Peer Assessment Analysis of Performance Appraisal Using Analytical Rubrics to Improve Critical Thinking Skills

Tuti Iriani[(✉)], Anisah, and Yusrina Luthfiana

Faculty of Engineering, Universitas Negeri Jakarta, Jalan Rawamangun Muka, Jakarta 13220, Indonesia
tutiiriani@unj.ac.id

Abstract. This study aims to analyze the results of peer evaluation by assessing students' ability to compile and analyze objective questions. The method in this study is the descriptive method. The study is conducted in Building Engineering Education Study Program Universitas Negeri Jakarta in the Evaluation Learning course. Each student is assessed by other students in the same group. The assessment is conducted when the student presents their compile and analyze objective questions task. The instrument in this study is an analytical rubric used when conducting peer assessment. The total respondent is 55 students who took the Evaluation Learning course. The analysis data technology used is Confirmatory Factor Analysis (CFA). The results showed that the measurement was declared valid (<0.7) and reliable ($CR = 0.96$). Based on the results of the calculation of Discriminant Validity (AVE), it can be seen that the peer assessment measurement has an AVE value of 0.929, which is quite high. The result of peer assessment of students' critical thinking skills showed a lower category. It can be understood that at the time of peer assessment there are many influencing factors such as lack of experience of students in conducting peer assessments, online assessments, lack of concentration, time, and lack of understanding of the rubric content and misconceptions.

Keywords: Peer assessment · Critical thinking · Confirmatory factor analysis

1 Introduction

Assessment is a key factor in learning that aims to measure the level of competency achievement and improve student learning processes [1]. One of the assessments used is the performance appraisal. According to [2], performance appraisal is an assessment that in its implementation involves students in an activity, which leads students to demonstrate their abilities in the form of processes and products. The performance appraisal is frequently paired with peer assessment and self-assessment [3]. Peer assessment is an assessment made by students based on criteria and standards in the work of their peers for assessment [4]. The advantages of the implementation of peer assessment include

requiring students to think more critically and improve learning skills based on the evaluation results obtained and improve their ability to manage themselves [5]. When the limitations in the implementation of the assessment, including students' distrust, and the results of the assessment are less accurate and less consistent when compared to the assessment given by the teacher [6]. In the implementation of peer assessment, feedback is needed. Feedback contains comments from teachers and peers [7]. In addition, another benefit of feedback increases the depth and breadth of information and comments from multiple perspectives [8]. In the study of [9], most students have the assumption that the assessments made by the teacher are more accurate than the assessments made by their peers. But, according to [10], there is no significant difference in the average score given by peers and teachers. Students' critical thinking skills are still lacking or low because students are not accustomed to being trained with indicators of critical thinking skills and there is still a lack of learning that is applied to empower students' ability to think critically. Several studies on peer assessment have been carried out, including [11] who conducted study with the aim of describing students' abilities in carrying out peer and self-assessment as a benchmark for performance assessment on coordination system materials in High School Class IX. The results of their study show that peer and self-assessment can be used as benchmarks for product performance assessment. Meanwhile, the comparison of teacher assessment and peer assessment shows a similarity of 84%, meaning that students are good at doing peer assessments. [12] in her study found that peer assessment had a positive impact on increasing the class average score on cognitive skills. Therefore, study conducted by [13] found that peer assessment can develop oral communication skills and understanding of concepts. If a previous study has found that peer assessment can be used as a benchmark for performance appraisal and peer assessment has a positive impact, then the formulation of the problem in this study is how the results of peer assessment analysis using analytical rubrics in assessing performance abilities?

Peer assessment is a process by which students use quantitative assessments or to assess the performance of an individual student in his team or in his class, which involves the use of assessment instruments or checklists prior to peer assessment exercises [14–17]. Peer assessment can encourage an individual's responsibility [5, 18], and can provide opportunities for students to reflect on their strengths and weaknesses and to assess, and compare their work with the work of others which enhances students' meta-cognitive perceptions [19–23]. In conducting peer assessment, [24] outline some guidelines for teachers to conduct peer assessment in the classroom. These guidelines are elaborated to be used as guidelines for achieving peer assessment potential. There are 4 guidelines for conducting peer assessment, namely: informing students of the purpose of the assessment, clearly defining assignments, encouraging impartial evaluation of performance or abilities, and ensuring fruitful returns through follow-up assignments. In addition, in conducting peer assessments, students are given several criteria to guide them in assessing the work of their peers. This is also supported by [25] who state that peer assessment involves students with criteria and standards and then applies them to make assessments. One of the assessments applied is the performance appraisal. According to [2], performance appraisal is especially suitable for assessing skills, such as observation skills, hypothesizing, applying concepts, planning and conducting research, etc. Performance

appraisal does not use answer keys in determining scores but uses scoring guidelines in the form of rubrics. In this study, the rubric used is an analytical rubric that has been compiled and has gone through the content validation stage with very decent results. The results show the content validation of 7 criteria measured values > 0.3, which means the instrument is valid. Based on Alpha Cronbach's reliability test obtained a value of 0.800, then the instrument is declared reliable.

2   Research Methods

The method used in this study is descriptive method. This study, conducted in Building Engineering Education Study Program, Universitas Negeri Jakarta. The subjects used are students who took Evaluation Learning course year 2021/2022.

The figure below shows a peer assessment chart, namely student A assesses the performance of B, and student B assesses the performance of C, student C assesses performance of D, student D assesses performance of E, student E assesses performance of A and so on. The assessment of this analytical rubric is filled by peer assessment through observation with zoom meetings. Analytical rubric using the google form. The components of the analytical rubric include 1) ability to arrange grid; 2) ability to compile questions; 3) calculate the Norm-Referenced Assessment (PAN) and Criterion-Referenced Assessment (PAP); 4) calculate validity and reliability; 5) calculate the difficulty level of the question analysis; 6) calculate discriminatory power; and 7) perform item analysis. All of these aspects use a rating scale of 1–4 (Fig. 1).

The data collection technique is carried out by assessing performance using an analytical rubric. Each student is asked to present his final assignment, namely compiling and analyzing objective questions. After that, each student will assess students in their groups. For example, student A evaluates the work of student B, then student B assesses the work of student C and so on. The analytic rubric was given to a friend through a google form. Then, the results of the assessment analysis were assessed by 3 lecturers using the critical thinking skills rubric consisting of 5 indicators with a Likert scale, 4 = very good; 3 = good; 2 = fair; 1 = poor.

Data analysis was performed using Confirmatory Factor Analysis (CFA) statistical techniques, aiming to test the validity and reliability of the measurement results using the LISREL 8.80 Trial. Followed by descriptive analysis on the different test data based on the level of critical thinking.
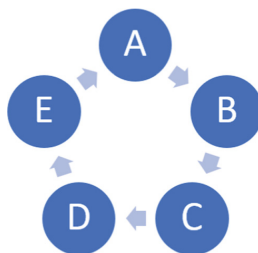

Fig. 1. Peer Assessment Scheme

3 Results and Discussion

Peer assessment is one of the ways to assess the performance of students in their group. Peer assessment is carried out one time after the students have finished their presentation activities. Students assess their group of friends to see their critical thinking skills. The data from the implementation of peer assessment using the analytical rubric can be described as shown in Table 1.

The five critical thinking indicators have a value of >0.7, so they are declared to meet valid (Table 2).

Peer Assessment (PA) which includes 5 indicators, with α (significant level) of 5% (0.05), has a standardize coefficient value or loading factor is declared fixed and has a p-value < 0.05, which is 0.000 so it can be concluded that the PA indicator significantly measures the PA variable (Fig. 2).

Table 1. Exogenous Convergent Validity

Critical Thinking Aspects	Loading Factor Standard
Interpretation	0.942
Analysis	0.830
Evaluation	0.937
Explanation	1.000
Self-Regulation	0.988

Table 2. Standardized Regression Weight from Critical Thinking Value

	Estimate	S.E	C.R.	P	Label
I ← PPA	1				
A ← PPA	0.851	0.088	9.639	***	par_1
E ← PPA	0.094	0.066	14.139	***	par_2
P ← PPA	1.01	0.049	20.524	***	par_3
RD ← PPA	0.997	0.053	18.867	***	par_4

Fig. 2. Peer Assessment Criteria

3.1 Construct Reliability

Construct reliability of peer assessment using a rubric show that the reliability obtained is >0.70 or has a CR value of 0.956, both small-scale and wide-scale tests on a sample of 55 students. The data obtained were analyzed by Confirmatory Factor Analysis (CFA), showing the peer assessment instrument on the performance appraisal of the preparation of the questions through the analysis of the questions, all indicators were reliable.

3.2 Discriminant Validity

Discriminant validity aims to test how far the latent construct is not in harmony with other latent constructs by looking at the Fornell Larcker value. Fornell Larcker is the value of the relationship between variables when the variable is used alone and the variable when the variable is used together with other variables. Invalid constructs are constructs where the value of the relationship between variables with other variables exceeds the value of other variables. Based on the results of the Discriminant Validity (AVE) calculation, it can be seen that in the peer assessment measurement, the AVE value is 0.929, which is quite high.

Based on Table 3, it shows that the suitability test of this model produces a good level of acceptance. Therefore, it can be concluded that peer assessment using analytical rubrics to measure critical thinking is acceptable.

The results of the analysis based on Confirmatory Factor Analysis (CFA) show that the validity and reliability provide precise and reliable results. Meanwhile, the results of the class average score showed low, namely 56. This can be interpreted that the results of observations on this indicator are very varied or "uneven". There are respondents who have very high scores and some other respondents have very low scores. It can be understood that during peer assessment, there are many influencing factors as stated by [26], that the factor of honesty or objectivity in the assessment. The lack of experience of students in conducting peer assessments causes student assessment results to be smaller or larger than the teacher's assessment, besides that the assessment is carried out online so many factors affect the assessment such as concentration, time, and lack of understanding of the content of the rubric. Different interpretations of the indicators or criteria assessed cause students to feel reluctant to participate in the assessment due to ignorance of
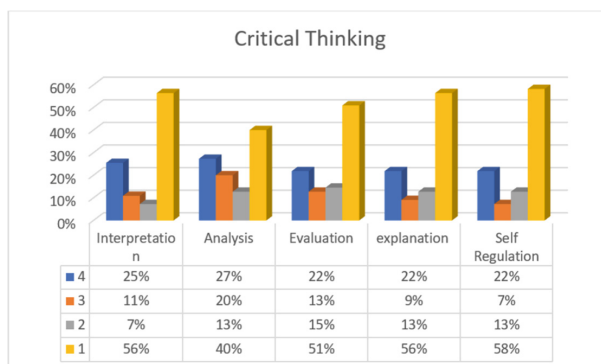
Table 3. Goodness of Fit Index

Criteria	Value	Cut Off	Description
Chi Square	27.415	≤ 54	Good
CMIN/Df	5.483	>2	Good
Probability	0.000	≥ 0.05	Not Good
GFI	0.864	≥ 0.90	Marginal
AGFI	0.591	≥ 0.90	Good
TLI	0.913	≥ 0.90	Good
NFI	0.948	≥ 0.90	Good
IFI	0.957	≥ 0.90	Good
RMSEA	0.288	≤ 0.085	Not Good

the criteria given [27]. In order for the implementation of the peer assessment to be successful as planned, it is necessary to provide explanations related to the assessment criteria/indicators so that students do not hesitate in conducting the assessment. The objectivity of assessment can also lead to differences in student and teacher assessment results. The results of student assessments are strongly influenced by certain feelings towards other students, both positive and negative.

Based on the critical thinking variable, below is the assessment results of peer assessment using critical thinking indicators (Fig. 3).

The graph above shows the results of peer assessment analysis based on critical thinking adopted from [28]: a) Interpretation, namely the ability of a person to understand and express the intent of a situation, data, assessment, rules, procedures, or various criteria. b) Analysis, namely the ability of a person to classify conclusions based on the relationship between information and concepts, with the questions in the problem. c) Evaluation, namely the ability of a person to assess the credibility of a statement or other representation of a person's opinion or judge a conclusion based on the relationship

Critical Thinking

	Interpretation	Analysis	Evaluation	explanation	Self Regulation
4	25%	27%	22%	22%	22%
3	11%	20%	13%	9%	7%
2	7%	13%	15%	13%	13%
1	56%	40%	51%	56%	58%

Fig. 3. Critical Thinking

between information and concepts, with questions that exist in a problem. d) Explanation, namely the ability of a person to state one's reasoning when giving reasons for the justification of proof, concept, methodology, and logical criteria based on existing information or data, where this reasoning is presented in the form of an argument. e) Self-regulation, namely a person's ability to have the awareness to examine self-cognitive activities, the elements used in these activities, and the results, using analysis and evaluation skills, in order to confirm, validate, and re-correct the results of reasoning that have been done previously.

Based on the graph above, it shows that the indicators of interpretation, analysis, evaluation, explanation, and self-regulation have students' critical thinking skills are classified as low, namely at 40%–58%. The high and low categories of critical thinking skills certainly have causal factors [29]. One of the causative factors identified as inhibiting factors for critical thinking skills is misconceptions. Several other studies also mention that misconceptions are an inhibiting factor for the development of critical thinking skills [30, 31]. The fact that students experience misconceptions is often found in the use of concepts that are still wrong when solving problems. Another point of view that arises when identifying information on the question is able to bring up other assumptions that are not in accordance with the context being discussed in the question, thus triggering misconceptions [32]. According to [33] the low critical thinking ability of students is thought to be caused by students not being trained enough to solve contextual problems that require students to use their logic and reasoning. Many students have difficulty solving, critical thinking questions. This difficulty results in low critical thinking skills of students. So that it is necessary to do an error analysis of the results of students' answers to find out the types of errors that students make. The results of the analysis of student's answers are expected to be used as reference material for educators in providing assistance to students appropriately so that in the future students' critical thinking skills learn better.

4   Conclusion

Based on the results of the research above, it can be concluded that the analytical rubric used by peer assessment to measure students' critical thinking skills is valid and reliable. However, based on the results of students' critical thinking skills have not given satisfactory results. This has many influencing factors such as lack of experience of students in conducting peer assessments, and online assessments, lack of concentration, time, and lack of understanding of the rubric content and misconceptions. Thus, the teacher's task is very important in preparing strategies and methods of appropriate learning to provide opportunities for students to practice problem-solving.

References

1. S. N. Suwaibah, E. Susilaningsih, and S. Sudarmin, "Pengembangan instrumen performance assessment praktikum kimia dengan estimasi reliabilitasnya menggunakan program genova," *Chem. Educ.*, vol. 5, no. 1, 2016.

2. I. Ardli, A. G. Abdullah, S. Mudalifah, and A. Ana, "Perangkat Penilaian Kinerja Untuk Pembelajaran Teknik Pemeliharaan Ikan," *invotec*, vol. 8, no. 2, 2012.

3. A. R. Wulan, N. Y. Rustaman, D. J. M. D. Kembara, and U. P. I. PRESS, *Menggunakan Asesmen Kinerja: untuk Pembelajaran Sains dan Penelitian*. UPI Press, 2020. [Online]. Available: https://books.google.co.id/books?id=et_3DwAAQBAJ

4. N. Falchikov, *Improving assessment through student involvement: Practical solutions for aiding learning in higher and further education*. Routledge, 2013.

5. C. Kulkarni *et al.*, "Peer and self assessment in massive online classes," *ACM Trans. Comput. Interact.*, vol. 20, no. 6, pp. 1–31, 2013.

6. H. Y. Liang, F. I. Tang, T. F. Wang, and S. Yu, "Evaluation of nurse practitioners' professional competence and comparison of assessments using multiple methods: Self-assessment, peer assessment, and supervisor assessment," *Asian Nurs. Res. (Korean. Soc. Nurs. Sci).*, vol. 15, no. 1, pp. 30–36, 2021.

7. D. Boud and E. Molloy, "Rethinking models of feedback for learning: the challenge of design," *Assess. Eval. High. Educ.*, vol. 38, no. 6, pp. 698–712, 2013, doi: https://doi.org/10.1080/02602938.2012.691462.

8. M. B. L. Donia, T. A. O'Neill, and S. Brutus, "The longitudinal effects of peer feedback in the development and transfer of student teamwork skills," *Learn. Individ. Differ.*, vol. 61, pp. 87–98, 2018, doi: https://doi.org/10.1016/j.lindif.2017.11.012.

9. M. Salehi and B. Daryabar, "Self-and peer assessment of oral presentations: Investigating correlations and attitudes," *English Specif. Purp. World*, vol. 15, no. 42, pp. 1–12, 2014.

10. L. Hsia, I. Huang, and G. Hwang, "A web-based peer-assessment approach to improving junior high school students' performance, self-efficacy and motivation in performing arts courses," *Br. J. Educ. Technol.*, vol. 47, no. 4, pp. 618–632, 2016.

11. M. Wulandari, S. Sriyati, and W. Purwianingsih, "Penerapan peer dan self assessment sebagai tolok ukur penilaian kinerja siswa pada materi sistem koordinasi kelas XI SMA," *Assim. Indones. J. Biol. Educ.*, vol. 3, no. 2, pp. 63–68, Sep. 2020, doi: https://doi.org/10.17509/aijbe.v3i2.28258.

12. R. Rochmiyati, "Model Peer Assessment pada Pembelajaran Kolaboratif Elaborasi IPS Terpadu di Sekolah Menengah Pertama," *J. Penelit. dan Eval. Pendidik.*, vol. 17, no. 2, pp. 333–346, Dec. 2013, doi: https://doi.org/10.21831/pep.v17i2.1704.

13. A. Juhanda, "Optimalisasi diskusi kelas melalui peer assessment dan self assessment untuk menilai kemampuan komunikasi lisan mahasiswa calon guru biologi," *J. Pendidik. Mat. dan IPA*, vol. 8, no. 2, pp. 1–9, 2017.

14. J. Lu and N. Law, "Online peer assessment: Effects of cognitive and affective feedback," *Instr. Sci.*, vol. 40, no. 2, pp. 257–275, 2012.

15. A. Havnes, B. Christiansen, I. T. Bjørk, and E. Hessevaagbakke, "Peer learning in higher education: Patterns of talk and interaction in skills centre simulation," *Learn. Cult. Soc. Interact.*, vol. 8, pp. 75–87, 2016.

16. H. Riese, A. Samara, and S. Lillejord, "Peer relations in peer learning," *Int. J. Qual. Stud. Educ.*, vol. 25, no. 5, pp. 601–624, 2012.

17. S. Kearney, "Improving engagement: the use of 'Authentic self-and peer-assessment for learning' to enhance the student learning experience," *Assess. Eval. High. Educ.*, vol. 38, no. 7, pp. 875–891, 2013.

18. P. Verma and S. K. Sood, "Internet of Things-based student performance evaluation framework," *Behav. Inf. Technol.*, vol. 37, no. 2, pp. 102–119, 2018.

19. E. Z.-F. Liu and C.-Y. Lee, "Using peer feedback to improve learning via online peer assessment.," *Turkish Online J. Educ. Technol.*, vol. 12, no. 1, pp. 187–199, 2013.
20. N. C. DiDonato, "Effective self-and co-regulation in collaborative learning groups: An analysis of how students regulate problem solving of authentic interdisciplinary tasks," *Instr. Sci.*, vol. 41, no. 1, pp. 25–47, 2013.
21. M. C. English and A. Kitsantas, "Supporting student self-regulated learning in problem-and project-based learning," *Interdiscip. J. Probl. Learn.*, vol. 7, no. 2, p. 6, 2013.
22. H. K. Suen, "Peer assessment for massive open online courses (MOOCs)," *Int. Rev. Res. Open Distrib. Learn.*, vol. 15, no. 3, pp. 312–327, 2014.
23. A. Ndoye, "Peer/Self Assessment and Student Learning.," *Int. J. Teach. Learn. High. Educ.*, vol. 29, no. 2, pp. 255–269, 2017.
24. S. A. Wind and M. E. Peterson, "A systematic review of methods for evaluating rating quality in language assessment," *Lang. Test.*, vol. 35, no. 2, pp. 161–192, 2018.
25. G. A. Brown, J. Bull, and M. Pendlebury, *Assessing student learning in higher education.* Routledge, 2013.
26. L. De Grez, M. Valcke, and I. Roozen, "How effective are self-and peer assessment of oral presentation skills compared with teachers' assessments?," *Act. Learn. High. Educ.*, vol. 13, no. 2, pp. 129–142, 2012.
27. C. Bryan and K. Clegg, *Innovative assessment in higher education: A handbook for academic practitioners.* Routledge, 2019.
28. J. E. McPeck, *Critical thinking and education.* Routledge, 2016.
29. S. R. Rosdiana, S. Sutopo, and S. Kusairi, "Kemampuan Berpikir Kritis Siswa SMA pada Materi Fluida Statis," *J. Pendidik. Teor. Penelitian, dan Pengemb.*, vol. 4, no. 6, pp. 731–737, 2019.
30. F. I. Permanawati, A. Agoestanto, and A. W. Kurniasih, "The students' critical thinking ability through problem posing learning model viewed from the students' curiosity," *Unnes J. Math. Educ.*, vol. 7, no. 3, pp. 147–155, 2018, doi: https://doi.org/10.15294/ujme.v7i3.25025.
31. L. F. Santos, "The role of critical thinking in science education.," *Online Submiss.*, vol. 8, no. 20, pp. 160–173, 2017.
32. P. Kowalski and A. K. Taylor, "Reducing students' misconceptions with refutational teaching: For long-term retention, comprehension matters.," *Scholarsh. Teach. Learn. Psychol.*, vol. 3, no. 2, p. 90, 2017.
33. A. Mubarok and E. Y. Dewi, "Analisis Kesalahan Peserta Didik SMP dalam Menyelesaikan Soal Kemampuan Berpikir Kritis Berdasarkan Newman," *J. Lebesgue J. Ilm. Pendidik. Mat. Mat. Dan Stat.*, vol. 2, no. 1, pp. 121–135, 2021.