



Predictive Validity of Peer Assessment in Micro-Teaching: Correlation with Teacher Ratings among Indonesian Preservice EFL Teachers

Santi Farmasari^{1,*}, Lalu Ali Wardana¹, La Ode Alfin Haris Munandar¹, Hartati Suryaningsih¹

¹English Education Department, University of Mataram, Nusa Tenggara Barat, Indonesia

Article Info

Article history:

Received Aug 29, 2025
Revised Sep 30, 2025
Accepted Oct 10, 2025
OnlineFirst Oct 31, 2025

Keywords:

Micro-Teaching
Peer Assessment
Predictive Validity
Pre-service Teachers
Teacher Ratings

ABSTRACT

Purpose of the study: This study investigates the predictive validity of peer assessment of teacher evaluations in English micro-teaching performance among preservice teachers.

Methodology: This study used a quantitative correlational-predictive design with 48 preservice teachers selected through random cluster sampling. The study used peer and teacher performance assessment rubrics covering eight teaching skills, which were previously validated by two experts (CVI = 1.0). Data were analyzed using Pearson correlation, linear regression, and paired-sample t-tests to examine predictive validity, alignment, and discrepancies between peer and teacher evaluations in micro-teaching performance.

Main Findings: Data reveal a moderate to strong correlation between peer and teacher scores ($r = 0.645$, $p < 0.001$), with peer assessments significantly predicting teacher evaluations ($R^2 = 0.416$). However, peer scores were consistently lower ($M = 34.02$ vs. 38.33 , $p < 0.001$), particularly in complex areas like classroom management and reinforcement. This highlights peer assessment's value as a supplementary tool for evaluating teaching and fostering reflection, while underscoring the need for assessor training and rubric calibration to ensure reliability.

Novelty/Originality of this study: This study brings a new perspective by exploring whether peer assessment in English micro-teaching can actually predict teacher evaluations. Unlike most research that sees peer review only as a learning aid, this study shows peers can meaningfully mirror teacher judgments, while also revealing where their views differ. The findings highlight the potential of peer assessment as both a learning and an evaluative tool in teacher education.

This is an open access article under the [CC BY](https://creativecommons.org/licenses/by/4.0/) license



Corresponding Author:

Santi Farmasari,
English Education Department, University of Mataram,
Jl. Majapahit No.62, Gomong, Kec. Selaparang, Kota Mataram, Nusa Tenggara Barat, 83115, Indonesia
Email: santifarmasari@unram.ac.id

1. INTRODUCTION

In recent years, the use of peer evaluation in higher education has received increasing attention as educators seek approaches that foster deeper learning, critical thinking, and students' capacity to make informed judgments about quality [1], [2]. Peer evaluation involves students assessing and providing feedback on their peers' work, performance, or contributions based on clearly defined criteria, typically under instructor supervision [3]. When carefully designed and facilitated, peer evaluation extends beyond a simple learning strategy and becomes a powerful tool for cultivating essential professional competencies such as constructive critique, analytical reasoning, and reflective pedagogical practice. Within teacher education programs, peer assessment has a particularly significant role, as it strengthens core competencies required of future educators,

including instructional analysis, evaluation of teaching methods, and the delivery of evidence-based feedback to peers [4], [5].

For pre-service teachers, peer assessment enhances feedback literacy, reflective thinking, and the ability to recognize indicators of effective teaching performance [6]. Through evaluating peers' instructional demonstrations, pre-service teachers are not merely assigning scores but are actively developing professional judgment. Micro-teaching activities that incorporate peer assessment allow preservice teachers to experiment with teaching strategies, practice lesson delivery, and obtain feedback from both instructors and peers. This process deepens understanding of pedagogical decision-making and strengthens reflective practice, as found in previous research highlighting its contribution to the development of instructional competencies [7], [8]. Consequently, peer assessment has become a cornerstone of many micro-teaching courses, providing a safe, formative environment for practicing essential teaching skills before entering real classrooms [9]. Although many studies emphasize the benefits of peer assessment for reflection and professional growth, limited evidence exists regarding the extent to which peer scores can accurately predict teacher judgments of teaching performance.

Peer assessment has also been recognized as an effective pedagogical strategy for promoting engagement, metacognitive development, and accountability across a variety of educational settings, including teacher education. When peer evaluation is structured with clear rubrics, anonymity, assessor training, and calibration procedures, the quality of feedback and reflective learning improves substantially [10]-[13]. Studies further suggest that peer assessments conducted systematically can generate outcomes closely aligned with teacher assessments, especially when supported by guided reflection and scaffolded evaluation tasks [14], [15]. In micro-teaching contexts, peer assessment enables preservice teachers to rehearse instructional strategies, refine classroom management techniques, and receive constructive peer-generated feedback without the pressures of high-stakes evaluation [16], [17]. Additionally, research indicates that micro-teaching peer assessments supported by training, explicit criteria, and supervision can facilitate reflective growth and demonstrate correlations with teaching efficacy in real classrooms [18], suggesting potential for predictive validity.

The concept of predictive validity examining whether an assessment can forecast future performance has become increasingly important in teacher education, where programs aim to identify early indicators of teaching readiness. Evidence shows that well-designed peer assessments can reliably predict subsequent academic performance and increase in validity as students gain more assessment experience, particularly for complex teaching skills [19], [20]. Yet during the Covid-19 pandemic, the shift to digital peer assessment raised concerns about usability, grading fairness, and the quality of feedback [21], [22]. Studies by Cheng and Warren highlight challenges associated with assessor variability, interpersonal dynamics, and potential biases, all of which threaten the predictive accuracy of peer evaluations [3]. These findings underscore the importance of structured training, transparent scoring rubrics, and systematic strategies to mitigate social and relational bias in order to protect the predictive value of peer assessment [23]-[25].

Predictive validity is particularly crucial because when peer evaluations strongly align with teacher ratings, peer assessment may serve not only as a formative learning tool but also as credible evidence of teaching competence [10][11]. This raises an important question: to what extent can peer scores meaningfully reflect or anticipate teacher evaluations? If peer assessment can reliably predict teacher judgments, it may hold substantial value for performance-based evaluation in teacher education, enabling programs to identify teaching strengths earlier and support professional development more effectively [4][12]. This issue is especially relevant in micro-teaching settings where authentic teaching practice and feedback play a central role in developing preservice teacher expertise.

Despite substantial scholarly attention, several gaps remain. First, although many studies explore the benefits of peer assessment, few have directly examined its predictive validity specifically, whether peer scores can statistically forecast teacher evaluations in micro-teaching contexts. Second, existing studies often discuss the alignment between peer and teacher assessments but rarely employ rigorous quantitative methods (e.g., regression models) to test predictive relationships. Third, empirical research comparing score discrepancies between peer assessors and instructors remains limited, even though such discrepancies can provide insight into judgment accuracy, assessor bias, and the reliability of peer-generated evaluations. Fourth, within the context of micro-teaching, evidence is particularly sparse regarding the conditions under which peer assessments function as valid indicators of teaching performance, particularly when influenced by training, rubrics, or social dynamics. These gaps highlight the need for empirical investigations that determine the extent to which peer evaluations can predict teacher assessments and identify factors that shape their alignment or divergence.

Addressing these gaps, the present study aims to investigate whether peer assessment scores in micro-teaching sessions can significantly predict teacher evaluation scores among preservice teachers. Using quantitative methods including correlational and regression analyses the study provides empirical evidence on the predictive validity of peer assessment and examines discrepancies between peer and teacher ratings. The findings are intended to inform assessment practices, enhance instructional strategies, and contribute to policy

decisions regarding the role of peer assessment in performance-based evaluation within teacher education programs.

2. RESEARCH METHOD

This study employs a quantitative, correlational-predictive design aimed at assessing the efficacy of peer-assessment scores in forecasting teacher-assigned ratings within a micro-teaching context. A linear regression model was utilised to analyse the predicted relationship between peer and teacher assessment, enabling the assessment of correlations among variables [26]. Pearson correlation and simple linear regression analysis were utilised to assess the degree and direction of the association between the two score sets [27], [28].

The population of this study included 192 students enrolled in the English Education Department at a public university in Indonesia. All of them were taking a Micro-Teaching course during the academic year. From this population, three intact classes were randomly selected using a cluster sampling technique. This approach ensured that the sample reflected naturally existing groups rather than individuals selected in isolation. The final sample consisted of 48 preservice teachers, aged between 19 and 22 years (Table 1), with a relatively balanced distribution of male and female students across the three classes. Participation was voluntary, and informed consent was obtained from all students.

Table 1. Research sample

Class	Students		Age
	Male	Female	
Class A	7	9	20-22
Class B	6	10	19-22
Class C	6	10	20-21

The data were collected throughout 16-week period or one course term during designated micro teaching sessions. A peer-assessment and teacher-assessment rubric were used to assess micro teaching performance. The rubric contains eight mandated-core teaching skills: opening a class, questioning, explaining, reinforcement, variation skills, small-group discussion, classroom management, and closing a lesson. Each dimension was evaluated using a 5-point Likert scale (1 = Very Poor to 5 = Excellent). Students performed their micro-teaching lessons in small groups. Following each class, the teacher and a student individually filled out the micro-teaching assessment rubric. A simulation of assessment using the rubric was conducted prior to the actual students' micro-teaching performance to provide actual assessment experience to students. Feedback was given to students on the processes and the output of the peer-assessment. The peer scores for each student were aggregated to provide a composite peer score. Both peer and teacher evaluations were documented and correlated for analysis.

The assessment rubrics were validated by two faculty members proficient in teaching methodologies and educational assessment using a four-scale (1 = not relevant to 4 = very relevant) content-validity rubric. The results (Table 2) showed that the item-level CVI (I-CVI) was 1.0, while the overall scale-level CVI (S-CVI/Ave) reached 1.0. This score is well above the commonly accepted threshold of 0.80, indicating that the instrument had strong content validity [29], [30]. Based on the experts' feedback, several minor revisions were made, that were simplifying technical terms for clarity and removing one redundant item. These adjustments ensured that the final version of the instrument was both clear and appropriate for the intended research context.

Table 2. Content validity index of the assessment rubric

Item No.	Expert 1	Expert 2	I-CVI
1	4	4	1.0
2	3	3	1.0
3	4	3	1.0
4	3	3	1.0
5	4	4	1.0
6	3	4	1.0
7	4	4	1.0
8	4	4	1.0
S-CVI (Ave)			1.0

The study focused on two main variables: peer-assessment scores and teacher-assessment scores. Both were measured using a rubric that covered eight essential teaching skills: opening a class, questioning, explaining, reinforcement, variation, small-group discussion, classroom management, and closing a lesson. Each

skill was rated on a 5-point Likert scale, ranging from 1 (very poor) to 5 (excellent). Peer scores were aggregated from individual ratings provided by classmates, while teacher scores were assigned independently by the course instructor. The reliability analysis (Table 3) indicated that both the teacher assessment instrument ($\alpha = 0.88$) and the peer assessment instrument ($\alpha = 0.88$) had strong internal consistency, signifying that the items within each scale consistently measure the construct. This indicates that all components are cohesively matched and complementary, and the instruments are dependable for evaluating teaching performance from both teacher and peer viewpoints.

Table 3. Cronbach's Alpha

Peer-assessment rubric		Teacher-assessment rubric	
Cronbach's Alpha	N of items	Cronbach's Alpha	N of items
.88	8	.88	8

The collected data were analyzed using IBM SPSS Statistics (Version 27). The analysis followed three main steps. First, descriptive statistics were used to summarize the distribution of peer and teacher scores. Second, Pearson's correlation coefficient (r) was applied to examine the strength and direction of the relationship between the two sets of scores. Third, simple linear regression was conducted to test whether peer scores could significantly predict teacher scores. The regression model displays the beta coefficient (β), significance level (p -value), and coefficient of determination (R^2) to assess predictive validity. In addition, a paired-sample t -test was used to compare mean differences between peer and teacher evaluations. These methods were chosen because they are widely used for examining predictive validity in educational research and provide clear evidence of both alignment and divergence between variables.

This work adheres to recognised ethical research standards. Ethical approval was secured from the Board of Research and Community Services at the affiliated university. Informed consents were obtained from all participants, guaranteeing their complete awareness of the study's goal, data confidentiality, and their right to participate willingly without academic repercussions. Participants' identities were kept confidential using codes and the data were securely maintained and were only accessible to the research team. Participation or lack thereof did not affect the students' academic assessment or grades.

3. RESULTS AND DISCUSSION

Table 4 below presents a comparison of peer and teacher assessment of students' micro teaching performance. The data show that the teacher scores ($M = 38.33$) surpass peer scores ($M = 34.02$) by almost 4.3 points which demonstrate a significant disparity. In relation to consistency of the scores, peer assessors demonstrate greater variability, evidenced by a higher standard deviation ($SD = 6.67$) in contrast to the teachers' ($SD = 4.18$). In addition, the peer scores fluctuate between 20 and 47, and the teacher scores vary from 25 to 46.

Table 4: Descriptive Statistics of Peer and Teacher Scores

	Peer Scores	Teacher Scores
Mean	34.02	38.33
SD	6.67	4.18
Min	20	25
Max	47	46

The scores above indicate that the teachers were generally more lenient or supportive in their assessments than the peers, who may be more stringent or reluctant to award better grades. Teachers, equipped with their professional judgement and a greater comprehension of evaluation criteria, may be more adept at identifying student effort or potential that peers would neglect. Besides, the broader variation in peer assessment may indicate discrepancies in the interpretation of the scoring criteria, individual biases, or insufficient assessment experience. The consistent grades from teachers indicate a mutual comprehension of standards and a more refined grading methodology. The wider spectrum of peer evaluations supports the notion that peers may assess more critically or erratically. The lowest peer score of 20 may signify an anomalous assessment, either resulting from personal interests, misinterpretation of performance, or emotional reactions. To obtain a more detailed understanding of the divergence of the peer and teacher evaluations, Table 5 and Figure 1 below show the mean differences in scores across individual rubric criteria. The data identify specific areas where peer-assessment evaluations differed from those of the teacher.

Table 5. Mean Difference by Criteria

Criteria	Mean Difference
Opening Skill	-0.02
The Use of Media	-0.11
English Language Accuracy & Communication	-0.25
Closing Skills	-0.30
Small Group Discussion Skill	-0.48
Explaining Skill	-0.51
Questioning Skill	-0.59
Variation Skill	-0.63
Reinforcement Skill	-0.65
Classroom Management Skill	-0.77

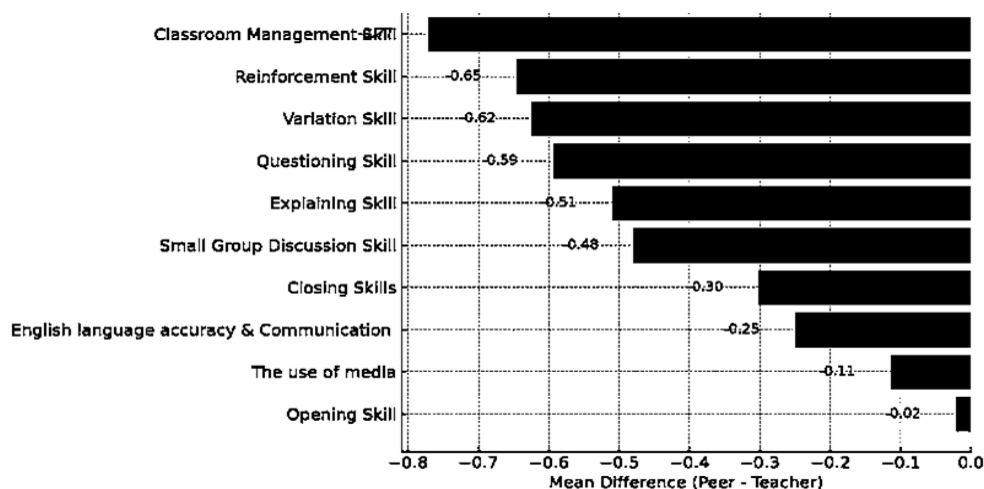


Figure 1. Visualization of the mean differences in scores by criterion

The data indicate that peers consistently assigned lower marks than the teacher across nearly all assessment parameters, although the magnitude of the disparities varied. The most significant deficiencies were observed in Classroom Management Skill (-0.77), Reinforcement Skill (-0.65), and Variation Skill (-0.63). These specific areas frequently entail more intricate and dynamic instructional behaviours, which may provide challenges for observers to evaluate with certainty and precision. The trend of persistently lower peer scores in these areas indicates that preservice teachers may be either more discerning or less confident when assessing instructional components that require more nuanced evaluation. Moderate discrepancies were seen in Questioning Skill (-0.59), Explaining Skill (-0.51), and Small Group Discussion Skill (-0.48). These skills encompass interactive components of classroom practice, highlighting the trend that more advanced pedagogical competence received less favourable evaluations from peers.

In contrast, criteria such as Opening Skill (-0.02), The Use of Media (-0.11), and English Language Accuracy & Communication (-0.25) showed only minor differences between peer and teacher scores, suggesting that when the performance elements were more concrete or visible, such as how lessons were introduced or which materials were used, peer evaluators felt more confident and aligned more closely with the teacher judgments. These findings highlight that while peer-assessments can approximate teacher evaluations in many respects, certain aspects of teaching performance remain difficult for pre-service teachers to assess consistently. Further, to provide the relationship between peer and teacher assessments, an inferential analysis was conducted as presented in Table 6 below. The Pearson correlation coefficient indicate a moderate to strong positive association ($r = 0.645$, $p < 0.001$), suggesting that elevated peer evaluation ratings were typically associated with higher teacher-assigned scores.

Table 6. Inferential analysis

Analysis	Statistic/Model	Result
Correlation	Pearson correlation coefficient (r)	0.645 ($p < 0.001$)
Linear Regression	Regression equation	Teacher Score = $24.60 + 0.40 \times \text{Peer Score}$
	R-squared	0.416
	Slope coefficient (β)	0.4036 ($p < 0.001$)
Paired-Sample t-Test	t-statistic	-5.86
	p-value	< 0.001

This finding demonstrates that, although peer evaluations are not identical to teacher assessments, there is a consistent pattern of alignment in how performances were judged. The linear regression analysis further underscored this association, showing that peer scores were a significant predictor of teacher scores, accounting for approximately 41.6% of the variance. The regression analysis yielded the equation: $\text{Teacher Score} = 24.60 + 0.40 \times \text{Peer Score}$, signifying that for each additional point awarded by peers, the teacher's score improved by an average of 0.40 points. This corroborates the predictive validity of peer-assessments yet indicates that more than fifty percent of the variance in teacher evaluations is affected by factors beyond peer scores. A paired-sample t-test revealed a statistically significant difference in average scores ($t = -5.86, p < 0.001$), indicating that teachers regularly provided higher ratings than their students which indicates that while peers and instructors typically concurred on relative performance levels, peers were surprisingly more critical in their evaluations. These findings indicate that peer-assessment can serve as a valuable adjunct for assessing teaching competency, provided its limits are acknowledged and it is utilised to enhance, rather than supplant, expert evaluation.

The findings from the correlation, regression, and paired-sample t-test analysis provide significant insights into the relationship between peer and teacher evaluations in micro-teaching course. The Pearson correlation value ($r = 0.645$) demonstrated a moderate to strong positive association, indicating that students with elevated peer scores generally obtained higher grades from their teachers. This degree of alignment indicates that, when directed by systematic rubrics, peer-assessments can identify numerous attributes that trained educators seek in teaching performance [31], [32]. The linear regression analysis reinforced this conclusion by showing that peer scores were a significant predictor of teacher scores, explaining approximately 42% of the variance. The regression equation, $\text{Teacher Score} = 24.60 + 0.40 \times \text{Peer Score}$, demonstrates that peer ratings contributed meaningful information about instructional quality, although more than half of the variability in teacher scores was still accounted for by other factors such as inter rater-related effects and interactions between teacher's knowledge of students' identities and teacher's belief [33]. Besides, the analytical rubric used in this study may also yield higher inter-rater consistency than holistic rubric which potentially influence teacher's stable scoring [19], [34]. This finding aligns with prior studies emphasizing the partial predictive validity of peer-assessments, particularly when rubrics and assessment criteria are clearly communicated [35].

The paired-sample t-test identified a clear and statistically significant difference in overall scoring tendencies. The analysis revealed that teachers consistently gave higher average scores than their students did ($t = -5.86, p < 0.001$), indicating that while peer and teacher evaluations tended to align in terms of relative performance, they differed in their absolute judgments. This shows that preservice teacher students may exhibit reluctance in assigning high ratings, possibly due to ambiguity regarding their evaluative competencies or a wish to evade perceived bias [18]. Conversely, teachers appeared to possess greater confidence in identifying progress, effort, and contextual problems that beginner assessors would neglect [12], [36]. This may also be provoked by teachers' intentions to encourage students' confidence, avoid discouragement so that students' motivation and rapport with students which may be absent in the student assessors' sides.

An examination of the average score disparities for each criterion yielded more insights. Significant disparities were observed in Classroom Management Skill, Reinforcement Skill, and Variation Skill, with peers assigning notably lower ratings compared to teachers. These domains generally encompass dynamic and complex instructional behaviours that may be challenging for less experienced observers to perceive and evaluate appropriately. This indicates that preservice teachers frequently exhibit a deficiency in confidence and calibration on the assessment of more intricate elements of pedagogy [37]. Student assessors usually possess limited pedagogical knowledge particularly when dealing with abstract constructs such as effectiveness of certain teaching techniques. Peer assessors may also be less confident of judging what counts as "good teaching" and may take neutral position when making judgements [24]. Difficulties in interpreting complex rubrics and lack of assessment practice may become other factors that contribute to the peer-assessment process [38]. However, moderate disparities were observed in Questioning Skill, Explaining Skill, and Small Group Discussion Skill, underscoring the idea that peers encounter greater difficulty in evaluating instructional approaches that require profound professional acumen [23]. Conversely, narrower discrepancies appeared in scores for Opening Skill, Use of Media, and Language Accuracy, indicating greater consensus among peers and teachers when assessing more tangible, observable elements of the session.

These findings highlight both the potential and the limitations of peer assessment in micro-teaching contexts. The moderate to strong correlation between peer and teacher scores supports the predictive validity of peer evaluations [19]. While peers did not assign identical ratings to instructors, their scores aligned closely enough to provide credible insights into teaching performance. This reinforces the view that peer assessment is not merely a formative exercise but can serve as a meaningful evaluative tool when supported by structured rubrics and training. The discrepancies observed in complex teaching skills also connect with the framework of assessment literacy [39] [37]. Preservice teachers often lack the confidence and expertise to judge nuanced aspects of pedagogy, such as classroom management or reinforcement. Their lower ratings in these areas may reflect uncertainty rather than strictness. This underscores the importance of developing assessment literacy in teacher education programs [40] so that preservice teachers can move beyond surface-level judgments toward

more accurate, consistent evaluations. Furthermore, the findings align with the concept of evaluative judgment, the ability to discern and apply standards of quality in professional practice [15], [41]. When preservice teachers engage in peer assessment, they begin to internalize criteria for good teaching. However, their variability and lower scoring suggest that evaluative judgment is still developing and requires scaffolding through feedback, exemplars, and guided calibration.

This research provides a distinctive contribution. Prior meta-analyses [42], [43] emphasize the benefits of peer assessment for learning and reflection, but often limited to short of testing its predictive validity. Similarly, much of the literature positions peer review primarily as a formative practice rather than as a valid assessment tool [21], [24]. In contrast, this study empirically demonstrates that peer scores can significantly predict teacher ratings in Indonesian micro-teaching courses. This novelty lies in showing that, with proper rubrics, peer evaluations can serve both formative and evaluative purposes in teacher preparation [44], as well as identifying areas where further training is required to improve accuracy and confidence [45]. Although peer evaluations can yield valuable insights and typically align with professional assessments, they ought to be utilised with teacher evaluations and systematic feedback to present a more comprehensive and equitable representation of teaching competencies particularly ability to make sense of complex teaching practices such as lesson coherence, scaffolding, and the use of feedback, has the potential to improve the value of peer assessment in both formative and summative ways [19]. When teacher education programs provide explicit training and opportunities for practice, preservice teachers gain greater confidence, broaden their understanding of what assessment is for, and produce judgments that are more consistent with those of experienced educators [46]. Thus, the results suggest that teacher education programs should not dismiss peer assessment as informal feedback but recognize its dual role [47]. Incorporating sufficient training, where students practice using rubrics with sample lessons and receive guidance on scoring, could reduce variability and increase alignment with instructor judgments. Such training would strengthen preservice teachers' assessment literacy [48], [49], and prepare them for future responsibilities that involve evaluating student learning [43].

Nevertheless, this study is limited to a relatively small sample and drawn from a single institution, which limits the generalizability of the findings. In addition, the analysis focused only on one semester of data; it did not capture how predictive relationships might evolve as students gain more assessment experience. Finally, the reliance on face-to-face micro-teaching contexts does not reflect how peer assessment might function in online or blended environments. Future studies should replicate this research across multiple institutions and larger samples to strengthen generalizability [50]. Longitudinal research would also be valuable to examine whether peer evaluations become more predictive of teacher ratings as preservice teachers' assessment literacy develops over time [51]. In addition, with the growing use of digital learning platforms, future studies could investigate how peer assessment operates in online micro-teaching contexts and whether technology-mediated feedback improves reliability and predictive validity.

1. CONCLUSION

This study is aimed to explore whether peer-assessment scores in micro-teaching could predict teacher ratings among preservice English teachers in Indonesia. The results show a clear pattern: peer and teacher evaluations were moderately to strongly correlated ($r = 0.645$, $p < 0.001$), and peer scores explained about 41.6% of the variation in teacher ratings. Peers tended to be stricter than instructors, giving consistently lower scores, especially in complex areas like classroom management, reinforcement, and variation. These findings suggest that while peer assessment has strong predictive value, it also comes with limitations that need to be acknowledged. From a theoretical perspective, the study adds to the growing conversation about predictive validity in teacher education. Whereas earlier research has often emphasized the reflective and developmental benefits of peer assessment, this study provides evidence that peer scores can also serve as a credible indicator of teaching performance. It highlights the need to strengthen preservice teachers' assessment literacy and evaluative judgment so they can engage more confidently with the subtler aspects of teaching practice. The practical implications are equally important. Teacher education programs can use peer assessment not only to encourage reflection but also to support evaluation when it is accompanied with clear rubrics and sufficient training. Building these practices into the curriculum can help preservice teachers become more capable assessors and prepare them for the professional demands of evaluating learning in their own classrooms. On a policy level, the findings suggest that peer assessment deserves more serious consideration as part of performance-based evaluation in teacher preparation.

The limitations of the study are on the sample size, a single institution, so the results cannot be generalized too widely. More research is needed across different universities and larger samples, and ideally over longer periods of time to track how peer assessment skills develop with experience. It would also be valuable to investigate peer assessment in digital or blended learning settings, where online platforms may change both the process and the reliability of feedback. Thus, these insights point toward a future where peer assessment plays a

stronger dual role in teacher education, that are as a tool for learning and reflection, and as a meaningful measure of teaching performance.

ACKNOWLEDGEMENTS

This study was conducted under self-funded research through the Research and Community Service Board of the University of Mataram. Therefore, we would like to acknowledge the board that had supported the conduct of the research. We also would like to acknowledge the participation of the micro teaching students who voluntarily contributed to the study.

REFERENCES

- [1] L. Li, Y. Xiong, C. V Hunter, and X. Guo, "Does peer-assessment promote student learning? A meta-analysis," *Assess. Eval. High. Educ.*, vol. 45, no. 2, pp. 193–211, 2020, doi: 10.1080/02602938.2019.1620679.
- [2] E. Panadero, "A review of self-regulated learning: Six models and four directions for research," *Front. Psychol.*, vol. 11, no. 234, 2020, doi: 10.3389/fpsyg.2020.00234.
- [3] W. Cheng and M. Warren, "Enhancing the reliability of peer-assessment through adaptive training: A study in higher education," *Assess. Eval. High. Educ.*, vol. 46, no. 5, pp. 709–724, 2021, doi: 10.1080/02602938.2021.1884842.
- [4] J. Tai, R. Ajjawi, D. Boud, and P. Dawson, "Developing evaluative judgement: A meta-analysis of peer-assessment," *Assess. Eval. High. Educ.*, vol. 46, no. 2, pp. 187–203, 2021, doi: 10.1080/02602938.2021.1875865.
- [5] M. Alqassab, J. W. Strijbos, E. Panadero, J. F. Ruiz, M. Warren, and J. To, "A Systematic Review of Peer Assessment Design Elements," *Educational Psychology Review*, vol. 35, no. 1, 2023, doi: 10.1007/s10648-023-09723-7.
- [6] C. N. Prilop, K. E. Weber, F. J. Prins, and M. Kleinknecht, "Connecting feedback to self-efficacy: Receiving and providing peer feedback in teacher education," *Stud. Educ. Eval.*, vol. 70, 2021, doi: 10.1016/j.stueduc.2021.101062.
- [7] M. J. González-Rodríguez, D. M. Perales-Palacios, F. J. Martínez-Segura, "Microteaching as an effective instructional strategy for pre-service teachers: A systematic review," *Int. J. Instr.*, vol. 13, no. 2, pp. 207–222, 2020, doi: 10.29333/iji.2020.13239a.
- [8] E. Panadero and M. Alqassab, "An empirical review of anonymity effects in peer assessment, peer feedback, peer review, peer evaluation and peer grading," *Assess. Eval. High. Educ.*, vol. 44, no. 8, 2019, doi: 10.1080/02602938.2019.1600186.
- [9] T. Nguyen, "Peer feedback in micro-teaching: Implications for pre-service teacher development," *J. Educ. Teach.*, vol. 48, no. 7, pp. 923–937, 2022, doi: 10.1080/02607476.2022.2047385.
- [10] R. Yuan and L. J. Zhang, "Teachers grow with their students': Examining the teacher identity development of EFL teachers in China," *System*, vol. 96, no. 102406, 2021, doi: 10.1016/j.system.2020.102406.
- [11] K. S. Double, J. A. McGrane, and T. N. Hopfenbeck, "The Impact of Peer Assessment on Academic Performance: A Meta-analysis of Control Group Studies," *Educ. Psychol. Rev.*, vol. 32, no. 2, 2020, doi: 10.1007/s10648-019-09510-3.
- [12] E. Panadero, H. Andrade, and S. Brookhart, "Fusing self-regulated learning and formative assessment: A roadmap of where we are, how we got here, and where we are going," *Aust. Educ. Res.*, vol. 48, no. 1, pp. 1–22, 2021, doi: 10.1007/s13384-019-00376-4.
- [13] M. Zhang, F., Schunn, C., Li, W., & Long, "Changes in the reliability and validity of peer assessment across the college years," *Assess. Eval. High. Educ.*, vol. 45, no. 6, pp. 841–854, 2020, doi: 10.1080/02602938.2019.1697166.
- [14] R. Shible and M. Sharma, "Designing effective peer assessment systems in higher education," *Assess. Educ. Princ. Policy Pract.*, vol. 31, no. 1, pp. 67–86, 2024, doi: 10.29333/iji.2023.16149a.
- [15] A. Jongsma, S. Vink, and A. Ros, "Scaffolding peer evaluation to enhance student reflection and learning outcomes," *Teach. High. Educ. Adv. online Publ.*, 2024, doi: 10.1080/13562517.2024.1160458.
- [16] L. T. T. Nguyen, "Peer feedback in higher education: A systematic review," *Assess. Eval. High. Educ.*, vol. 47, no. 4, pp. 510–525, 2022, doi: 10.1080/02602938.2022.2052507.
- [17] N. Liu and D. Carless, "Peer feedback: The learning element of peer-assessment," *Teach. High. Educ.*, vol. 25, no. 1, pp. 1–15, 2020, doi: 10.1080/13562517.2019.1599302.
- [18] S. Lertsakulbunlue and A. Kantiwong, "Evaluating the dependability of peer assessment in project-based learning for pre-clinical students: A generalizability theory approach," *BMC Med. Educ.*, vol. 25, no. Article 260, 2025, doi: 10.1186/s12909-025-06772-0.
- [19] Y. Zhang, "Predictive validity in educational assessment: Current issues and future directions," *Stud. Educ. Eval.*, vol. 73, no. 101161, 2022, doi: 10.1016/j.stueduc.2022.101161.
- [20] C. C. Deneen and H. T. Hoo, "Connecting teacher and student assessment literacy with self-evaluation and peer feedback," *Assess. Eval. High. Educ.*, vol. 48, no. 2, 2023, doi: 10.1080/02602938.2021.1967284.
- [21] R. Craig and D. Kay, "Online peer assessment in higher education: A systematic review," *Comput. Educ.*, vol. 203, no. 104708, 2024, doi: 10.1016/j.compedu.2023.104708.
- [22] S. Taghizadeh Kerman, N. Banihashem and M. Karami, "Online peer feedback in higher education: A synthesis of the literature," *Educ. Inf. Technol.*, vol. 29, pp. 763–813, 2023, doi: 10.1007/s10639-023-12273-8.
- [23] O. Noroozi, H. J. A. Biemans, and M. Mulder, "Factors affecting the quality of peer feedback in online education: A review and implications for practice," *Educ. Res. Rev.*, vol. 36, no. 100423, 2022, doi: 10.1016/j.edurev.2022.100423.
- [24] T. Tanti, D. Darmaji, A. Astalini, D. A. Kurniawan, and M. Iqbal, "Analysis of user responses to the application of web-based assessment on character assessment," *Journal of education technology*, vol. 5, no. 3, pp. 356–364, 2021, doi: 10.23887/jet.v5i3.33590.
- [25] K. Misiejuk and B. Wasson, "Backward evaluation in peer assessment: A scoping review," *Comput. Educ.*, vol. 175, 2021, doi: 10.1016/j.compedu.2021.104319.

- [26] J. W. Creswell and J. D. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* - John W. Creswell, J. David Creswell - Google Books. 2018.
- [27] T. Tanti, W. Utami, D. Deliza, and M. Jahanifar, "Investigation in vocation high school for attitude and motivation students in learning physics subject", *Jor. Eva. Edu*, vol. 6, no. 2, pp. 479-490, 2025, doi: 10.37251/jee.v6i2.1452.
- [28] D. Edelmann, T. F. Móri, and G. J. Székely, "On relationships between the pearson and the distance correlation coefficients," *Stat. Probab. Lett.*, vol. 169, 2021, doi: 10.1016/j.spl.2020.108960.
- [29] F. Polit and C. T. Beck, *Nursing Research: Generating and Assessing Evidence for Nursing Practice*, 11th ed. Lippincott Williams & Wilkins., 2021.
- [30] M. S. B. Yusoff, "ABC of Content Validation and Content Validity Index Calculation," *Educ. Med. J.*, vol. 11, no. 2, 2019, doi: 10.21315/eimj2019.11.2.6.
- [31] Y. Han and Y. Xu, "The development and validation of a rubric for assessing peer feedback in EFL writing," *Assess. Writ.*, vol. 39, no. 14, pp. 14–28, 2019, doi: 10.1016/j.asw.2018.12.002.
- [32] L. van der M. Dasoo, N., & Muller, "Teacher advocacy for the enhancement of professional learning and development in continuous professional teacher development programmes," *South African J. High. Educ.*, vol. 34, no. 4, pp. 45–59, 2020, doi: 10.20853/34-4-3485.
- [33] L. Doyle, P. . Harris, and M. . Easterbrook, "Quality and quantity: How contexts influence the emergence of teacher bias," *Soc. Psychol. Educ.*, vol. 27, pp. 1855–1882, 2024, doi: 10.1007/s11218-023-09882-z.
- [34] D. A. Kurniawan, A. Astalini, D. Darmaji, T. Tanti, and S. Maryani, "Innovative learning: Gender perception of e-module linear equations in mathematics and physics," *Indonesian Journal on Learning and Advanced Education (IJOLAE)*, vol. 4, no. 2, pp. 92-106, 2022, doi: 10.23917/ijolae.v4i2.16610.
- [35] M. Wong and W. C. Liu, "Enhancing pre-service teachers' identity development through basic psychological need satisfaction in coursework," *Teach. Teach. Educ.*, vol. 135, no. 104415, 2024, doi: 10.1016/j.tate.2023.104415.
- [36] S. Kang, M. O'Reilly, G. Lancioni, T. S. Falcomata, J. Sigafos, and Z. Xu, "Comparison of the predictive validity and consistency among preference assessment procedures: A review of the literature," *Research in Developmental Disabilities*, vol. 34, no. 4. 2013, doi: 10.1016/j.ridd.2012.12.021.
- [37] S. Farmasari, "Pre-service EFL teachers' language assessment literacy satisfaction and assessment preparedness," *Int. J. Lang. Educ.*, vol. 7, no. 4, 2023, doi: 10.26858/ijole.v7i4.47445.
- [38] T. T. P. Nguyen and C. D. Nguyen, "An evaluation of researcher–teacher collaboration with practitioners engaged in instructing English language with young learners," *Educ. 3-13*, 2024, doi: 10.1080/03004279.2024.2323648.
- [39] B. Mansouri, K. Molana, and M. Nazari, "The interconnection between second language teachers' language assessment literacy and professional agency: The mediating role of institutional policies," *System*, vol. 103, 2021, doi: 10.1016/j.system.2021.102674.
- [40] M. Barahona and X. Ibaceta-Quijanes, "Chilean EFL student teachers and social justice: ambiguity and uncertainties in understanding their professional pedagogical responsibility," *Teach. Teach. Theory Pract.*, vol. 30, no. 2, 2024, doi: 10.1080/13540602.2022.2062726.
- [41] W. Cheng and M. Warren, "Peer assessment of language learning in higher education: Practices and challenges," *Assess. Eval. High. Educ.*, vol. 46, no. 4, pp. 623–637, 2021, doi: 10.1080/02602938.2020.1819743.
- [42] X. Li and P. Ke, "Becoming feedback literate: A case study of a Chinese University EFL teacher," *Asia-Pacific Educ. Res.*, vol. 33, no. 2, 2024, doi: 10.1007/s40299-023-00743-z.
- [43] E. Caffrey, D. Fuchs, and L. S. Fuchs, "The predictive validity of dynamic assessment: A review," *Journal of Special Education*, vol. 41, no. 4. 2008, doi: 10.1177/0022466907310366.
- [44] K. Sugihartini, N., Hadi, S., Wahyuni, D. S., & Agustini, "Peer-assessment integrated microteaching course with e-learning system," *EAI Endorsed Trans. e-Learning*, 2022, doi: 10.4108/eai.6-10-2022.2327353.
- [45] H. Y. Phuong *et al.*, "The impact of training vietnamese EFL learners for self-assessment on writing performance. , 17(2), 237-258," *Int. J. Instr.*, vol. 17, no. 2, pp. 237–258, 2024, doi: 10.29333/iji.2024.17214a.
- [46] J. Axt, N. Buttrick, and R. Y. Feng, "A Comparative Investigation of the Predictive Validity of Four Indirect Measures of Bias and Prejudice," *Personal. Soc. Psychol. Bull.*, vol. 50, no. 6, 2024, doi: 10.1177/01461672221150229.
- [47] G. Luria and Y. Kalish, "A social network approach to peer assessment: Improving predictive validity," *Hum. Resour. Manage.*, vol. 52, no. 4, 2013, doi: 10.1002/hrm.21541.
- [48] M. Ishak and F. Hussin, "Predictive validity study of sustainable leadership for learning questionnaire," *Int. J. Instr.*, vol. 15, no. 2, 2022, doi: 10.29333/iji.2022.15233a.
- [49] A. G. Greenwald, T. A. Poehlman, E. L. Uhlmann, and M. R. Banaji, "Understanding and using the implicit association test: III. Meta-Analysis of predictive validity," *J. Pers. Soc. Psychol.*, vol. 97, no. 1, 2009, doi: 10.1037/a0015575.
- [50] D. E. Morisky, L. W. Green, and D. M. Levine, "Concurrent and predictive validity of a self-reported measure of medication adherence," *Med. Care*, vol. 24, no. 1, 1986, doi: 10.1097/00005650-198601000-00007.
- [51] B. P. D. Riyanti, A. O. Suryani, C. W. Sandroto, and S. Y. Soeharso, "The construct and predictive validity testing of Indonesian entrepreneurial competence inventory-situational judgment test model," *J. Innov. Entrep.*, vol. 11, no. 1, 2022, doi: 10.1186/s13731-022-00202-x.