brief history and how to improve its accuracy. *Current Directions in Psychological Science, 16*, 228-232.
- Dunlosky, J., & Metcalfe, J. (2009). *Metacognition*. Thousand Oaks, CA: SAGE.
- Dunlosky, J., Mueller, M., & Tauber, S. K. (2014). The contribution of processing fluency (and beliefs) to people's judgments of learning. In D. S. Lindsay, C. M. Kelley, A. P. Yonelinas, & H. L. Roediger III (Eds.), *Remembering: Attributions, processes, and control in human memory: Papers in honor of Larry L. Jacoby* (pp. 46–64). New York: Psychology Press.
- Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction, 22*(4), 271-280.
- Dunlosky, J., Rawson, K. A., & Middleton, E. L. (2005). What constrains the accuracy of metacomprehension judgments? Testing the transfer-appropriate-monitoring and accessibility hypotheses. *Journal of Memory and Language, 52*(4), 551-565.
- Ehrlinger, J., Mitchum, A. L., & Dweck, C. S. (2016). Understanding overconfidence: Theories of intelligence, preferential attention, and distorted self-assessment. *Journal of Experimental Social Psychology, 63*, 94-100.
- Foster, N. L., Was, C. A., Dunlosky, J., & Isaacson, R. M. (2016). Even after thirteen class exams, students are still overconfident: The role of memory for past exam performance in student predictions. *Metacognition and Learning, 12*, 1-19.
- Gorodnichenko, Y., & Roland, G. (2012). Understanding the individualism-collectivism cleavage and its effects: Lessons from cultural psychology. In M. Aoki, T. Kuran, & G. Roland (Eds.), *Institutions and comparative economic development* (pp. 213-236). UK: Palgrave Macmillan.
- Hartwig, M. K., Was, C. A., Isaacson, R. M., & Dunlosky, J. (2012). General knowledge monitoring as a predictor of in-class exam performance. *The British Journal of Educational Psychology, 82*, 456-468.
- Haun, D. E., Zeringue, A., Leach, A., & Foley, A. (2000). Assessing the competence of specimen processing personnel. *Laboratory Medicine, 31*, 633–637.
- Hayes, S. C. (1987). A contextual approach to therapeutic change. In N. Jacobson (Ed.), *Psychotherapists in clinical practice*. New York: Guilford Press.
- Heine, S. J., & Hamamura, T. (2007). In search of East Asian self-enhancement. *Personality and Social Psychology Review, 11*, 4–27.
- Heine, S. J., Kitayama, S., Lehman, D. R., Takata, T., Ide, E., Leung, C., & Matsumoto, H. (2001). Divergent motivational consequences of success and failure in Japan and North America. *Journal of Personality and Social Psychology, 81*, 599–615.

- Heine, S. J., Takata, T., & Lehman, D. R. (2000). Beyond self-presentation: Evidence for Japanese self-criticism. *Personality and Social Psychology Bulletin*, *26*, 71-78.
- Koriat, A. (2008). Subjective confidence in one's answers: The consensually principle. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *34*, 945-59.
- Koriat, A. (2012). The self-consistency model of subjective confidence. *Psychological Review*, *119*, 80-113.
- Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, *103*, 490-517. doi: 10.1037/0033-295X.103.3.490.
- Koriat, A., Ma'ayan, H., & Nussinson, R. (2006). The intricate relationships between monitoring and control in metacognition: Lessons for the cause-and-effect relation between subjective experience and behavior. *J. of Exper. Psyc.: General*, *135*, 36-69.
- Koriat, A., & Shitzer-Reichert, R. (2002). Metacognitive judgments and their accuracy: Insights from the processes underlying judgments of learning in children. In P. Chambres, M. Izaute, & P-J. Marescaux (Eds.), *Metacognition: Process, function, and use* (pp. 1- 17). New York, NY: Kluwer.
- Maki, R. H. (1998). Test predictions over text material. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds), *Metacognition in educational theory and practice* (pp. 117-144). Mahwah, NJ: Erlbaum.
- Marttoli, R. A., & Richardson, E. D. (1998). Confidence in, and self-rating of, driving ability among older drivers. *Accident Analysis and Prevention*, *30*(3), 331-336.
- Middleton, W., Harris, P., & Surmom, M. (1996). Give 'am enough rope: Perception of health and safety risks in bungee jumpers. *J. of Social and Clinical Psyc.*, *15*(1), 68-79.
- Metcalfe, J., & Finn, B. (2008). Familiarity and retrieval processes in delayed judgments of learning. *J. of Exper. Psyc.: Learn., Memo. and Cognition*, *34*, 1084-1097.
- 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*, 267-270.
- Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 125-173). New York: Academic Press.
- Nietfeld, J. L., Cao, L., & Osborne, J.W. (2005). Metacognitive monitoring accuracy and student performance in the postsecondary classroom. *The Journal of Experimental Education*, *74*(1), 7-28.
- Nisbett, R. (2008). *Intelligence and how to get it: Why schools and cultures count*. New York, NY: Norton.
- Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic vs. analytic cognition. *Psychological Review*, *108*, 291-310.

- Norenzayan, A., Smith, E. E., Kim, B. J., & Nisbett, R. E. (2002). Cultural preferences for formal versus intuitive reasoning. *Cognitive Science*, 26, 653-684.
- Papleontiou-Louca, E. (2003). The concept and instruction of metacognition. *Teacher Development*, 7(1), 9-30.
- Perfect, T. J., Hollins, T. S., & Hunt, A. L. (2000). Practice and feedback effects on the confidence-accuracy relation in eyewitness memory. *Memory*, 8, 235-244.
- Rattan, A., Savani, K., Naidu, N. V. R., & Dweck, C. S. (2012). Can everyone become highly intelligent? Cultural differences in and societal consequences of beliefs about the universal potential for intelligence. *J. of Personality and Social Psych.*, 103, 787-803.
- Rawson, K., & Dunlosky, J. (2007). Improving students' self-evaluation of learning for key concepts in textbook materials. *European J. of Cognitive Psychology*, 19, 559-579.
- Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology. General*, 137, 615-625. <http://dx.doi.org/10.1037/a0013684>.
- Rhodes, M. G., & Castel, A. D. (2009). Metacognitive illusions for auditory information: Effects on monitoring and control. *Psychonomic Bul. & Rev.*, 16, 550-554.
- Robinson, M. D., Johnson, J. T., & Herndon, F. (1997). Reaction time and assessments of cognitive effort as predictors of eyewitness memory accuracy and confidence. *Journal of Applied Psychology*, 82, 416 - 25.
- Schwartz, B. L., Benjamin, A. S., & Bjork, R. A. (1997). The inferential and experiential bases of metamemory. *Current Directions in Psyc. Sciences*, 6, 132-137.
- Sternberg, R. J. (2004). Culture and intelligence. *American Psychologist*, 59, 325-338.
- Stevenson, H., & Stigler, J. W. (1994). *The learning gap: Why our schools are failing and what we can learn from Japanese and Chinese education*. New York, NY: Simon & Schuster.
- Swann, W. B., Jr., & Gill, M. J. (1997). Confidence and accuracy in person perception: Do we know what we think we know about our relationship partners? *Journal of Personality and Social Psychology*, 73, 747-757.
- Thiede, K. W. (1999). The importance of monitoring and self-regulation during multitribal learning. *Psychonomic Bulletin and Review*, 6, 662-667.
- Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *J. of Edu. Psychology*, 95, 66-73.
- Tweng, J. M. (2009). Generational changes and their impact in the classroom: teaching Generation Me. *Medical Education*, 43, 398-405.
- Wells, A. (1995). Meta-cognition and worry: a cognitive model of generalized anxiety disorder. *Behavioral and Cognitive Psychotherapy*, 23, 301-320.