Enhancing Assessment of Student Engagement in Face-to-Face Global Project-Based Learnings (gPBLs): Adding Peer Assessment to Improve Slack-based Evaluation

Mr. Leo Kimura, Shibaura Institute of Technology

Leo Kimura received his Bachelor of Engineering in Science and Mechanics from Shibaura Institute of Technology (SIT), Japan, in March 2024. He is currently a master's program student at SIT, majoring in Mechanical Engineering. His main research area is control systems for microrobots.

Prof. Hatsuko Yoshikubo Ph.D., SHIBAURA INSTITUTE OF TECHNOLOGY

Dr. Hatsuko Yoshikubo obtained Ph. D. in English Literature from Chiba University in Japan in 2002. She is currently a Professor of Humanities and a senior departmental administrator, currently attached to the Innovative Global Program in the College of Engineering at Shibaura Institute of Technology, a highly-ranked engineering university in Tokyo. Innovative Global Program (IGP) is an engineering degree program incorporating humanities components, with a student body consisting mainly of international students (https://igp.shibaura-it.ac.jp).

As part of her current role as Deputy Director of the IGP, she is the Principal Investigator for the following research grants: - Japan Society for the Promotion of Science Research: Grant 24K06133 (2024-2027) - Shibaura Institute of Technology Grant for Educational Reform and Research Activity (AY2024).

Her research interests include: (1) Impacts and potential benefits of including humanities components in science and engineering degree programs (2) Innovative methods of assessment in science and engineering education, especially in the context of remote learning, hybrid courses, and collaborative international programs (3) Solving systematic issues that impact the effectiveness of science and engineering education programs, in both in-person and remote learning contexts.

In recent years she has been a presenter at the following international conferences: World Educational Research Association (WERA): WERA Conference 2019, 2022, 2023, 2024 Asia-Pacific Association for International Education (APAIE): APAIE Conference 2022, 2023, American Society for Engineering Education (ASEE): Annual Conference 2022, 2023, 2024 International Institute of Applied Informatics (IIAI): International Conference on Data Science and Institutional Research (DSIR) 2023

Awards: SIT's Presidential Award 2023 ASEE 2023: Multidisciplinary Engineering Division's 'Best Diversity Paper' Award DSIR 2021: 'Outstanding Paper' Award

Prof. Sumito Nagasawa Ph.D. in Engineering, Shibaura

Prof. Dr. Sumito Nagasawa received Ph.D. in Engineering from the University of Tokyo in 2001. He is a Professor in Department of Engineering Science and Mechanics at SIT. His research interests include miniaturized robots using Micro-Electro-Mechanical Systems technologies and robot education for STEAM.

Enhancing Assessment of Student Engagement in Face-to-Face Global Project-Based Learnings (gPBLs): Adding Peer Assessment to Improve Slack-based Evaluation

Abstract
This study aims to introduce and evaluate a peer assessment method as a complementary approach to Slack-based engagement assessment in face-to-face group work. The participants were students who attended a Japan-Thailand joint global project-based learning (PBL) workshop held in 2024. Posts made on Slack were classified into three types—A, B, and C—according to their level of contribution. Additionally, students conducted peer assessments using a four-point Likert scale to evaluate contributions across three dimensions: idea generation, discussion, and task execution. Multiple regression analysis revealed that the number of Type A posts on Slack had a significant positive correlation with peer-assessed contributions in both idea generation and task execution. However, no significant relationship was observed with discussion. Furthermore, Type C posts had a significant negative impact on task execution. These findings indicate that while Slack-based evaluation has certain validity even in face-to-face settings, it has limitations in capturing oral discussions, which are less likely to be documented.

Key words
Peer assessment, Slack-based evaluation, Student engagement in the group project, Face-to-face/Online global project-based learning, Miville-Guzman Universality-Diversity Scale (MGUDS-S), Group activity analysis, Accurate grading

1. Introduction

1.1 Research Objectives

This study aims to experimentally introduce peer assessment as a complementary method to address the limitations of Slack-based engagement evaluation in face-to-face group work, and to examine its effectiveness. In particular, the study investigates whether combining activity scores derived from Slack posts with peer-assessed scores can enable objective evaluation of student engagement, even in a face-to-face setting. In this context, "Slack-based evaluation" refers to a method of assessing student engagement by analyzing their posts recorded on the communication platform Slack, with attention to both the quantity and quality of contributions. Student posts are categorized into three types—Type A, B, and C—based on the degree of contribution to the project. These categories are assigned different weights to generate activity scores, thereby enabling the quantitative visualization of "process contributions" that are often overlooked in conventional evaluations based solely on final outcomes [1][2].

This study analyzes data collected during a face-to-face global Project-Based Learning (gPBL) workshop conducted in 2024 and addresses the following two research questions (RQs):

- **RQ1:** Is there a correlation between Slack-based activity scores and peer assessment scores?
- **RQ2:** Can peer assessment be used to compensate for the limitations of Slack-based evaluation?

1.2 Implementation of PBL in Engineering Education

Project-Based Learning (PBL) is an instructional approach that fosters deep understanding of subject matter by engaging students in real-world projects. It cultivates problem-solving abilities, collaboration skills, and practical competencies for application in real-life contexts [3][4]. Due to these advantages, PBL has been widely adopted in the field of engineering education [5][6], where students acquire hands-on skills by tackling actual technical challenges. In recent years, Global Project-Based Learning (gPBL) programs have also been implemented, with the aim of promoting intercultural understanding and international collaboration skills through teamwork among students from diverse cultural backgrounds [7][8].

At Shibaura Institute of Technology (SIT) in Japan, where the authors are affiliated, gPBL programs are organized to cultivate globally competent science and engineering professionals. Among these initiatives, the authors have annually conducted a collaborative robotics workshop with partner universities abroad. In Academic Years (AY) 2020 and 2021, joint online workshops were held with University Tun Hussein Onn Malaysia, and in AY2022 with King Mongkut's University of Technology Thonburi (KMUTT), Thailand. Although these robotics workshops were originally planned as face-to-face events, the COVID-19 pandemic necessitated a transition to a HyFlex (hybrid flexible) format, in which Japanese students participated physically at the SIT campus, while international students joined remotely. Following the global easing of early COVID-19 restrictions and the resumption of in-person activities, face-to-face workshops were reinstated. In the AY2023 program, KMUTT students were invited to the SIT campus in Tokyo, and in AY2024, SIT students visited KMUTT's campus in Bangkok.

In both the AY2023 and AY2024 robotics workshops, students from both institutions were assigned to mixed teams and engaged in building mechatronics systems using a rescue robot kit developed by the authors [9] (Figures 1-1 and 1-2). These workshops were open to students across disciplines and academic years, with the primary goal of learning foundational concepts in mechatronics through hands-on experience with robot kits. The ability to control the robots as intended helped sustain students' motivation, and the tangible systems—such as mobility mechanisms—enabled intuitive understanding of mechatronic principles. The workshops followed the schedule shown in Table 1, and students were required to complete the team project outlined in Table 2 within a very short period of less than four days. This demanded that participants apply their international collaboration skills to work efficiently and effectively. At SIT, academic credit was granted to participating students. (KMUTT students did not receive academic credit for participation.)

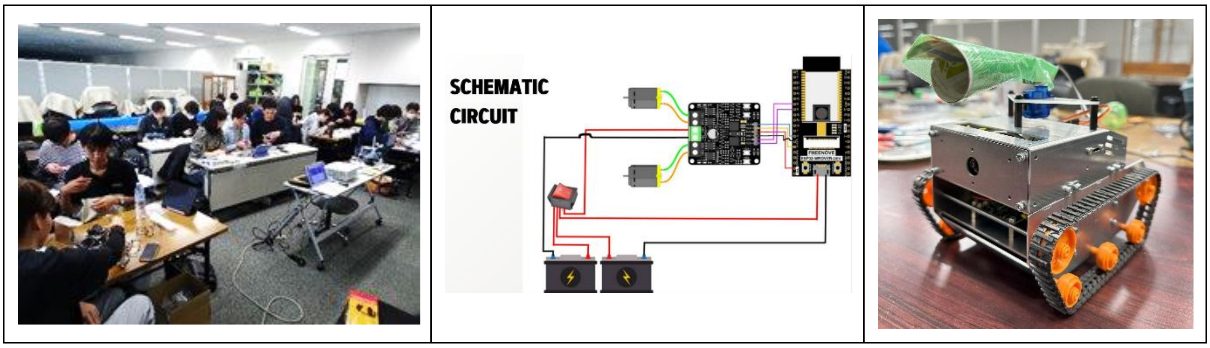

Figure 1-1: Face to Face Robotics workshop in AY2023 at the SIT campus

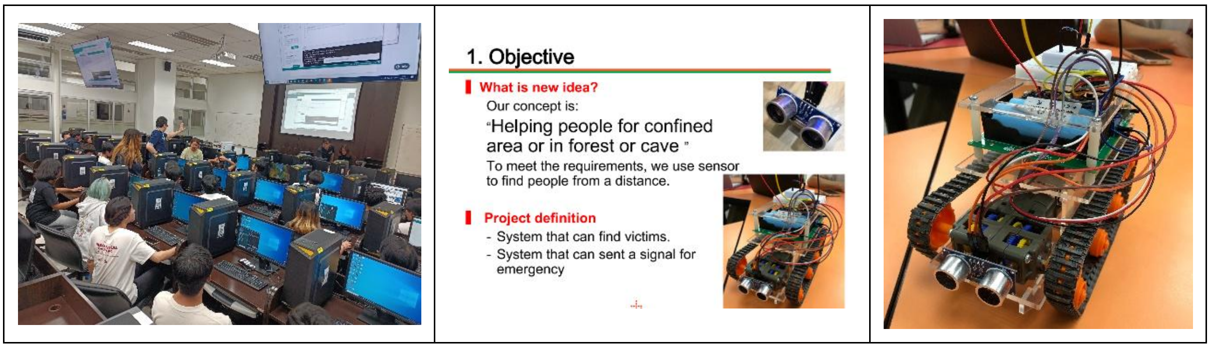

Figure 1-2: Face to Face Robotics workshop in AY2024 at the KMUTT campus

Table 1: Schedule for robotics workshop in AY2024

Day	Hours	Time	Activity
13 Sep (Fri)	1.5h	18:00-19:30	Opening Ceremony / Filling out a survey form / Guidance / Self-Introduction / Cultural Lecture / Team Building
14 Sep (Sat)	6h	10:00-12:00	Robot Kit Assembling
		12:00-13:00	Lunch Break
		13:00-15:00	Programing Exercise 1: Basics of Arduino / ESP32 Programing
		15:00-15:30	Break
		15:30-17:30	Programing Exercise 2: Sensor-Computer-Actuator using Robot Kit
15 Sep (Sun)	8h	10:00-19:00 (Lunch Break 1 hour)	Curtual Activity 1
16 Sep (Mon)	6h	10:00-12:00	Programing Exercise 3: Robot Control with Smartphone
		12:00-13:00	Lunch Break
		13:00-15:00	Programing Exercise 4: Robot Programing with ROS
		15:00-15:30	Break
		15:30-17:30	Team Project 0: Guidance for Team Project
17 Sep (Tue)	8h	10:00-12:00	Team Project 1: Considering Team Concept
		12:00-13:00	Lunch Break
		13:00-15:00	Team Project 2: Deciding Roles / Planning Schedule
		15:00-15:30	Break
		15:30-17:30	Team Project 3: Survey
18 Sep (Wed)	6h	10:00-19:00 (Lunch Break 1 hour)	Curtual Activity 2 (Ayudhaya trip)
19 Sep (Thu)	6h	10:00-12:00	Team Project 4: Feasibility Study
		12:00-13:00	Lunch Break
		13:00-15:00	Team Project 5: System Building 1 / Mechanism-Electronics-Programing
		15:00-15:30	Break
		15:30-17:30	Team Project 6: System Building 2 / System Integration
20 Sep (Fri)	6h	10:00-12:00	Team Project 7: Preparing for Presentation
		12:00-13:00	Lunch Break
		13:00-15:00	Team Project 8: Practice for Presentation
		15:30-17:30	Project Presentation / Closing Ceremony / Certificate

Table 2: Contents of team project

● Theme: Expansion rescue robot’s function. (1) Consider the concept of your team project. (2) Decide the member’s role and time schedule. (3) Perform feasibility study. (4) Programing by PictoBlox or Arduino language (is locked by hardware) . (5) Develop your system. (6) Evaluate your system. ● Make presentation slides using Power Point, Google Slides, and Canva.

Table 3: Attributes of Participants

Study major	Robotics workshop in AY2023		Robotics workshop in AY2024	
	Japanese students	Thai students	Japanese students	Thai students
Mechanical Engineering	13	11	8	15

1.3 Problems in gPBL and background on development of Slack-based evaluation

Several issues have been pointed out regarding both PBL and gPBL[10][11][12][13]. Among these, the authors have focused on a particular issue: the decline in global competency scores, as measured by the Miville-Guzman Universality-Diversity Scale, Short Form (MGUDS-S) [14][15], among some students who participated in gPBL [1]. This issue was observed during the online robotics workshops conducted in Academic Years (AY) 2021 and 2022. The decline in MGUDS-S scores indicates that, despite participating in gPBL, some students did not sufficiently develop intercultural understanding or international collaboration skills, which constitutes a significant concern that undermines the fundamental purpose of the program.

To investigate the underlying causes of this issue, the authors proposed (1) an evaluation framework that combines the MGUDS-S with an original student satisfaction survey, and (2) a method for analyzing online activity logs using Slack (Slack-based evaluation) [1][2]. The MGUDS-S was administered both before and after the workshop, while the satisfaction survey was conducted after the workshop. The relationship between the scores and trends in student feedback was examined.

Slack, in particular, facilitates smooth communication among students and between students and instructors, while also enhancing the efficiency of information sharing [16][17]. Furthermore, the activity logs recorded on Slack make it possible to visualize students' level of participation and workload even in the absence of a facilitator. This enables instructors to conduct more objective and equitable evaluations. As a result, disparities in group contributions—which are often overlooked in conventional evaluation methods that rely solely on final deliverables or peer assessment [18][19]—can be identified. This, in turn, helps reduce perceptions of unfairness among students and contributes to maintaining their motivation.

In practice, students whose MGUDS-S scores declined were found to have provided negative feedback, and Slack logs revealed that tasks had been disproportionately assigned to certain individuals. This suggests that the students' subjective evaluations accurately reflected the actual situation, thereby demonstrating the effectiveness of the proposed methods [1][2].

1.4 Problems with Slack-based evaluation in Face-to-Face group work and implementation of peer assessment

As initial restrictions related to COVID-19 were lifted and face-to-face activities resumed, the limitations of using Slack-based evaluation to capture student engagement became increasingly evident. The authors observed a notable decrease in the total number of Slack posts: while there were 1,042 posts during the online AY2022 workshop, the number dropped to only 164 during the face-to-face AY2024 workshop—an approximate reduction of 84%. This finding suggests that Slack activity alone is no longer sufficient to capture the full scope of student engagement and communication in an in-person environment. However, due to its unique advantages—particularly its ability to provide an accurate understanding of student engagement and to support fairness in assessment and motivation among students—Slack-based evaluation is still desirable for use not only in online but also in face-to-face group work settings. In response, this study experimentally introduces peer assessment as a supplementary evaluation method to address the limitations of Slack-based evaluation in face-to-face settings, and examines its effectiveness.

2. Methodology

2.1 Slack-based evaluation

The Slack platform used in this study is a communication tool that allows users to share text messages and files in real time via a chat-based interface. Slack automatically records the message history and enables users to search and review past communications, making it well-suited for documenting and analyzing interactions in educational settings. In this workshop, each team was assigned a dedicated Slack channel for discussions and sharing deliverables. Students were expected to use these channels to collaborate with their team members and advance their projects. Prior to the start of group work, students were informed that their communication on Slack would be factored into their course grades, and the criteria used for this evaluation were clearly explained.

Our analysis focused on assessing both the quantity and quality of each student's contributions. To measure quantity, we counted the total number of messages posted by each student. For quality, each message was classified into one of three categories—Type A, B, or C—based on its contribution to the project's progress [1]. For example, in the robotics workshop, sharing a new circuit design proposal created in Tinkercad would be classified as Type A. This classification approach provides deep insights into individual student behavior within teams and helps identify potential issues, such as low engagement or difficulty in understanding project requirements. The classification of messages into Type A, B, and C was conducted by two master's-level research assistants under the supervision of the program organizers, following the rules outlined in Table 4 [1].

Table 4: Classification criteria for Slack-based evaluation

Group A	● Words and actions essential to the progress of the group work. ● Words and actions that contribute to the group activities, such as sharing circuits and references regarding their project. ● Expressions of ideas and opinions which directly contribute to or influence the direction of the group project. ● Providing program code. ● Video clips of the work produced by the group.
Group B	● Any comments which cannot be clearly categorized into either Group A or Group C. ● Comments which our assessors judged facilitated the progress of group work, without directly triggering a 'next step' in the project. For example, just expressing encouragement to other group members.
Group C	● Casual conversation unrelated to the project, emoji, and other reactions as well as simple 'yes' 'no' type responses. ● Sharing Zoom links, Google slides, etc. ● Posting presentation materials in the group chat (as all students have access to these already). ● *A post where a large number of photos and videos were shared, but it was determined that the person only took the photos or videos and shared them, rather than creating the content themselves (The second and subsequent postings at the same time are not categorized.).* ● Statements with no identifiable meaning. ● Repetition of a statement the same or very similar to one the student had already made. ● 'Negative' statements which might demotivate the team.

In this study, the evaluation criteria for a specific item emphasized under Group C in the proposed classification rules [1] were modified. This adjustment was made in response to a recurring issue observed in both past and the current workshops: students who shared a large number of photos and videos were often classified as making high-level contributions (Type A), despite the absence of clear evidence that they had completed the associated tasks themselves. Under the previous criteria, such posts were automatically categorized as Type A, which led to potential overestimation of individual contributions.

2.2 Peer assessment

Peer assessment is a widely used method for evaluating student contributions in group work settings [20][21][22][23]. In many cases, peer assessments are structured around questions using the Likert scale. One well-known framework is CATME (Comprehensive Assessment of Team Member Effectiveness), which is designed to assess individual contributions within teams and support effective team functioning [24]. In this study, two reasons led us to develop a customized peer assessment form: (1) the participants were non-native English speakers from Thailand and Japan, raising concerns that the English phrasing and intent of CATME items might not be fully understood; and (2) peer assessment was being introduced on an experimental basis. Therefore, the evaluation items were adapted to match the students' English proficiency.

In the AY2024 workshop, students were asked to assess their own and their teammates' levels of engagement using a four-point Likert scale. The evaluation items were as follows:

(1) Brainstorming ideas for the team concept (i.e., whether the student contributed to generating ideas for enhancing the functionality of the rescue robot),
(2) Discussion during group work (i.e., whether the student actively participated in discussions and contributed to smooth group operations), and
(3) Actual work (i.e., whether the student contributed to tangible tasks such as programming, circuit design, or slide preparation).

Each item was rated on a four-point scale:

4. Contributed a lot.
3. Contributed to some extent.
2. Didn't contribute much.
1. Didn't contribute at all.

The four-point scale was adopted to avoid central tendency bias [25], which can occur when students choose a neutral midpoint in a five- or three-point scale. In addition, to prevent leniency bias [26], which could result in inflated scores for students with low engagement, only instructors were given access to the peer assessment results. This allowed students to provide candid feedback without fear of judgment from their teammates.

Regarding the ethical considerations of this study involving human subjects, due to the use of personal data from participants in an international collaborative program, the study was conducted in compliance with established research ethics guidelines, including ethical standards, codes of conduct, and responsibilities. Specifically, (1) prior consultation was held with program coordinators to obtain permission for conducting the survey and using student data, and (2) students were informed of the research purpose and content, and their consent was obtained before participation in the survey.

2.3 Multiple regression analysis

The relationship between Slack-based evaluation, as assessed by the program organizers, and students' self-perceived engagement levels was investigated. Specifically, a multiple regression analysis was conducted to determine whether the number of posts categorized as Type A, B, and C (based on Slack evaluation) could predict students' peer-assessed scores for idea generation, discussion, and task execution. This analysis aimed to examine whether there is a meaningful relationship between the two types of assessment, and thereby whether combining them provides added value. Multiple regression analysis is a statistical method used to predict or explain a single dependent variable using multiple independent variables. It enables quantitative evaluation of the magnitude, direction, and reliability of the effect of each independent variable on the dependent variable [27][28][29][30].

In this study, the peer assessment items were measured on a four-point Likert scale, and the mean values of the responses were treated as continuous variables. This approach is based on previous studies that support the validity of using average Likert scores in regression and correlation analyses [31][32][33]. To eliminate the potential influence of multicollinearity among the independent variables, the Variance Inflation Factor (VIF) for each explanatory variable was calculated prior to conducting the regression analysis. All VIF values were confirmed to be below 10, indicating that multicollinearity was not a significant concern in the model. Both the VIF calculations and the multiple regression analyses were conducted using the Analysis Toolpak in Microsoft Excel.

Table 5: Variables of the study

Variable	Variable Code	Data (Table 14 in Appx. 2)
Independent variable		
Number of posts classified as group A	A	Column 3
Number of posts classified as group B	B	Column 4
Number of posts classified as group C	C	Column 5
Dependent variable		
Evaluation of ‘Brainstorming ideas for the team concept’	Idea	Column 7
Evaluation of ‘Discussion during group work’	Discussion	Column 8
Evaluation of ‘Actual work’	Work	Column 9

* For example, in ‘Brainstorming ideas for the team concept’ in Appendix 1, Table 13, Student G gave oneself a rating of 4 and received a rating of 4 from Student H, 3 from Student I, 3 from Student J, 4 from Student K, and 4 from Student L. Student G's evaluation of ‘Brainstorming ideas for the team concept’ is calculated by

$$\frac{(4+3+3+4+4+4)}{5} = 3.60$$

excluding evaluation of oneself.

3. Results and Discussion

3.1 Results of multiple regression analysis of Slack-based evaluation and peer assessment

First, the relationship between each of the three dependent variables—Idea, Discussion, and Work—and the three explanatory variables—A, B, and C—was examined using a multiple regression model that included all three explanatory variables. When assessing the statistical significance of each variable, any explanatory variable in the initial model with a p-value exceeding the 5% significance level was sequentially removed, and the regression model was re-estimated accordingly. Table 6 presents the descriptive statistics used in this study.

Table 6: Descriptive statistics

Variable Code	Mean	Std. Deviation	N
A	1.91	1.73	23
B	0.391	1.27	23
C	4.83	6.54	23
Idea	3.65	0.282	23
Discussion	3.70	0.275	23
Work	3.75	0.238	23

As a result, for the dependent variable *Idea* with explanatory variables A, B, and C, the coefficient of determination (R^2) was 0.249, the adjusted R^2 was 0.130, and the significance F value from ANOVA was 0.134. The p-values for the partial regression coefficients were 0.0673 for A, 0.961 for B, and 0.554 for C. Since the p-values for B and C greatly exceeded the 0.05 significance level, they were not considered good predictors and were therefore removed from the model. A revised regression analysis excluding B and C was conducted, and the results are presented in Section 3.2.
For the dependent variable *Discussion* with explanatory variables A, B, and C, the R^2 was 0.119 and the adjusted R^2 was -0.0204. The significance F value was 0.482, and the p-values for the partial regression coefficients were 0.157 for A, 0.693 for B, and 0.697 for C. Since the p-values for all three variables were well above 0.05, none were considered effective predictors. It was concluded that there was no significant relationship between the explanatory variables A, B, C and the dependent variable *Discussion*.

For the dependent variable *Work* with explanatory variables A, B, and C, the R^2 was 0.654 and the adjusted R^2 was 0.600. The significance F value was 0.000125, and the p-values for the partial regression coefficients were 0.00531 for A, 0.566 for B, and 0.0000436 for C. Since the p-value for B exceeded the 0.05 threshold, B was not considered a good predictor and was removed from the model. A revised regression analysis excluding B was conducted, and the results are presented in Section 3.3.

3.2 Results of multiple regression analysis of explanatory variable 'A' and objective variable 'Idea'

The results of the multiple regression analysis are presented in Table 8. The R^2 was 0.233, indicating that variable A explains approximately 23.3% of the variance in *Idea*. The adjusted R^2 was 0.197, which is relatively high, suggesting that the model is not significantly affected by overfitting. In addition, the F-value from the ANOVA test, which indicates the overall significance of the model, was 0.0196. This result was statistically significant at the 5% level, indicating that the model as a whole significantly predicts *Idea*. Furthermore, the p-value for the partial regression coefficient of variable A was also 0.0196, demonstrating that A is a significant predictor of *Idea* even when considered alone. The corresponding t-value was 2.53, which exceeds the critical value for a two-tailed test at the 5% significance level ($|t_{0.025}(23)| = 2.069$), further supporting the statistical significance of this result. In addition, the standardized regression coefficient for variable A was 0.483, indicating that A has a moderate positive effect on the dependent variable *Idea*. Thus, it can be concluded that variable A has a statistically significant influence on *Idea*. The standardized regression coefficient was calculated using the standard deviations shown in Table 6 and the unstandardized regression coefficient from Table 8, based on the following formula:

$$SPRC = PRC \times \frac{SDEV}{SDRV}$$

Table 7: Calculation of standardized partial regression coefficients

	Meaning
SPRC	Standardized partial regression coefficient
PRC	Partial regression coefficient
SDEV	Standard deviation of explanatory variable
SDRV	Standard deviation of response variable

Table 8: Results of multiple regression analysis of explanatory variable ‘A’ and objective variable ‘Idea’

Regression Statistics	
Multiple R	0.483
R Square	0.233
Adjusted R Square	0.197
Standard Error	0.253
Observations	23

ANOVA

	df	SS	MS	F	Significance F
Regression	1	0.409	0.409	6.39	0.0196
Residual	21	1.35	0.0641		
Total	22	1.75			

	Coefficients	Std. Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	3.49	0.0797	43.9	3.86E-22	3.33	3.66
A	0.0789	0.0312	2.53	0.0196	0.0140	0.144

3.3 Results of multiple regression analysis for explanatory variables ‘A’, ‘C’ and objective variable ‘Work’

The results of the multiple regression analysis are presented in Table 9. The R^2 was 0.648, indicating that variables A and C together explain approximately 64.8% of the variance in *Work*. The adjusted R^2 was also high at 0.613, suggesting that the model possesses substantial explanatory power. The ANOVA revealed a significance probability of 0.0000293, which is well below the 1% level, confirming that the model as a whole is statistically significant. Regarding the individual coefficients, the p-value for variable A was 0.000225, and that for variable C was 0.0000224—both statistically significant at the 5% level. This indicates that both A and C are significant predictors of *Work*. The t-values were 4.49 for A and -5.49 for C, both of which substantially exceed the critical value for significance at the 5% level ($|t_{0.025}(23)| = 2.069$), further supporting the reliability of the results. Moreover, the standardized regression coefficients indicate that A has a moderate to strong positive effect on *Work*. In contrast, the coefficient for C was -0.789, suggesting a strong negative impact. These findings demonstrate that both explanatory variables A and C have statistically significant and substantial effects on the dependent variable *Work*.

Table 9: Results of multiple regression analysis for explanatory variables 'A', 'C' and objective variable 'Work'

Regression Statistics	
Multiple R	0.805
R Square	0.648
Adjusted R Square	0.613
Standard Error	0.148
Observations	23

ANOVA

	df	SS	MS	F	Significance F
Regression	2	0.808	0.404	18.4	2.93E-05
Residual	20	0.439	0.0220		
Total	22	1.25			

	Coefficients	Std. Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	3.72	0.0479	77.7	2.74E-26	3.62	3.82
A	0.0888	0.0198	4.49	0.000225	0.0475	0.130
C	-0.0287	0.00523	-5.49	2.24E-05	-0.0397	-0.0178

3.4 Discussion

In the relationship between Type A in the Slack-based evaluation and *Idea* in the peer assessment, it was found that A had a moderate positive effect on *Idea*. This suggests that students who contributed more idea proposals classified as Type A may have received higher scores in the peer assessments.

On the other hand, no significant relationship was observed between the Slack-based evaluation types A, B, and C and *Discussion* in the peer assessment. This result indicates that the Slack-based evaluation may not be an appropriate indicator for measuring contributions to discussion in a face-to-face environment. A possible reason is that most of the discussions occurred orally and were therefore not recorded on Slack. In fact, a review of the Slack logs from the AY2024 workshop revealed very few posts corresponding to discussion activities.

Regarding the relationship between Types A and C in the Slack-based evaluation and *Work* in the peer assessment, A showed a moderate to strong positive effect, while C showed a strong negative effect. This suggests that students who engaged in and posted work-related activities classified as Type A were rated more highly in peer assessments. In contrast, students who contributed little to the actual tasks and instead made low-contribution posts—such as simple reactions to deliverables shared by other team members—were classified as Type C and tended to receive lower scores in peer assessments. In other words, the Slack-based evaluation in face-to-face programs appears to be an effective indicator for capturing the level of student engagement.

Based on these results, a correlation was observed between Slack activity scores and peer assessment scores—excluding *Discussion*—suggesting that the Slack-based evaluation method remains effective even in face-to-face program settings (RQ1). At the same time, the limitations of Slack-based evaluation also became apparent. In particular, the challenge in face-to-face environments is that discussion content is less likely to be recorded on Slack. However, the peer assessment results confirmed that actual discussions did take place, even when they were not documented on Slack. By incorporating peer assessment, it becomes possible to visualize such unrecorded activities as well (RQ2). To address the limitations of Slack-based evaluation, the authors are currently working on integrating real-time audio recording and automatic speech recognition alongside peer assessment.

4. Conclusion

In this study, peer assessment was experimentally introduced as a method to complement the limitations of Slack-based evaluation in face-to-face group work, and its effectiveness was examined. Focusing on a global PBL workshop conducted in AY2024, a multiple regression analysis was conducted to explore the relationship between activity scores derived from Slack posts and scores from student peer assessments.

The results showed that the number of Type A posts in the Slack-based evaluation had a significant positive correlation with the *Idea* score in peer assessments. Furthermore, the number of Type A and Type C posts was also significantly related to the *Work* score; Type A had a moderate to strong positive effect, while Type C had a strong negative effect. On the other hand, no significant relationship was observed between the Slack-based variables and the *Discussion* score. This is likely because, in a face-to-face environment, much of the discussion occurs orally and is not recorded on Slack.

These findings indicate that Slack-based evaluation is moderately effective in capturing contributions such as idea generation and hands-on work but has limitations when it comes to evaluating discussions that primarily take place through spoken interaction. The results also suggest that incorporating peer assessment makes it possible to visualize activities that are not recorded on Slack, thereby enabling a more comprehensive understanding of student engagement.

As a future direction, the study aims to develop a framework that combines Slack-based post analysis and peer assessment with real-time audio recording and automatic speech recognition, to objectively visualize verbal discussion in group work. Additionally, the development of peer assessment items aligned with participants' CEFR levels, based on CATME criteria, is also under consideration. This study proposes an approach toward fair and multifaceted student evaluation in group work, applicable in both face-to-face and online learning environments.

Acknowledgement
This research is supported by JSPS Research Grants 24K06133 and the 2024 SIT's Grant for Educational Reform and Research Activity.

References

[1] H. Yoshikubo, S. Nagasawa, and H. Ishizaki, "Creating Innovation for Interdisciplinary Robotics Workshops: Solving Issues in the Online Project-Based Learnings in Engineering Education," in *Proc. 2023 ASEE Annual Conf. and Exposition*, Baltimore, MD, USA, Jun. 25–28, 2023.
[2] Y. Iwata, L. Kimura, H. Yoshikubo, and S. Nagasawa, "Enhancing Student Participation in Online Global Project-Based Learnings (gPBLs) Through a Slack-Based Evaluation: A Student Perspective," in *Proc. 2024 ASEE Annual Conf. and Exposition*, Portland, OR, USA, Jun. 23–26, 2024.
[3] J. W. Thomas, *A Review of Research on Project-Based Learning*. 2000.
[4] J. S. Krajcik and P. C. Blumenfeld, "Project-Based Learning," in *The Cambridge Handbook of the Learning Sciences*, R. K. Sawyer, Ed. Cambridge, U.K.: Cambridge Univ. Press, 2005, pp. 317–334.
[5] M. Frank, I. Lavy, and D. Elata, "Implementing the Project-Based Learning Approach in an Academic Engineering Course," *Int. J. Technol. Des. Educ.*, vol. 13, pp. 273–288, 2003.
[6] B. Gray, G. M. Fragoso-Diaz, O. Ogunrinde, and T. Rider, "Implementation of a Project-Based Learning Approach in an Upper Level Course in Engineering Technology," in *Proc. 2024 ASEE Annual Conf. and Exposition*, Portland, OR, USA, Jun. 2024.
[7] N. Avsheniuk, O. Lutsenko, O. Lutsenko, N. Seminikhyna, and T. Svyrydiuk, "Fostering Intercultural Communicative Competence and Student Autonomy through Project-Based Learning," *Arab World English Journal*, Special Issue on Communication and Language in

Virtual Spaces, pp. 130–143, Jan. 2023.
[8] T. Fortune, S. Borkovic, A. Bhopti, R. Somoza, H. C. Nhan, and S. Rangwala, “Transformative Learning Through International Project-Based Learning in the Global South: Applying a Students-as-Partners Lens to a ‘High-Impact’ Capstone,” *J. Stud. Int. Educ.*, vol. 23, no. 1, pp. 49–65, Nov. 2018.
[9] H. Ishizaki, S. Nagasawa, H. Yoshikubo, and H. Nakamura, “Affordable robotics toolkits for equitable and interdisciplinary education, transformable to searching nodes for disaster onsite investigations,” in *Proc. 2023 ASEE Annual Conf. and Exposition*, Baltimore, MD, USA, Jun. 25–28, 2023.
[10] W. Sumarni, “The Strengths and Weaknesses of the Implementation of Project-Based Learning: A Review,” *Int. J. Sci. Res.*, vol. 4, no. 3, pp. 478–484, 2015.
[11] J. Chen, A. Kolmos, and X. Du, “Forms of Implementation and Challenges of PBL in Engineering Education: A Review of Literature,” *Eur. J. Eng. Educ.*, vol. 46, no. 1, pp. 90–115, 2021.
[12] P. P. Srinivasa, N. C. Niranjan, and B. R. Shrinivasa, “Project-Based Learning (PBL): Issues Faced by Faculty for Its Effective Implementation,” *J. Eng. Educ. Transform.*, vol. 31, no. 3, pp. 9–16, Jan. 2018.
[13] T. Yokemura and M. Inoue, “A Method to Solve PBL Issues and to Improve Project Management Competencies of Students,” *Manag. Stud.*, vol. 6, no. 3, pp. 147–166, May-Jun. 2018.
[14] M. L. Miville, P. Holloway, C. Gelso, R. Pannu, et al., “Appreciating Similarities and Valuing Differences: The Miville-Guzman Universality-Diversity Scale,” *J. Counseling Psychol.*, vol. 46, no. 3, pp. 291–307, Jul. 1999.
[15] J. N. Fuertes, M. L. Miville, J. J. Mohr, W. E. Sedlacek, and D. Gretchen, “Factor Structure and Short-form of the Miville-Guzman Universality-Diversity Scale,” *Meas. Eval. Counseling Dev.*, vol. 33, no. 3, pp. 157–169, 2000.
[16] S. Müller, "How Slack Facilitates Communication and Collaboration in Seminars and Project-Based Courses," *Journal of Educational Technology Systems*, vol. 51, no. 3, pp. 1-14, Jan. 2023. doi: 10.1177/00472395231151910.
[17] F. R. Kates, S. K. Samuels, J. B. Case, and M. Dujowich, "Lessons Learned from a Pilot Study Implementing a Team-Based Messaging Application (Slack) to Improve Communication and Teamwork in Veterinary Medical Education," *Journal of Veterinary Medical Education*, vol. 47, no. 1, pp. 18-26, Feb. 2020. doi: 10.3138/jvme.0717-091r2.
[18] W. M. Davies, “Groupwork as a form of assessment: Common problems and recommended solutions,” High. Educ., vol. 58, pp. 563–584, 2009. doi: 10.1007/s10734-009-9216-y.
[19] M. R. Fellenz, “Toward fairness in assessing student groupwork: A protocol for peer evaluation of individual contributions,” J. Manag. Educ., vol. 30, no. 4, pp. 570–591, 2006.

doi: 10.1177/1052562906286713.
[20] M. Freeman, “Peer assessment by groups of group work,” Assess. Eval. High. Educ., vol. 20, no. 3, pp. 289–300, 1995. doi: 10.1080/0260293950200305.
[21] S. Clarke and M. Blissenden, “Assessing student group work: Is there a right way to do it?,” Law Teach., vol. 47, no. 3, pp. 368–381, 2013. doi: 10.1080/03069400.2013.851340.
[22] F. Kilickaya, “Peer assessment of group members in tertiary contexts,” in Innovations in Languages for Specific Purposes - Present Challenges and Future Promises, M. Sowa and J. Krajka, Eds. Frankfurt am Main: Peter Lang, 2017, pp. 329–343.
[23] S. Dutta et al., “Enhancing students’ engagement and learning through peer assessment in group projects,” J. Educ. Res. Rev., vol. 11, no. 6, pp. 93–104, Oct. 2023. doi: 10.33495/jerr_v11i6.23.120.
[24] C. P. Pung and J. Farris, “Assessment of the CATME peer evaluation tool effectiveness,” in Proc. 2011 ASEE Annu. Conf. Expo., Vancouver, BC, Jun. 2011. doi: 10.18260/1-2--17542.
[25] I. Douven, "A Bayesian perspective on Likert scales and central tendency," Psychonomic Bulletin & Review, vol. 25, pp. 1203–1211, 2018. doi: 10.3758/s13423-017-1344-2.
[26] R. Tandiono and A. Limijaya, "Understanding Rater Bias in Self and Peer Assessment among University Students: A Cultural Psychology Perspective," Asia-Pacific Education Researcher, 2025. doi: 10.1007/s40299-024-00967-7.
[27] Z. Zhiquan and J. Jieyin, "Multiple Regression Analysis of Class Teacher’s Effect on Students 'Incremental' of Academic Performance," *American Research Journal of Humanities and Social Sciences*, vol. 1, no. 5, pp. 1–9, 2015.
[28] Z. Zekarias, N. Aba-Milki, and F. Mikre, "Predictors of academic achievement for first year students: The case of Wolaita-Soddo University, Ethiopia," *European Scientific Journal*, vol. 11, no. 28, pp. 63–76, Oct. 2015.
[29] A. Bora and S. Ahmed, "Parents' socioeconomic status and pupils' mathematics achievement: Stepwise multiple regression analysis approach," *International Journal of Technical Innovation in Modern Engineering & Science*, vol. 4, no. 11, pp. 316–322, Nov. 2018.
[30] N. S. M. Pazil and N. Mahmud, "Multiple regression in determining affecting factors student success in a statistics subject," *International Journal of Academic Research in Progressive Education and Development*, vol. 12, no. 3, pp. 492–499, 2023.
[31] J. S. Mueller, “Why individuals in larger teams perform worse,” *Organizational Behavior and Human Decision Processes*, vol. 117, no. 1, pp. 111–124, 2012.
[32] P. Rylander, A. Heden, T. Archer, and D. Garcia, “Will the peer leader please stand up? The personality of the peer leader in elite and non-elite sport teams,” *International Journal of Research Studies in Psychology*, vol. 3, pp. 65–74, 2014, doi: 10.5861/ijrsp.2013.451.
[33] C. C. Martin and K. D. Locke, “What do peer evaluations represent? A study of rater consensus and target personality,” *Frontiers in Education*, vol. 7, 2022.

Appendix 1

All peer assessment results collected during the AY2024 workshop are presented below. For example, in Table 1 under *Idea*, Student A rated themselves with a score of 3, and gave scores of 3 to Students B and C, 4 to Student D, and 3 to both Students E and F. Likewise, Student A received scores of 4 from Student B, 4 from Student C, 3 from Student D, and 2 from Student F. In other words, the diagonal cells (highlighted with bold borders) in the table represent the self-assessments of each student.

Table 10: Results of Team 1 Peer Assessment

* The questionnaire response from Student E was not obtained.

[Brainstorming ideas for the team concept]

Evaluatee / Evaluator	Student A	Student B	Student C	Student D	Student E	Student F
Student A	**3**	**3**	**3**	**4**	**3**	**3**
Student B	**4**	**2**	**3**	**3**	**3**	**3**
Student C	**4**	**4**	**4**	**4**	**4**	**4**
Student D	**3**	**3**	**4**	**4**	**4**	**4**
Student E						
Student F	**2**	**2**	**3**	**4**	**3**	**3**
Average excluding evaluation of oneself	**3.25**	**3.00**	**3.25**	**3.75**	**3.40**	**3.50**

[Discussion during group work]

Evaluatee / Evaluator	Student A	Student B	Student C	Student D	Student E	Student F
Student A	**4**	**3**	**4**	**4**	**4**	**4**
Student B	**4**	**3**	**3**	**3**	**3**	**3**
Student C	**4**	**4**	**4**	**4**	**4**	**4**
Student D	**3**	**3**	**3**	**4**	**3**	**3**
Student E						
Student F	**2**	**2**	**3**	**4**	**3**	**3**
Average excluding evaluation of oneself	**3.25**	**3.00**	**3.25**	**3.75**	**3.40**	**3.50**

[Actual work]

Evaluator \ Evaluatee	Student A	Student B	Student C	Student D	Student E	Student F
Student A	3	3	4	3	4	4
Student B	4	3	3	4	3	4
Student C	4	4	4	4	4	4
Student D	4	4	4	3	4	4
Student E						
Student F	3	3	3	4	3	3
Average excluding evaluation of oneself	3.75	3.50	3.50	3.75	3.60	4.00

Table 11: Results of Team 2 Peer Assessment

[Brainstorming ideas for the team concept]

Evaluator \ Evaluatee	Student G	Student H	Student I	Student J	Student K	Student L
Student G	4	4	4	4	4	4
Student H	4	3	4	4	4	4
Student I	3	3	4	3	3	2
Student J	3	4	4	3	3	2
Student K	4	4	4	4	4	4
Student L	4	4	3	4	4	4
Average excluding evaluation of oneself	3.60	3.80	3.80	3.80	3.60	3.20

[Discussion during group work]

Evaluatee / Evaluator	Student G	Student H	Student I	Student J	Student K	Student L
Student G	4	4	4	4	4	4
Student H	4	3	4	4	4	4
Student I	3	4	4	4	3	3
Student J	3	3	3	4	3	3
Student K	4	4	4	4	4	4
Student L	4	4	4	3	4	4
Average excluding evaluation of oneself	3.60	3.80	3.80	3.80	3.60	3.60

[Actual work]

Evaluatee / Evaluator	Student G	Student H	Student I	Student J	Student K	Student L
Student G	3	4	4	4	4	4
Student H	4	3	4	4	4	4
Student I	3	4	3	4	4	3
Student J	4	3	3	3	4	3
Student K	4	4	4	4	4	4
Student L	4	4	4	4	4	4
Average excluding evaluation of oneself	3.80	3.80	3.80	4.00	4.00	3.60

Table 12: Results of Team 3 Peer Assessment

[Brainstorming ideas for the team concept]

Evaluator \ Evaluatee	Student M	Student N	Student O	Student P	Student Q
Student M	4	4	4	4	4
Student N	4	3	4	3	3
Student O	4	4	4	4	4
Student P	4	4	4	4	4
Student Q	4	4	4	4	3
Average excluding evaluation of oneself	4.00	4.00	4.00	3.75	3.75

[Discussion during group work]

Evaluator \ Evaluatee	Student M	Student N	Student O	Student P	Student Q
Student M	4	4	4	4	4
Student N	3	4	4	3	4
Student O	4	4	4	4	4
Student P	4	4	4	4	4
Student Q	4	4	4	4	3
Average excluding evaluation of oneself	3.75	4.00	4.00	3.75	4.00

[Actual work]

Evaluator \ Evaluatee	Student M	Student N	Student O	Student P	Student Q
Student M	3	4	4	4	4
Student N	3	3	4	3	4
Student O	4	4	4	4	4
Student P	4	4	4	4	4
Student Q	4	4	4	4	4
Average excluding evaluation of oneself	3.75	4.00	4.00	3.75	4.00

Table 13: Results of Team 4 Peer Assessment

[Brainstorming ideas for the team concept]

Evaluator \ Evaluatee	Student R	Student S	Student T	Student U	Student V	Student W
Student R	2	4	4	4	3	4
Student S	4	4	2	2	4	3
Student T	4	4	4	4	4	4
Student U	4	4	4	4	4	4
Student V	4	4	4	4	4	3
Student W	4	4	4	4	3	4
Average excluding evaluation of oneself	4.00	4.00	3.60	3.60	3.60	3.60

[Discussion during group work]

Evaluator \ Evaluatee	Student R	Student S	Student T	Student U	Student V	Student W
Student R	3	4	4	4	4	4
Student S	4	4	4	4	4	3
Student T	4	4	4	4	3	3
Student U	4	4	4	4	4	4
Student V	4	4	4	4	3	4
Student W	4	3	4	4	4	3
Average excluding evaluation of oneself	4.00	3.80	4.00	4.00	3.80	3.60

[Actual work]

Evaluatee / Evaluator	Student R	Student S	Student T	Student U	Student V	Student W
Student R	3	4	4	4	4	4
Student S	4	3	4	4	3	2
Student T	4	4	4	4	4	3
Student U	4	4	4	4	3	3
Student V	4	4	3	3	3	3
Student W	3	4	3	3	4	4
Average excluding evaluation of oneself	3.80	4.00	3.60	3.60	3.60	3.00

Appendix 2

A table summarizing the results of the Slack-based evaluations and peer assessments is presented below.

Table 14: Results of Slack-based evaluations and peer assessments

Col.1	Col.2	Col. 3	Col. 4	Col. 5	Col.6	Col.7	Col.8	Col.9
Group	Full name	Slack-based evaluation * Number of posts classified as Type A, B, or C.				Peer assessment * Average excluding evaluation of oneself.		
		Type A	Type B	Type C	Total Posts	Brainstorming ideas for the team concept	Discussion during group work	Actual work
1	Student A	1	0	3	4	3.25	3.25	3.75
1	Student B	1	0	2	3	3.00	3.00	3.50
1	Student C	1	0	9	10	3.25	3.25	3.50
1	Student D	0	0	1	1	3.75	3.75	3.75
1	Student E	0	0	1	1	3.40	3.40	3.60
1	Student F	0	0	0	0	3.50	3.50	4.00
2	Student G	2	0	1	3	3.60	3.60	3.80
2	Student H	3	0	3	6	3.80	3.80	3.80
2	Student I	3	0	4	7	3.80	3.80	3.80
2	Student J	1	0	2	3	3.80	3.80	4.00
2	Student K	5	0	6	11	3.60	3.60	4.00
2	Student L	0	0	0	0	3.20	3.60	3.60
3	Student M	2	0	1	3	4.00	3.75	3.75
3	Student N	2	0	6	8	4.00	4.00	4.00
3	Student O	3	0	0	3	4.00	4.00	4.00
3	Student P	1	0	0	1	3.75	3.75	3.75
3	Student Q	1	0	0	1	3.75	4.00	4.00
4	Student R	3	1	8	12	4.00	4.00	3.80
4	Student S	7	6	14	27	4.00	3.80	4.00
4	Student T	2	0	8	10	3.60	4.00	3.60
4	Student U	1	0	0	1	3.60	4.00	3.60
4	Student V	4	1	15	20	3.60	3.80	3.60
4	Student W	1	1	27	29	3.60	3.60	3.00