〈Others〉

Evaluation for the effectiveness of peer assessment activities in biochemistry education

Kei MIYANO[1), Akira YAMAUCHI[2), Futoshi KURIBAYASHI[2), Shuichiro OKAMOTO[2)

1) Department of Natural Sciences,
2) Department of Biochemistry, Kawasaki Medical School

ABSTRACT **Background**: Peer assessment, a form of active learning, has been widely implemented in educational institutions with noted effectiveness in achieving educational goals. However, as many factors have been found to influence peer assessment efficacy, it is unclear which combinations promote the achievement of educational goals.

Methods: This study describes a peer assessment activity used in a biochemistry lecture at the Department of Medical Engineering students, with the educational goals being the acquisition of the "knows" component and aspects related to the "knows how" component of Miller's pyramid of competence. Three types of tests were used to analyze the achievement of these objectives: fill-in-the-blank, multiple choice, and open-ended.

Results: The results suggest that peer assessment might be related to teaching the application of knowledge and the acquisition of the "knows how" component.

Conclusions: These results suggest that traditional lectures are effective for increasing knowledge, while peer assessment helps promote its application. For these reasons, educators should choose the lecture style that best suits their educational goals.

doi：10.11482/KMJ-E202652007　*(Accepted on January 18, 2026)*

Key words：Active Learning, Peer Assessment, Miller's pyramid of competence

INTRODUCTION

Learning biochemical metabolic pathways is essential for students, but it can be quite challenging. Various educational techniques, including active learning, have been developed to facilitate understanding. Active learning has received considerable attention in medical education in recent years. It requires students to engage in meaningful learning activities and think carefully about their choices and rationale[1)]. There are various reports on the many kinds of active learning and their effectiveness[2)]. Among them, peer assessment allows learners to evaluate the level, value, or quality of a product or performance of their peers and improve their learning by giving detailed feedback and discussing their judgments with peers to achieve a shared conclusion. Peer assessment is also known as peer review, peer

Corresponding author
Shuichiro Okamoto
Department of Biochemistry, Kawasaki Medical School, 577 Matsushima, Kurashiki, 701-0192, Japan

Phone : 81 86 462 1111
Fax : 81 86 462 7897
E-mail: shuokamoto@med.kawasaki-m.ac.jp

feedback, and peer response[3]. Although many studies have discussed the efficacy of peer assessment in education, many elements influence this efficacy, including student level and course subject[4]. Few reports examine the effectiveness of peer assessment in biochemistry subject, and its effectiveness has not been appropriately analyzed. This study therefore explored the efficacy of peer assessment for achieving educational goals in a biochemistry class, which were the achievement of the "knows" and "knows how" aspects of Miller's pyramid of competence[5].

MATERIALS AND METHODS

Research collaborators

The participants in this study were 46 sophomore students from the Department of Medical Engineering at Kawasaki University of Medical Welfare. A comparative study group was created by selecting students who had been absent from the lectures.

Curriculum

The content of the first eight out of fifteen biochemistry lectures was evaluated. The lecture covered carbohydrate, lipid, and amino acid metabolism. The first six lectures were traditional lectures (TLs), the seventh comprised peer assessment (PA), and the eighth lecture was an examination.

Peer assessments

The PA session was divided in half. The first half consisted of a presentation and assessment activity, and the second was a feedback activity. Students were divided into teams of four, and all students participated in all parts of the process. All students were tasked with creating materials for their presentations. In their materials, they were instructed to summarize the lecture by summarizing carbohydrate metabolism and lipid metabolism on a single map. Students gave presentations based on their materials. Each presentation was 10 minutes. The other three students assessed the presentations using the given assessment rubric (Table 1). In the second half of the session, they provided feedback and gave each other advice on improving their presentations.

Examination

Three types of tests were administered in the eighth lecture. Because the lecture objectives were to learn the basics of metabolism and to understand metabolic diseases, we evaluated the achievement of the "knows" component and aspects related to the "knows how" component of Miller's pyramid[5]. Fill-in-the-blank and multiple choice questions (MCQs) were included to assess the aspects of the "knows how" component, i.e., the factual recall of information. Open-ended questions were included to assess the "knows how" component, i.e., the application of knowledge to problem-solving and decision-making. Assessing the achievement of "knows how" is very difficult because it requires evaluating responses generated by thinking[6]. To address this issue, we applied the Mosenthal Taxonomy[7-8], which classifies questions into five levels based on their degree of abstraction. The Fill-in-the-blank questions consisted of eight items that asked the most concrete questions, such as the names of metabolites in the metabolic pathway. The MCQs consisted of 28 items that asked highly concrete and intermediate questions like the names of hormones that promote metabolic pathways. The open-ended questions consisted of three items that assessed higher levels of abstraction by asking about the roles of metabolic pathways. The exam had a maximum score of 45 points, with each Fill-in-the-blank question and MCQ worth 1 point, and each open-ended question worth 3 points. To evaluate the "knows how" level, these items included questions that required integrative reasoning, such as "Explain

Table 1. Active Learning Evaluation Sheet

Please evaluate each criterion on a 5-level scale (0-4 points).
The maximum total score is 28 points.

Evaluation Items	Evaluation Scale				
	Excellent (4)	Good (3)	Acceptable (2)	Needs Improvement (1)	Fail (0)
Glycolysis	Explains the role of glycolysis in overall metabolism. Its relationship with the TCA cycle, etc.	It correctly explains the production and consumption of ATP during glycolysis.	The metabolic products necessary to explain glycolysis are described.	The starting and finishing products of the metabolic pathway are correctly described.	No metabolic pathway is described at all.
TCA Cycle	Explains the role of the TCA cycle within overall metabolism.	The metabolic products necessary to explain the TCA cycle are described.	It explains what citric acid is synthesized from.	Explains where the TCA cycle takes place. (e.g., cytoplasm, mitochondria, etc.)	No metabolic pathway is described at all.
Electron transport chains	It explains ATP production in the Electron transport chain.	The metabolic products necessary to explain the Electron transport chain are described.	The starting and finishing products of the metabolic pathway are correctly described.	Explains where the Electron transport chain takes place. (e.g., cytoplasm, mitochondria, etc.)	No metabolic pathway is described at all.
Anaerobic glycolysis	The role of anaerobic glycolysis is explained, including the Cori cycle.	It explains the purpose of synthesizing lactic acid using LDH in anaerobic glycolysis.	It's pointed out that anaerobic glycolysis causes lactic acid to accumulate in muscles.	It's explained that ATP synthesis is possible through glycolysis even in anaerobic conditions.	Fails to explain anaerobic glycolysis or ATP production under anaerobic conditions.
Glycogen metabolism	Explains the differences in glycogen metabolism between fasting and fed states.	It explains the mechanisms of hormone-induced regulation of glycogen metabolism.	The relationship between the glycogen metabolic pathway and glycolysis is shown on the metabolic pathway map.	Explains in which organs glycogen is stored.	Fails to explain that glycogen is a polysaccharide or provides no relevant explanation.
Gluconeogenesis	It explains the metabolic products that serve as materials for gluconeogenesis. (3 types of substances)	The relationship between gluconeogenesis and glycogen metabolism can be explained as a mechanism for maintaining blood glucose levels.	The gluconeogenesis pathway is correctly depicted in the metabolic pathway map.	Explains in which organs gluconeogenesis is carried out.	Fails to explain that glucose is produced via gluconeogenesis.
Lipid metabolism	Explains the role of ketone bodies in overall metabolism.	Explains how the carbon skeleton of a fatty acid is metabolized during beta-oxidation.	Explains the relationship between glycerol metabolism and gluconeogenesis.	Explains the relationship between the breakdown of triglycerides and adrenaline.	Fails to explain that triglycerides consist of glycerol and fatty acids.

Instructions: Assign a score (0-4) for each of the seven criteria. Calculate the final score by summing all criterion scores.

how glycogen metabolism changes from the fed to the fasted state." This question format may have enabled a partial assessment of the "knows" and "knows how" levels of Miller's pyramid.

Statistical analysis

Statistical analysis was conducted using Student's t-test or one-way ANOVA, followed by Bonferroni correction for multiple comparisons. Bonferroni-adjusted p-values were calculated by multiplying the original p-values by the number of comparisons, and a Bonferroni-adjusted p-value of < 0.05 was considered statistically significant.

RESULTS

Because some students were absent from the PA session, the Fill-in-the-blank question, MCQ, and Open-ended question scores were first compared between students who had attended and those who had not. The first group consisted of 15 students who had been absent from the PA session (PA non-Attendance). The second included the 31 students

Table 2. Number of students by attendance at PA lectures

Total number of students	PA non-Attendance	PA Attendance
46	15	31

All 46 students were classified according to their attendance at PA lectures.

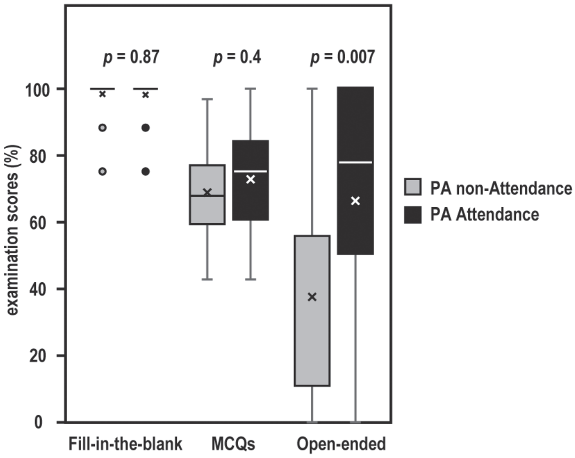

Fig. 1. Comparison of test scores (two groups)
The PA non-Attendance group is indicated by gray bars (n = 15). The PA Attendance group is indicated by black bars (n = 31). The highest possible score of each test was 100%. The error bars indicate the SD. Statistical analysis was performed using Student's t-test.

Table 3. Number of students by type of lecture absence

Number of Absences	Number of people	PA non-Attendance	TL non-Attendance	ALL Attendance	Both Absent
0	21	0	0	21	0
1	19	11	8	0	0
2	4	0	2	0	2
3	1	0	0	0	1
4	1	0	0	0	1
(Total)	46	11	10	21	4

Students were categorized according to the number of absences (left column), and each group was further subdivided based on the type of lecture missed. Eleven students who were absent from PA sessions were classified into the PA only non-Attendance group. Among the ten students absent from TL sessions, eight who missed only one session were classified as the TL-only non-attendance group.

who had been present (PA Attendance) (Table 2). No differences were found between the PA Attendance and PA non-Attendance groups in their average scores for the Fill-in-the-blank questions and MCQs. Conversely, the average score of the Open-ended questions in the PA non-Attendance group was lower than that in the PA Attendance group (Fig. 1). However, students in the PA non-Attendance group were more likely to have also been absent from the TLs than those in the PA Attendance group. Students in the PA non-Attendance groups missed an average of 1.47 out of the total eight lectures, while students in the PA Attendance groups missed an average of 0.38 times. This was a significant difference between the two groups ($p = 0.00002$). Because the observed effects may have been influenced by the overall lecture absence rate rather than by the PA lectures themselves, a new grouping was created (Table 3).

In the subsequent analysis, students who had missed multiple lectures were excluded to control for the effect of absence frequency. The first group included 11 students who were absent from only the PA session (PA only non-Attendance). The second group included eight students who were absent from TLs (TL only non-Attendance). Finally, 21 students who attended PA and all TLs were included in a third group (All-Attendance).

Among the three reformulated groups (PA only non-Attendance, TL only non-Attendance, All-Attendance), there was no substantial difference in the average Fill-in-the-blank question scores (Fig. 2). In the MCQs, there was no difference between the average scores of the PA only non-Attendance and All-Attendance groups. However, the average MCQ score of the TL only non-Attendance group was significantly lower than that of the All-Attendance group (p = 0.008). For the Open-ended questions, the PA only non-Attendance group's average score was significantly lower than that of the All-Attendance group (p = 0.007). However, there was no difference between the average scores of the TL only non-Attendance and the All-Attendance groups.

DISCUSSION

This study assessed the efficacy of PA activities for meeting biochemistry educational goals. First, examination results were compared between students who had missed the PA session and those who had not. Although the scores of the PA non-Attendance group were significantly lower than that of the PA Attendance group, this difference may have also been influenced by missed TLs. Thus, students who were absent from either the PA session or only one time TL were placed in the PA only non-Attendance or TL only non-Attendance groups, respectively. There were no differences between these groups' average examination scores for all three tests. The third group (All-Attendance) consisted of students who were present for the PA session and all TLs, which allowed for the further analysis of the influence of missing lectures on examination scores.

There were no differences between the three groups in their average Fill-in-the-blank question scores. As the average score was 97%, this could be

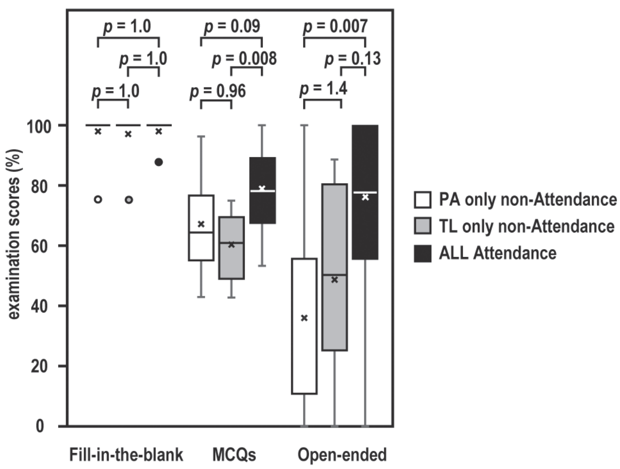

Fig. 2. Comparison of test scores (three groups)
The PA only non-Attendance group (n = 11) is indicated by the white bar. The TL only non-Attendance group (n = 8) is indicated by the gray bar. The All-Attendance group (n = 21) is indicated by the black bar. The highest possible score for each test was 100%. The error bar indicates the SD. Statistical analysis was performed using one-way ANOVA followed by the Bonferroni adjustment.

because the questions were too easy to accurately assess knowledge. For the MCQs, the average score of the TL only non-Attendance group was significantly lower than the All-Attendance group. As the MCQs assessed the "knows" component of Miller's pyramid, these results suggested that TLs play a crucial role in building theoretical knowledge. For the Open-ended questions, the average score of the PA only non-Attendance group was significantly lower than that of the All-Attendance group. As the Open-ended questions assessed higher-order knowledge integration related to the "knows how" level of Miller's pyramid, these results suggested that participation in PA activities may be associated with students' ability to apply theoretical knowledge taught in the TLs within more practical contexts.

These results indicate that attendance at TLs was associated with higher knowledge scores and may facilitate the achievement of the "knows" component of Miller's pyramid of competence. Meanwhile, PA activities assist in learning how to apply information and may be associated with aspects related to the "knows how" component of Miller's pyramid. However, a concern remains regarding the group assignment based on lecture attendance. Students who attended the lectures may have had higher motivation for learning, which could have influenced the results of this study. Therefore, it cannot be ruled out that the higher test scores observed in the lecture-attending group were due to student motivation rather than the effect of the lectures themselves. Considering these issues, we plan to conduct future studies using a pre-post comparison design.

LIMITATIONS

This study was conducted to verify the effectiveness of PA, but several issues remain. One of the crucial issues is the group assignment, as explained in the Discussion section. Another issue is the rubric used to evaluate PA. The learning goal of PA was to understand the overall metabolic relationships through an understanding of metabolic maps. However, many of the evaluation items used in the rubric solely assessed understanding of metabolic pathways, such as glycolysis and the TCA cycle, and were not capable of assessing understanding of the relationships between metabolic pathways. Furthermore, content validity and construct validity were not conducted, and improvements are needed. We plan to continue this research while improving these points.

ACKNOWLEDGEMENTS

We are grateful to Hiroko Sakamoto of the Kawasaki University of Medical Welfare for her assistance with the course preparation.

REFERENCES

1) Bonwell CC, Eison JA: Active Learning: Creating Excitement in the Classroom. ASHEERIC Higher Education Report No. 1 George Washington University Washington DC 1991. pp1-50.
2) Prince M: Does Active Learning Work? A Review of the Research. J. Eng. Educ. 2004; 93 (3): 223-231.
3) Topping KJ: Peer Assessment: Learning by Judging and Discussing the Work of Other Learners. Interdiscip. Educ. Psychol. 2017; 1 (1): 1-17.
4) Topping KJ: Trends in Peer Learning. Educ. Psychol. 2005; 25 (6): 631-645.
5) Miller GE: The Assessment of Clinical Skills / Competence / Performance. Acad. Med. 1990; 65 (9): S63-S66.
6) Wass V, Van der Vleuten C, Shatzer J, Jones R: Assessment of Clinical Competence. Lancet. 2001; 357 (9260): 945-949.
7) Ozuru Y, Briner S, Kurby CA, Mcnamara DS: Comparing Comprehension Measured by Multiple-Choice and Open-Ended Questions. Can. J. Exp. Psychol. Rev. Can. Psychol. 2013; 67 (3): 215-227.
8) Graesser AC. Ozuru Y. Sullins J: What is a good question? In M. McKeown & L. Kucan (Eds.), Bringing reading research to life. New York, NY: Guilford Press; 2010. 112-141.