Iyer, S. et al. (Eds.) (2022). Proceedings of the 30th International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

# Metacognitive Accuracy in Homework Assignments, Time-Limited Quizzes, and Learning Objectives

**Huanyi CHEN[a*] & Paul A.S. WARD[a]**
*Department of Electrical and Computer Engineering, University of Waterloo, Canada*
*huanyi.chen@uwaterloo.ca

**Abstract:** How confident students are about their solutions is essential to instructors. It reflects how students understand the course materials and how they learn in the course. Studies usually use "Metacognitive Accuracy" to refer to the relationship between a student's confidence or estimated performance and actual performance.

In our study, 170 students estimated their performances for individual tasks in homework assignments and their total performances in time-limited quizzes. The study was conducted in an upper-year database course. In this paper, we propose research questions related to students' estimations in homework assignments, time-limited quizzes, and different learning objectives. We found that students can estimate their homework assignments and time-limited quizzes differently. They also expose different over- and underestimation patterns under different learning objectives.

**Keywords:** Metacognitive Accuracy, Metacognition, Learning Objective, Self-assessment, Student Performance

## 1. Introduction

Understanding one's ability to solve a problem is one's metacognitive skill. A good metacognitive skill can guide the self-regulated learner to learn well. Relevant studies are gaining more and more interest in computing education in recent years (Prather et al., 2020). A commonly accepted hypothesis is that high-performing students have higher metacognitive accuracy than low-performing students (Aghababyan, Lewkow, & Baker, 2018; Hacker, Bol, & Bahbahani, 2008; Hacker, Bol, Horgan, & Rakow, 2000; Harrington, Peng, Jin, & Khan, 2018; Murphy & Tenenberg, 2005). A possible reason why low-performing students tend to have poor metacognitive accuracy is that they do not know what they do not know, and they tend to overestimate their marks compared to those who perform well (Harrington et al., 2018; Kruger & Dunning, 2000). Instructors may conduct early intervention if they have such observations and thus improve students learning.

Although there are several valuable observations, we consider some aspects that can be further explored. First, we wondered if the evaluation format will affect students' performance estimation. For example, homework assignments usually give students days to finish, while the time is usually limited for quizzes. Limit time may lead to stress. Under stress, students may make an inaccurate estimation more often than usual (Reyes, Silva, Jaramillo, Rehbein, & Sackur, 2015). Second, prior studies rarely look into different learning objectives in a course. For example, in a database course, the learning objectives may include aspects from conceptual and theoretical knowledge to practical development and implementation skills (Pahl, Barrett, & Kenny, 2004). Given that the difficulties of different learning objectives may vary from each other, students may have different estimations for different types of learning objectives (John L. Nietfeld, Li Cao, & Jason W. Osborne, 2005). It may suggest learners to take different self-regulated learning strategies on different learning objectives.

Our study was conducted within a database course, and we focused on two research questions:

***RQ1:*** *Is there any difference in the metacognitive accuracy between homework assignments and time-limited quizzes?*

***RQ2:*** *What is the metacognitive accuracy for different types of learning objectives?*

## 2. Related Work

A commonly accepted hypothesis is that high-performing students have higher metacognitive accuracy than low-performing students (Hacker et al., 2008, 2000; Harrington et al., 2018; Murphy & Tenenberg, 2005). For example, Harrington et al. asked students to predict their grades on each question of their final examination in a CS1 course (Harrington et al., 2018). They analyzed the summed predicted grade and found a relatively strong correlation that students who performed poorly in the course were more likely to overestimate their marks than those who performed well. Stephens-Martinez asked students to predict before and after exams, and they found that low performers were worse at predicting than other performers (Stephens-Martinez, 2021). Lee & Liao also recognized similar findings that weak students tend to be more confident than strong students when given identical actual scores (Lee & Liao, 2021).

Those studies primarily collect the metacognitive data with the same time-restriction condition. For example, Lee & Liao used Google Forms without time restrictions (Lee & Liao, 2021); while Hacker et al. collected the data from exams (Harrington et al., 2018), which had a time restriction for students to make their score estimations. However, we consider that students may estimate differently under time stress. Therefore, in our study, we have two different types of time restrictions. Homework assignments allowed students to finish within days, while time-limited quizzes were typically 50 minutes to 1 hour each. Students may show different behavior in their performance estimation under these two time-restriction conditions. We will discuss it in this paper.

John L. Nietfeld et al. found that the difficulty of the question is related to students' metacognitive accuracy (John L. Nietfeld et al., 2005). However, a learning objective may contain test items at multiple difficulty levels, and it is not easy to categorize them accordingly. To our knowledge, we did not see any paper mentioning how students estimate their performance under different types of learning objectives. Our paper will present our findings on how students estimate their performance over coding practice, theory, theory & practice, and design questions.

## 3. Course Background

This study was conducted in a database course in an R1 institution (name omitted for ethics reason) in Canada. The learning objectives of this course cover a wide range of topics that include but are not limited to entity-relationship models, relational algebra, functional dependencies, basics of SQL, and performance.

There were three homework assignments, and three quizzes were used in the study from 170 students who made performance estimations for all assignments and quizzes. The sample solutions to these homework assignments were provided soon after the due date, with a detailed explanation of the thinking process regarding how the answers were written up, hoping to help students gain helpful feedback before the marking is done. The quizzes were used similarly to evaluate students' knowledge using exams and help the instructional team better understand how students learned.

The assignments and quizzes were released alternatively, meaning assignment 1, quiz 1, assignment 2, quiz 2, assignment 3, then quiz 3. There was a fourth assignment in the course. However, since we cannot access the individual score of its tasks while we can for other assignments, so it was excluded from the study. The questions in a quiz were primarily related to the same learning objectives covered in its previous assignment. The only exception is that quiz 3 contains one question to cover some learning objectives in the fourth assignment. Since we focused on metacognitive accuracy within assignments and quizzes separately, we did not exclude that quiz. The solutions for quizzes were not provided to students.

Depending on the difficulties, homework assignments gave students a different number of days ranging from 10 days to 14 days to finish. There were multiple tasks in each assignment. A task in an assignment will have a clear section name that connects to a learning objective. The last task of each assignment is a metacognition task, which asks students to provide their estimated performances out of 100 percent for every other task.

All quizzes were open-book. The first two quizzes were 50 minutes each, and the last quiz was 60 minutes. A short extension period was provided for every quiz, but with a penalty (1% per minute). The last question of each quiz is a metacognition question, which asks students to provide their estimated performance out of 100 percent for the entire quiz. Students could see how many points there were for each question in a quiz, but they cannot see that in assignments, even the total points.

The metacognition task was a blip in the grade, and its effect was lost in the final grade as the instructor rounded up the grade to the least large letter grade equivalent. However, this process is hidden from students.

## 4. Results

For all the actual performance discussed in this section, penalties were not included, as we consider students were not likely to include the penalty in their estimation.

### *4.1 RQ1: Is there any difference in the metacognitive accuracy between homework assignments and time-limited quizzes?*

This research question aims to see if students estimate their performances differently under different assignment formats, with or without time constraints. We used two sample two tailed Welch's t-test in our experiments.

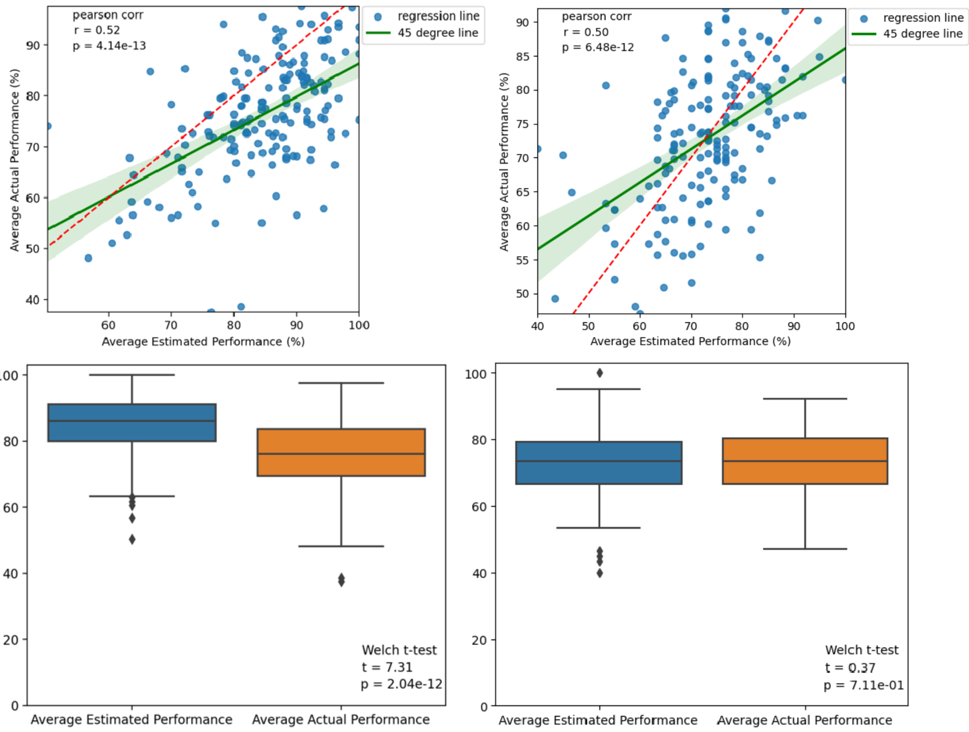


*Figure 1.* Average Estimated Performances and Average Actual Performances of Each Student on Assignments (left) and Quizzes (right) separately

Figure 1 (top-left and top-right) tell us that the correlations between students' average estimated performances and their average actual performances is at a similar level for homework assignments (r = 0.52) and time-limited quizzes (r = 0.50). A predictive model may be built based on this correlation.

The interesting finding comes from Figure 1 (bottom-left and bottom-right), which are box plots of the absolute average values of students' estimated performances and their actual performances. From Figure 1 (bottom-left), it is showing that when estimating their assignment performances, students have a significant trend to overestimate their performance ($t = 7.31$, $p < 0.01$); however, students tend to make accurate estimation when it comes to time-limited quizzes ($t = 0.37$, $p > 0.05$), as shown in Figure 1 (bottom-right). A significant difference in the decrease between students' actual performance and their estimation was also observed ($p = 0.005 < 0.05$).

For homework assignments, since students were given more time to finish, they could refer to course materials or ask for help from the instructional team whenever they faced any difficulties in completing the assignment. The online discussion forum shows that the average response time in the course was half an hour. Students may gain further understanding as well as confidence as they get help. Therefore, they may have more confidence in their assignment answers, although the answers may not be entirely correct.

However, in time-limited quizzes, students must quickly pick up their knowledge and apply it to the questions. There will not be much extra time to seek help or review course materials. Under such conditions, students may have less confidence in their answers, thus resulting in low estimation. The more significant decrease in their estimation from assignments to quizzes also provides evidence that students were not so confident about their answers. However, the high metacognitive accuracy in the quizzes is inconsistent with prior work (Harrington et al., 2018). We will discuss it in the discussion section later.

### *4.2 RQ2: What is the metacognitive accuracy for different types of learning objectives?*

Table 1 shows us the types of different learning objectives for assignment tasks. We grouped tasks based on their type, and we ended up with three Coding Practice tasks, four Theory & Coding Practice tasks, one Design task, and one Theory task.

Table 1. *The Types of Learning Objectives for Assignment Tasks*

| Assignment | Task | Short Description | Type |
|---|---|---|---|
| 1 | 1 | CLI and SQL Commands | Coding Practice |
| | 2 | Embedding SQL in a client program | Coding Practice |
| | 3 | Relational Algebra Queries | Theory |
| 2 | 1 | Relational Algebra to SQL | Theory & Coding Practice |
| | 2 | Attribute Decomposition | Coding Practice |
| | 3 | Primary and foreign keys | Theory & Coding Practice |
| 3 | 1 | Entity-Relationship Model | Design |
| | 2 | Translation from ER Model to Relational Schema | Theory & Coding Practice |
| | 3 | Dependencies | Theory & Coding Practice |

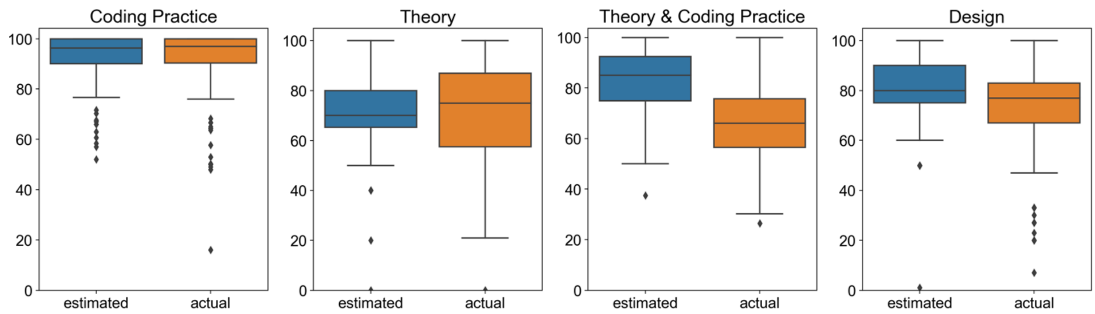


*Figure 2.* Average Estimated Performance and Actual Performance for different types of Learning Objectives

The coding practice task does not need students to understand complex concept. All students needed to do was coding. The theory task is a math-like question. There is no coding component. Students need to understand the concept to get the answer correct. The theory & coding practice task needs students to write code. However, they can only write correct code if they understand the underlying theory correctly. The design task asks students to create an entity-relationship diagram. There can be multiple correct solutions.

Figure 2 shows us the average estimated performance and actual performance of each student for different types of learning objectives. We can see that in both coding practice and the theory tasks, students made accurate estimations; however, for theory & coding practice tasks and the design task, overestimation can be observed.

Given that effort and practice are related to mindset for programming, programming-efficacy, and performances (Scott & Ghinea, 2014), it is reasonable that students achieved a reasonable metacognitive accuracy for coding practice tasks ($p = 0.68 > 0.05$, Welch's t-test on students' estimation and their actual performance). Students could simply try out their solutions using the provided database server to see whether a solution was correct. Therefore, it is expected that they achieved both high metacognitive accuracy and high performance for those coding practice tasks.

Surprisingly, students did not make significant over- or underestimation in the theory-only task ($p = 0.25 > 0.05$). It may suggest that students have good metacognition skills in estimating theory questions in general. However, we can still see that the lowest estimation was still higher than the lowest actual performance. It suggests that low-performing students tend to overestimate their performance, which is consistent with prior work (Harrington et al., 2018; Lee & Liao, 2021).

Students were generally overestimating for theory & coding practice and design tasks. For the design task ($p < 0.01$), it is reasonable since students rarely had such type of question. The correlation between their estimated performance and their actual performance was also low ($r = 0.03$, Figure not shown in the paper). Thus, we consider they were overestimating it under the Dunning–Kruger effect (Kruger & Dunning, 2000). In fact, most students tend to believe their solutions were (at least half) correct, thus overestimating the task blindly. For the theory & coding practice tasks ($p < 0.01$), we consider it is the great combined difficulty of such tasks that leads students to overestimate. It is found in other fields that people tend to overestimate their actual performance when tasks are difficult (Krawczyk & Wilamowski, 2019; Moore, Healy, & Hall, 2008).

## 5. Discussion

Our result (Section 4.1) implies that there is a difference in metacognitive accuracy between students' assignment performances and students' quiz performances. The data in our study shows that students could make reasonably accurate estimations of their actual performance in quizzes without receiving markings or feedback for the previous assignment. It has been found that students' estimation can be calibrated on easy items (John L. Nietfeld et al., 2005). As we provided the sample solution of assignments to students, we consider the sample solution played an important role in calibrating students' metacognitive accuracy. However, we do not consider all the quizzes easy. A preliminary experiment (not shown due to page limit) shows that students only show an inaccurate estimation in quiz 3, which mainly contains design questions. However, they had accurate estimations for quiz 1 and quiz 2. It is accurate for quiz 1 because quiz 1 mainly contains coding practice and theory questions. Students tend to make accurate estimations of these types of learning objectives (see Figure 2). However, quiz 2 mainly contains the theory & coding practice questions, which are considered difficult; thus, calibration may not exist. Further analysis is needed to reveal the real reason.

## 6. Threats to Validity

The first threat to validity would be that students cooperate sometimes. It was not likely that students over-cooperated in quizzes since those quizzes had a very intense time setup. However, it was possible that some students over-cooperated for some assignments in the course. Therefore, some of the answers for assignments may not reflect their real metacognition. However, those cases should be considered trivial. Otherwise, they would be recognized by the teaching assistants who marked their assignments.

Second, which could potentially be the biggest threat, it is possible that students did not submit their thoughtful estimation, but instead, they filled up a random number. Those numbers should be considered outliers. However, since their estimation was worth a few points (although its effect would be lost when calculating the final grades, students do not know that), we consider that most of them would try their best to make accurate estimations.

## 7. Conclusion

This paper investigated students' metacognitive accuracy in homework assignments, time-limited quizzes, and different learning objectives in a database course. We found that the correlations between students' average estimated performances and their average actual performances are similar; however, when estimating their assignment performances, students have a significant tendency to overestimate their performance, but they tend to make accurate estimations when it comes to time-limited quizzes, although the reason behind remains hidden. We found that students made accurate estimations for the coding practice and the theory tasks; however, overestimation can be observed for the theory & coding practice and the design tasks.

## References

Aghababyan, A., Lewkow, N., & Baker, R. S. (2018). Enhancing the Clustering of Student Performance Using the Variation in Confidence. In R. Nkambou, R. Azevedo, & J. Vassileva (Eds.), *Intelligent Tutoring Systems* (pp. 274–279). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-91464-0_27

Hacker, D. J., Bol, L., & Bahbahani, K. (2008). Explaining calibration accuracy in classroom contexts: The effects of incentives, reflection, and explanatory style. *Metacognition and Learning*, *3*(2), 101–121. https://doi.org/10.1007/s11409-008-9021-5

Hacker, D. J., Bol, L., Horgan, D. D., & Rakow, E. A. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology*, *92*(1), 160–170. https://doi.org/10.1037/0022-0663.92.1.160

Harrington, B., Peng, S., Jin, X., & Khan, M. (2018). Gender, confidence, and mark prediction in CS examinations. *Proceedings of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education*, 230–235. Larnaca Cyprus: ACM. https://doi.org/10.1145/3197091.3197116

John L. Nietfeld, Li Cao, & Jason W. Osborne. (2005). *Metacognitive Monitoring Accuracy and Student Performance in the Postsecondary Classroom*. 7–28.

Krawczyk, M., & Wilamowski, M. (2019). Task difficulty and overconfidence. Evidence from distance running. *Journal of Economic Psychology*, *75*, 102128. https://doi.org/10.1016/j.joep.2018.12.002

Kruger, J., & Dunning, D. (2000). *Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments*. 14.

Lee, P., & Liao, S. N. (2021). Targeting Metacognition by Incorporating Student-Reported Confidence Estimates on Self-Assessment Quizzes. *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*, 431–437. Virtual Event USA: ACM. https://doi.org/10.1145/3408877.3432377

Moore, D., Healy, P. J., & Hall, A. (2008). *The Trouble with Overconfidence*. 61.

Murphy, L., & Tenenberg, J. (2005). *Do Computer Science Students Know What They Know?: A Calibration Study of Data Structure Knowledge*. 5.

Pahl, C., Barrett, R., & Kenny, C. (2004). *Supporting Active Database Learning and Training through Interactive Multimedia*. *Vol.36*(3), 27–31.

Prather, J., Becker, B. A., Craig, M., Denny, P., Loksa, D., & Margulieux, L. (2020). What Do We Think We Think We Are Doing?: Metacognition and Self-Regulation in Programming. *Proceedings of the 2020 ACM Conference on International Computing Education Research*, 2–13. Virtual Event New Zealand: ACM. https://doi.org/10.1145/3372782.3406263

Reyes, G., Silva, J. R., Jaramillo, K., Rehbein, L., & Sackur, J. (2015). Self-Knowledge Dim-Out: Stress Impairs Metacognitive Accuracy. *PLOS ONE*, *10*(8), e0132320. https://doi.org/10.1371/journal.pone.0132320

Scott, M. J., & Ghinea, G. (2014). On the Domain-Specificity of Mindsets: The Relationship Between Aptitude Beliefs and Programming Practice. *IEEE Transactions on Education*, *57*(3), 169–174. https://doi.org/10.1109/TE.2013.2288700

Stephens-Martinez, K. (2021). A Study of the Relationship Between a CS1 Student's Gender and Performance Versus Gauging Understanding and Study Tactics. *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*, 679–685. Virtual Event USA: ACM. https://doi.org/10.1145/3408877.3432365