Kasetsart Journal of Social Sciences 44 (2023) 267–272

Kasetsart Journal of Social Sciences

journal homepage: http://kjss.kasetsart.org

Investigation of Thai students' metacognitive monitoring in biochemistry

Witawas Handee

Department of Chemistry, Faculty of Science, Silpakorn University, Mueang, Nakhon Pathom 73000, Thailand

Article Info

Article history:
Received 9 February 2021
Revised 20 April 2022
Accepted 21 April 2022
Available online 15 March 2023

Keywords:
domain-specific judgment,
judgement of learning,
metacognitive monitoring

Abstract

This research aimed to study metacognitive monitoring centered on judgment of learning in Thai undergraduates taking a biochemistry course. The participants' performances were evaluated based on their predicting ability on the topic recall and content prediction after being presented with visual image clues in biochemistry. The relative accuracy measured by gamma, G, and diagnostic accuracy analyzed by confusion matrices were conducted to explore patterns of metacognitive monitoring and examine the relationship with academic achievement. The outcomes from relative and diagnostic accuracy revealed that most students had overconfidence in both topic recall and content prediction, but the patterns varied from task to task suggesting that students' metacognitive monitoring was more likely a domain-specific judgment. Finally, the accuracy of students' prediction was correlated with exam scores at r_s =.624, p < .001, indicating a positive relationship between metacognitive monitoring and learning outcomes. The findings of this study could potentially establish future metacognitive prompt tools suitable for Thai students.

Introduction

Metacognition can be defined as the ability to monitor and control one's thoughts, commonly known as 'thinking about thinking' (Livingstone, 2003). Many studies have revealed that metacognition involves three sub-components: metacognitive knowledge, monitoring, and control (Dunlosky & Metcalfe, 2009). Metacognitive monitoring (Monitering, refer to as metacognitive accuracy, sensitivity, or performance (Siedlecka et al., 2016), is described as the ability to evaluate the current cognitive stage of an activity. Most of the time, it is measured by asking participants to judge how well they learn (Dunlosky & Metcalfe, 2009) leading to two types of metacognitive judgments: prospective and retroprospective judgments (Siedlecka et al., 2016).

In a classroom setting, prospective prediction describes the student's ability to predict their performance on future tasks based on previous knowledge (e.g., gauging likely success in completing upcoming homework after finishing a class session); this process is also known as judgement of learning (JOLs). Retro-prospective prediction describes the student's evaluation of already accomplished tasks (e.g. estimating the accuracy of completed homework), a process known as feeling of knowing (FOKs) (Kelemen et al., 2000).

E-mail address: handee_w@silpakorn.edu.

https://doi.org/10.34044/j.kjss.2023.44.1.30
2452–3151/© 2023 Kasetsart University.

In this study, we focused on JOL since it is one of the crucial skills for learners' success in academic settings. Lacking this skill could potentially downgrade students' performance since they might think they could do well in the test, but their actual performances are, in fact, lower.

It has been shown that contexts could play a role manipulating metacognition. For example, the metacognitive monitoring measured by memory confidence tests is shown to fluctuate across tasks or individual repetition (Kelemen et al., 2000). Other factors such as question difficulty or participant's interest and intelligence also have demonstrated to be predictors for metacognitive accuracy (Scott & Berman, 2013). These findings suggest that metacognitive monitoring is more likely a domain-specific process different than other metacognitive components such as metacognitive knowledge, which is more general to the context (Scott & Berman, 2013). However, it is still under debate whether metacognitive accuracy is domain-general or domain-specific to the environmental contexts, since there are numerous findings supporting both sides.

In addition to context, it has been revealed that learning culture has a substantial impact on metacognitive monitoring performance. This idea has arisen from the fact that human learning process relies heavily on learning cultures such as types of communication between teachers and pupils or surrounding social pressures. Therefore, cultural selection could have a great impact on how student think, feel, and perform a task (Heyes et al., 2020). A recent finding displayed that people who grew up in different countries performed differently on metacognition (Coutinho et al., 2020).

There are only a few metacognition studies specific to Thailand's culture and learning environment. This research, therefore, aimed to; (1) explore patterns of metacognitive monitoring in Thai students and (2) inspect the correlation of metacognitive monitoring to learning outcomes. There are multiple ways to measure metacognitive monitoring in various settings. In this study, Thai undergraduate students performed JOLs in two tasks: topic recall and content prediction based on visual image clues from materials that had been covered in class. The differences between confidence and performance–referred to as calibration (Fischhoff et al., 1977) was measured based on the signal detection theory (Fleming & Lau, 2014) via two types of measurement including diagnostic accuracy (by confusion matrix) and relative accuracy (by Goodman and Kruskal' gamma) (Lingel et al., 2019). Later, the calibration data were categorized based on students' grades to illustrate the relationship between performances and learning achievement.

Methodology

Participants

Participants in this study were purposively sampled from students enrolled in Biochemistry I of the first semester of 2017, Faculty of Science, Silpakorn University. Of 260 total enrollees, 255 agreed to participate; 16.1 percent were male, 83.8 percent female. Participants' majors included chemistry (31.2%), environmental science (33.1%), biology (26.9%), and microbiology (8.8%).

Data Collection

An online survey was used as the research instrument. Three stimulus items (illustrations selected from previous textbook exercises) were presented to participants, one at a time: an RNA molecular model, a line graph depicting the energy level of an enzymatic biochemical reaction, and the glyoxylate cycle (a side-road pathway of the citric acid cycle).

There were two sections in the questionnaire. In the first section, students had to predict their ability or rate their future performance for the upcoming tests in the second section. The survey was constructed so that students were unable to turn back to change their ratings from previous sections. The accuracy of figures and questions in the tests were approved by three independent teachers in the field with an IOC of 1.00.

In the first section, for each item, participants had to predict two tasks: whether they recognized the illustration from which topics were taken (yes-no questions) and whether they could tell which contents or types of problems associated with the figure (confidence rating scale from 1–5). The answers from the first section were counted as predicting scores. Later in the second section, they had to identify the names of the chapters containing the figures (matching questions) and answer open-ended questions related to each illustration. Their answers in the second section were graded as correct or incorrect and served actual performance scores. The internal consistency between the two questions was highly correlated ($r = .858$, $p < .001$) indicating high consistency in the individual answers.

Data Analysis

To answer the first research question if students can accurately recall and predict associated contents from three biochemistry visual images, the descriptive statistics were applied to the continuous data including rating scales and answer scores. Subsequently, Goodman and Kruskal' gamma (G) was conducted between confident and actual performance scores during the content prediction task to measure relative accuracy. Since the data from the recall task were obtained as binary data, G correlation was ignored.

To gain an in-depth view of students' calibration, confusion matrix analysis was applied. A confusion matrix (or a 2 × 2 contingency table) is a tool for measuring the diagnostic accuracy of metacognitive monitoring (Lingel et al., 2019). Both performance and judgment are separated into either correct or incorrect and filled in a 2 × 2 table. Therefore, four possible outcomes were obtained: hit (true positive, TP), miss (false positive, FP), false alarm (false negative, FN), and correct rejection (true negative, TN). Subsequently, two key confusion matrix parameters were calculated as follows: sensitivity = TP/(TP+FN) and specificity = TN/(TN+FP).

After classifying students' calibration in the matrices, Matthew's correlation coefficient (MCC) and Receiver operating characteristic (ROC) analysis were analyzed to find whether students' predictions were better than a random chance. The MCC is a common measurement of binary classifications comparable to Pearson's correlation. The MCC of 1 or -1 indicates perfect agreement, whereas an MCC closer to 0 refers to a prediction no better than random. Based on the adaptation by (Fleming & Lau, 2014), the ROC curves were generated by plotting hit and false alarm for calculating Area Under ROC or AUROC. The AUROC has the possible value of 0–1 by which a value closer to 1 indicates high accuracy with high sensitivity and a value lower than 0.5 suggests a random prediction.

The final part of the analysis aimed to answer the second research question if there was a relationship between metacognitive monitoring and individual who received a grade. To achieve this, a correlation between overall prediction scores and total exam scores was performed. Subsequently, the average of individual prediction was categorized based on their grades and tested by a non-parametric one-way ANOVA, Kruskal-Wallis, H test since the data had failed the assumption test.

Results

Students' Performances on the Topic Recall and Content Prediction

Table 1 presents the contrast between students' prediction (confidence levels) and actual performances. The overall students' confidence levels of each item (figure) from both questions were slightly positive (approximately 3.0–3.3 out of 5) except for the last item, where the students were rather confident (above 3.5 out of 5). However, the percent of overall students who made their accurate prediction was slightly moderate levels in some items (around 50–80%) but was significantly low in some items (approximately 8%). This discrepancy demonstrated the issue of overconfidence occurring unevenly among the items, suggesting that metacognitive monitoring was likely to be specific to question types and items.

To acquire more details about diagnostic accuracy, students' predictions and performances were classified into the confusion matrices and the outcomes of each parameter are shown in Table 2.

In the topic recall, the relative accuracy was disregarded because it was unable to calculate G from binary data. However, diagnostic accuracy was analyzed. For the first item, participants demonstrated moderate confidence and accuracy (Table 1, item 1). This is in agreement with the outcomes from matrix parameters which demonstrate decent sensitivity, specificity, and AUROC. However, the MCC = .33 was low suggesting that overall students' predictions were only slightly better than a random guess.

For the second item, the accuracy outcome shows very low accuracy, which is also consistent with all parameters from the matrix. Therefore, it was a clear example of overconfidence: students displayed very high confidence about what they thought they knew, but in fact they did not know. Based on the open-ended answers, students were confused between the reaction energy diagram and the enzyme kinetics graphs. These two graphs were both crucial diagrams for learning enzymatic characteristics. This indicated that students understood this topic superficially.

Table 1 Descriptive statistics of confidence and accuracy levels from each item

Question	Item 1	Item 2	Item 3
1. Can you identify the related topic of this figure?			
- Average confidence level ± *SD* (out of 5)	3.12 ± 1.10	3.00 ± 1.05	3.53 ± 1.07
- Percent of correct recall	62.0%	7.8%	79.2%
2. Can you tell which contents or types of problems are associated with this figure?			
- Average confidence level ± *SD* (out of 5)	3.24 ± 1.22	3.29 ± 1.16	3.75 ± 1.07
- Percent of correct prediction	36.2%	57.5%	8.0%

Table 2 Diagnostic accuracy parameters from the confusion matrix

Item	Sensitivity	Specificity	MCC	AUROC
Topic recall				
Item 1	0.82	0.48	0.33	0.72
Item 2	0.80	0.26	0.04	0.63
Item 3	0.92	0.34	0.30	0.66
Content prediction				
Item 1	0.84	0.47	0.34	0.73
Item 2	0.80	0.42	0.23	0.63
Item 3	1.00	0.18	-0.01	0.66

Students rated their confidence for the third item at a somewhat high level (3.53), which agreed with a high percentage of accuracy (79.2%). Other parameters from the matrix also displayed high values. However, the MCC score at .30 and AUROC at .66 were considered a low level of prediction efficiency indicating that students' prediction was slightly better than random. This disagreement tends to reveal that even though students had a better correct prediction, this correctness was not better than a random chance.

For the content prediction, the gamma correlation between student prediction score and content scores was calculated to measure relative accuracy. The Gs for the three items were .495 (p < .001), .305 (p < .001), .385 (p < .001), respectively. This result displays that the overview agreement between the prediction score and content scores was correlated at low to medium levels. Later, the diagnostic accuracy was analyzed by a confusion matrix and displayed in Table 2.

For the first item, it was found that students showed a moderate level of confidence (3.24 out of 5) with 36.2 percent correct answer. When considering all parameters in the matrix including MCC and AUROC, such show slightly positive. It could be concluded that students were unsure about their answers. The qualitative data also demonstrated that many students were confused about the structures of DNA, RNA, and nucleotides. Students were able to merely recall where they did recognize the structures, but they did not know how to distinguish them or provide related details about them.

For the second item, students also displayed moderate level of confidence (3.59 out of 5). However, this time they could answer more correctly at 57.5 percent and slightly higher matrix parameters. Surprisingly, the MCC and AUROC scores were lower. This may suggest that the higher average accuracy could have resulted from a random chance rather than their actual knowledge. On the other hand, it is indicated that students were off-target about their prediction.

The participants had higher confidence for the last item as they rated 3.75 out of 5. However, only 8 percent of students could answer correctly, which is consistent with all matrix parameters. It is clear that students were hugely overconfident in this item. Without a doubt, most of them were unable to recognize the differences between glyoxylate and citric acid cycles, which share many similarities.

The Correlation between Students' Prediction and Academic Achievement

In the second part of the study, we wanted to examine if overconfidence was distributed evenly among students with different grades. It is possible that students with better grades might demonstrate higher metacognitive monitoring than weaker students. Therefore, we tested the correlation between exam scores and prediction scores and found a significant moderate correlation (topic recall, r_s =.495, p < .001; content prediction, r_s = .452, p < .001). When considering the accuracy of the prediction, the exam scores showed a better correlation, r_s = .624, p < .001, suggesting that students who excelled in the topics could also potentially become good predictors. In order to investigate this postulation, we classified data from the confusion matrix based on students' grade as shown in Figure 1.

It was noticeable that students with C+ and above had the ability to perform more 'hit' predictions (more than 50% on the average) whereas students with lower grades had sequentially decreased the hit and were lowest in F group. However, the H-test did not confirm the difference (H(7,48) = 8.776, p = .269). Student in the F group also had highest 'correct rejection' rate at 30 percent. This rate gradually decreases in higher grade groups and appears statistically significant (H(7,48) = 30.548, p < .001) when compared with grade A-C. Finally, we found that students of all grades demonstrated the same level of 'false alarm' or underconfidence and 'miss' or overconfidence, which was confirmed by the H-test H(7,48) = 2.981, p = .887 and H(7,48) = 10.794, p = .148, respectively.

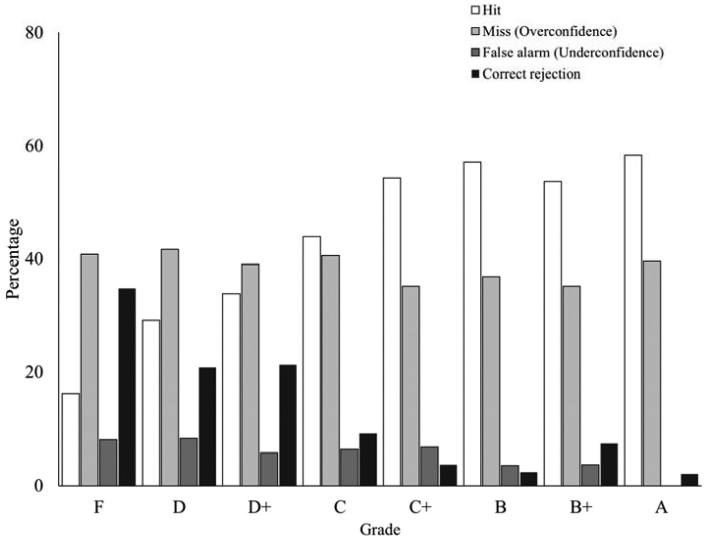

Figure 1 Percentage of average responses distributed by their received grades

Discussion

In this research, the metacognitive monitoring of students in a biochemistry class was measured based on the signal detection theory via the confusion matrix technique for the in-depth analysis of two judgement of learning questions. The participants were requested to perform JOLs by recalling topics and predicting questions associated with three biochemistry figures. Our goals were to investigate JOLs based on visual pictures and categorize patterns of metacognitive performances based on their actual grades. Some interesting aspects were shown here.

First, participants displayed an alarming level of overconfidence. When encountering visual clues, many students believed that they knew what those figures were, but in reality, they did not. It has been long recognized that overconfidence is commonly noticed in college students (de Bruin et al., 2017) or known as the Dunning-Kruger effect (Kruger & Dunning, 1999) by which new learners are likely to create bubbles of ignorance and overconfidence. This finding is not surprising but could be problematic since biochemistry heavily relies on visualization and representation (Mnguni et al., 2016).

At a closer look, the metacognitive monitoring reflected from correct recalls and predictions was not consistent across three tasks. The MCC values substantially varied from task to task, even though the responses were obtained from the same subject. This highly indicated that metacognitive monitoring possesses context specificity. This finding was consistent with the Kelemen et al. (2000) study in which only 8 percent of metacognitive accuracy was correlated with tasks. Therefore, our results support the theory by which metacognitive monitoring is domain-specific rather than generalized. In fact, our data emphasized that even lessons from the same subject could display different metacognitive monitoring as seen from the different subjects (Scott & Berman, 2013).

Lastly, we found that final grades exhibit mixed pattern of metacognitive monitoring; that was grades influenced the correct prediction, but not the errors (Figure 1). There are two types of correct prediction: hit and correct rejection. The hit rate was higher in the strong performance students, but on the other hand, the correct rejection rate was more noticeable in the weaker learners. This finding could be explained by the fact that better students could carry higher confidence in general resulting in higher hit rate. Likewise, the weaker group was likely to negate their ability to predict, or in another word, admit to their lacking of knowledge or refusing to make a sensible prediction. This part of the results strongly agreed with the meta-analysis by Ohtani and Hisasaka (2018) that metacognition is positively correlated with academic outcome.

The error prediction data from the confusion matrix were classified into two terms: overconfidence and underconfidence. In both cases, we found that final grades had no impact on their inaccurate predictions. However, participants in all grades exhibited dramatically higher overconfidence rate than underconfidence. This outcome universally agreed with other findings on which overconfidence in novices is ubiquitously observed in all types of people regardless of their academic performance, levels (Potgiete et al., 2009), and majors (Scott & Berman, 2013). Interestingly, it is known that cultural and ethnic differences could become a factor influencing metacognitive monitoring (Coutinho et al., 2020), and cultural learning also affects metacognition (Heyes et al., 2020). Unfortunately, metacognition studies in Thai educational settings are rarely conducted, and future research about the effects of cultural learning in Thailand is strongly needed.

Conclusion and Recommendation

This study showed that metacognition monitoring is highly specific to surrounding contexts and correlated to learning achievement. In our case, students were mostly overconfident and performed defectively on predicting their abilities. We, as educators, always noticed that many students in our classes were unable to monitor their learnings properly, which this study confirmed. As a result, it is important to develop a method for improving metacognition for Thai students in future studies. In addition, it would be interesting to find more unique effects of Thai cultural learning on metacognition and how to appropriately metacognitive prompt students under different learning styles.

Conflict of Interest

The author declares that there is no conflict of interest.

Acknowledgments

This work was supported by a teaching development grant, Faculty of Science, Silpakorn University. The author also would like to express gratitude to Chatree Faikhamta for proofreading and providing comments for the article.

References

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

de Bruin, A. B. H., Kok, E. M., Lobbestael, J., & de Grip, A. (2017). The impact of an online tool for monitoring and regulating learning at university: Overconfidence, learning strategy, and personality. *Metacognition and Learning*, *12*(1), 21–43. https://doi.org/10.1007/s11409-016-9159-5

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

Fischhoff, B., Slovic, P., Lichtenstein, S. (1997). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(4), 552–564. https://doi.org/10.1037/0096-1523.3.4.552

Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience*, *8*, 443. https://doi.org/10.3389/fnhum.2014.00443

Heyes, C., Bang, D., Shea, N., Frith, C., & Fleming, S. (2020). Knowing ourselves together: The cultural origins of metacognition. *Trends in Cognitive Sciences*, *24*(5), 349–362. https://doi.org/10.1016/j.tics.2020.02.007

Kelemen, W. L., Frost, P. J., & Weaver, C. A. (2000). Individual differences in metacognition: Evidence against a general metacognitive ability. *Memory & cognition*, *28*(1), 92–107. https://doi.org/10.3758/BF03211579

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

Lingel, K., Lenhart, J., & Schneider, W. (2019). Metacognition in mathematics: Do different metacognitive monitoring measures make a difference?. *ZDM Mathematics Education*, *51*, 587–600. https://doi.org/10.1007/s11858-019-01062-8

Livingston, J. (2003). Metacognition: An overview. Psychology, 13, 259–266.

Mnguni, L., Schönborn, K., & Anderson, T. (2016). Assessment of visualisation skills in biochemistry students. *South African Journal of Science*, *112*(9–10), 1–8. https://doi.org/10.17159/sajs.2016/20150412

Potgieter, M., Ackermann, M., & Fletcher, L. (2010). Inaccuracy of self-evaluation as additional variable for prediction of students at risk of failing first-year chemistry. *Chemistry Education Research and Practice*, *11*(1), 17–24. https://doi.org/10.1039/C001042C

Ohtani, K., & Hisasaka, T. (2018). Beyond intelligence: A meta-analytic review of the relationship among metacognition, intelligence, and academic performance. *Metacognition Learning*, *13*, 179–212. https://doi.org/10.1007/s11409-018-9183-8

Scott, B. M., & Berman, A. F. (2013). Examining the domain-specificity of metacognition using academic domains and task-specific individual differences. *Australian Journal of Educational & Developmental Psychology*, *13*, 28–43. https://www.newcastle.edu.au/__data/assets/pdf_file/0008/100241/V13_Scott_Berman.pdf

Siedlecka, M., Paulewicz, B., & Wierzchon, M. (2016). But I was so sure! Metacognitive judgments are less accurate given prospectively than retrospectively. *Frontiers in Psychology*, *7*, 218. https://doi.org/10.3389/fpsyg.2016.00218