International Journal of Instruction *July 2022 ● Vol.15, No.3*
e-ISSN: 1308-1470 ● www.e-iji.net *p-ISSN: 1694-609X*
pp. 751-766

Article submission code: 20210817103352
Received: 17/08/2021 Revision: 10/03/2022
Accepted: 05/04/2022 OnlineFirst: 27/05/2022

How do Students Perform a Peer Assessment?

Kistantia Elok Mumpuni
Department of Biology Education, Faculty of Teacher Training and Education, Universitas Sebelas Maret Surakarta, Indonesia, *kistantiaelok@staff.uns.ac.id*

Desy Fajar Priyayi
Department of Biology Education, Faculty of Teacher Training and Education, Universitas Kristen Satya Wacana Salatiga, Indonesia, *desy.priyayi@uksw.edu*

Sri Widoretno
Department of Biology Education, Faculty of Teacher Training and Education, Universitas Sebelas Maret Surakarta, Indonesia, *sriwidoretno@staff.uns.ac.id*

Peer assessment is an evaluation in which the students are assessing each other. However, do they understand what to assess, how to assess, and the objectivity of the assessment? This study aims to analyse students' ability to conduct peer assessments. The research applied a quantitative research design using questionnaire. The participant was a pre-service teacher who assessed their peers' teaching practices focused on basic teaching skills using questionnaire. The data of ten skills were obtained and analysed using a multi-rater Rasch model with Many-Facet Rasch Measurement. Findings revealed that students' assessment was reliable (0.99), with a high separation index (9.46). The chi-square test showed a significant difference between the levels of the assesses. The exact agreement score was 47.3%, and the expected score was 46.8%. The data showed that the assessors did the scoring independently, the distribution of respondents and items' difficulty levels with the same scales were clear. In conclusion, the results indicated that the students were able to conduct peer assessment objectively. These findings provide support for using peer assessment as an objective and an authentic assessment. For recommendation, the grading rubric should be arranged as clear as possible, conduct assessor training, and more often tests to reduce bias.

Keywords: peer assessment, students' ability, rasch model, teaching skills, evaluation

INTRODUCTION

Peer assessment is an activity that provides opportunities for students to consider and determine the grades, value, product qualities, or performances of the other peer students (K. J. Topping, 2009). Students can perform the peer assessment toward some activities such as science laboratory activities, presentation skills, and project products. Peer assessment Peer assessment is one of the authentic assessments that provides

Citation: Mumpuni, K. E., Priyayi, D. F., & Widoretno, S. (2022). How do students perform a peer assessment? *International Journal of Instruction*, *15*(3), 751-766. https://doi.org/10.29333/iji.2022.15341a

opportunities for the students to contribute their roles during the learning process (Karami & Rezaei, 2015).

Peer assessment facilitates the students to help each other in the learning process by identifying their strengths and weaknesses. They can give each other suggestions and reinforcements to achieve the learning targets and improve their metacognitive skills (K. J. Topping, 2009). Peer assessment can be a critical source of feedback for student learning. Meaningful feedback makes the students be more responsible for their behavior while providing personal improvement and development (Fete, Haight, Clapp, & McCollum, 2017). The understanding of assessment criteria also improves understanding in achieving higher standards. Students can reflect on the learning process and consider a peer assessment to improve their learning. (Bloxham & West, 2004; Orsmond, Merry, & Reiling, 2002).

Alzaid (2017) and Sahin (2008) stated that in the implementation of performance assessments, there were similarities in the results of self-assessment and peer assessment between the assessments carried out by students and the assessments of lecturers. Conversely, Kartono (2011) found that found that students' peer assessment and self-assessment were not equivalent to the teachers' summative assessment. Kartono further suggested that the student's understanding of the assessment rubric for summative tests should be checked. This was because the assessors (students) had clear tendencies and interests to boost their assessments, especially during the self-assessment or assessing their close friends. There was also some indication that there was no consistent relationship between the students' ability to assess others and their grade for the assignment (Bloxham & West, 2004).

The problem with peer assessment is the difficulty in determining the validity and reliability of the students' assessments or whether they gave the correct scores or not. The previous report showed the lower reliability found for the assessment in practice course than the assessment for the academic product (Topping, 2009). The solution is to use strict assessment criteria with multiple raters (assessors). Generally, peer assessment requires more initial preparation than its implementation (Bostock, 2000). It is critical that the marking criteria have to be clearly defined, unambiguous, specific, and transparent to the students (Purchase, 2000; Rust, 2002). Therefore, the students are provided the information regarding what standards must be obtained to achieve different grades (Rust, Price, & O'Donovan, 2003).

Another problem is scoring bias, or when the assessors are hesitant to decide their assessments. Doubts arose when the assessors did not understand the assessed content or material, especially if they were complex. Another bias occurred when the students find that there is a mismatch between the assessment guidelines and the practices. There is also the tendency to give good scores so that the assessment did not reflect the actual learning outcomes. Gu (2020) explained some challenges in conducting peer assessment, such as the students' hesitation to criticize their peers and the students' doubtfulness in each other's competency for proper judgment. Therefore, whether students give objective assessments or not needs to be examined in-depth. It will affect the quality of the scores.

Furthermore, some results of peer assessment are dubious (Izgar & Akturk, 2018). However, Topping (1998) argued that undergraduate students have mature age and are able to think and to be trusted. Stefani (1998) also stated that expressing criteria in ways that students can understand is enormously difficult because of the students' interpretations of criteria influenced by their social and cultural backgrounds. Therefore, it is crucial to know the ability of the undergraduate student to conduct the peer assessment. To measure the students` ability, the peer assessment research often employs the questionnaire as the instrument (Patchan, Schunn, & Clark, 2018; Ratminingsih, Artini, & Padmadewi, 2017).

The use of questionnaires to measure learning outcomes is ubiquitous, but information on reliability and validity is often based on problematic Classical Test Theory approaches. Based on Item Response Theory, the Rasch Analysis offers a better alternative for exploring the quality of rating scales and informing scale improvements (Van Zile-Tamsen, 2017). So far, most data from questionnaires are usually seen in percentages. Percentages summarize the data into numbers. This raw score could not show students' abilities in detail. The Rasch Analysis can change the raw scores into a processed score that provides more accurate information (Sumintono & Widhiarso, 2015). In Sumintono & Widhiarso (2015), Rasch explained that "A person having a greater ability than other persons should have a greater opportunity to answer one item correctly. With the same principle, the more difficult items, the chances of a person answering correct will be smaller". The implementation of the Rasch model is based on the specific characteristic implied in the model: that both person and item parameters are aligned on the same scale. It can identify aspects of further development as well as indicators of biases (if any) (Maseko, Luneta, & Long, 2019; Wei, Liu, & Jia, 2014). Thus, the resulting data are valid, and data analysis can be discussed in more depth. It is also suitable as a tool in assessment for learning (Sumintono, 2018).

However, there are still limited numbers of Rasch model analysis for peer assessment. In addition, the ability of undergraduate students to perform peer assessment is unclear. This study focuses on revealing the pre-service teachers' ability in peer assessment on teaching practices. More importantly, it is expected to know the strengths and weaknesses of peer assessment related to the respondents' ability. In turn, the results can be used to improve peer assessment methods. Also, it is projected that the study would significantly help teachers and students in knowing the development of skills and planning for further learning improvements.

METHOD

Design

This research follows the quantitative design with a questionnaire as the instrument. Respondents answered the questionnaire then the resulted data generalized the description of populations. It is widely known that the quantitative or numerical description of samples can be generalized to claim the condition of population (Creswell, 2009). In this study, the curated data were analyzed using Rasch model (Sumintono, 2018; Sumintono & Widhiarso, 2014).

Procedure

This research was conducted on students who enrolled on the Basic Teaching Skills Course at the Biology Education Department, UNS. The learning activities were student-centered. The students were practicing the teaching techniques focused on ten basic teaching skills. Those ten skills were to provide explanations, ask questions, provide reinforcement, provide apperception and motivation, guide discussions, manage classes, teach small groups and individuals, evaluate learning, and give assignments. The lecturer acted as a facilitator, not the primary source of learning. The questionnaire as peer assessment instrument was designed by the lecturer based on the indicators that would be assessed. Therefore, the students did not participate in the process of instrument development. This study was conducted for a semester.

First, the students constructed the lesson plans, which consisted of ten basic teaching skills. The lecturer examined the lesson plans. In the following meetings, students practiced their teaching according to the revised lesson plans. Two students were practicing for each session. At the end of each practice session, students were allowed to assess their practices and friends. The assessment was given as the online forms. Thus, students could see the summary of the peer assessment right after the assessment.

Participant

Seventeen students who enrolled in the Basic Teaching Skills courses were selected as participants. There were 14 females, and three males with ages ranged from 20-21 years old.

Material

The data were collected using the questionnaire describing the implementation of ten basic teaching skills. The instrument was modified from the assessment sheet for Field Teaching Practice, specifically on the learning objectives to be achieved. The scores were on a scale of four. Apart from ratings (scores), the participants also provided feedback. The assessment was conducted when the participant completed the teaching practice. The items to be assessed and the rating scales were shown in Table 1. The items' order was the same as the codes for the analysis. For example, Item #2 on analysis was the skill to give explanations.

Table 1
Rating scale to assess the teaching skills that have been utilized for the analysis

No	Skills	Definition
1	Opening the Learning Process	Provide the apperceptions to focus the students on learning and motivate them to be more enthusiastic.
2	Explaining	Present information using structured and systematic oral speech that students can understand.
3	Questioning	Ask questions to increase participation and encourage students to discuss the answers.
4	Giving Reinforcement	Give the feedback (support or correction) based on the students' behavior which can motivate students to be better.
5	Making Variations of Learning Activities	The students are not bored, and the excellent atmosphere is maintained by adjusting the learning methods and media.
6	Guiding the Discussion	Students are actively involved in proposing opinions to solve the problem with directions from the teacher.
7	Managing Classroom	Optimal and conducive learning conditions are generated
8	Teaching Small Groups and Individuals	Students are active in learning, and there is a relationship between teachers and students, as well as students and students.
9	Evaluating Learning	Assessments are practical and following the learning objectives and topics that have been

No	Skills	Definition
	Process	determined.
10	Giving Assignment to Students	The tasks are relevant to improve students' conceptual understanding.

Data Analysis

The data were analyzed using the multi-rater Rasch Model with the Facets (Many-Facet Rasch Measurement) V. 3.83.2. The Rasch Model changed the ordinal raw data by looking at the odds and then using the logarithmic function, resulting in data with the same interval called logit (log odds unit) (Linacre, 1999; Sumintono, 2018). It can be used to solve the ordinal raw score of the Likert rating, which probably has different intervals among the scores (San Martín & Rolin, 2013).

Seventeen participants did ten assessment items. Thus, the total data were 2890 data (17x17x10). The results of the Rasch Analysis described the students' abilities. The students who gave scores were called assessors, and the assessed students were called assessees. Table 2 showed the codes for assessors and asseessees. The same line described the same person. Assessor A was the same person as assessee 1. The different codes were used to differentiate between assessors and assessees.

Table 2
Codes for assessor and assessee involved in this study. The assessors were coded as letters (A to Q), while the assessees were coded by numbers (1 to 17)

Student	Assessor	Assessee
Student 1	A	1
Student 2	B	2
Student 3	C	3
Student 4	D	4
Student 5	E	5
Student 6	F	6
Student 7	G	7
Student 8	H	8
Student 9	I	9
Student 10	J	10
Student 11	K	11
Student 12	L	12
Student 13	M	13
Student 14	N	14
Student 15	O	15
Student 16	P	16
Student 17	Q	17

FINDINGS AND DISCUSSIONS

Peer Assessment has some shortcoming that makes possible for errors to arise. So, more assessors were needed to minimize the errors. In this study, 17 students were involved in the Peer Assessment process, where they assessed all their friends and then assessed themselves. This method was expected to minimize errors. The data were analyzed using

the Many-Facet Rasch Measurement. The analysis was focused on the assessor measurement reports, probability curves, unexpected responses, and bias/interaction reports. The data illustrated the students' ability to conduct peer assessments.

Assessor Measurement Report

Assessor is the important component in the peer assessment to determine the quality of the scores. Things to be considered are the accuracy of the assessors, the values of infit and outfit, reliability, separation and strata, and the agreement.

Table 3
Statistic on assessor fit of peer assessment

Assessor	Logit	Standard Error Measurement	Infit Mnsq	Outfit Mnsq	Point Measurement Correlation
E	0.05	0.18	1.07	1.06	0.38
I	-0.86	0.16	1.44	1.44	0.27
M	-1.22	0.17	0.63	0.60	0.23
H	-1.42	0.19	0.61	0.55	0.62
D	-1.78	0.19	0.69	0.64	0.49
B	-1.88	0.17	0.93	0.92	0.22
J	-1.93	0.18	0.35	0.33	0.50
L	-1.93	0.19	1.01	1.00	0.52
F	-1.99	0.17	1.23	1.27	0.49
P	-2.31	0.17	0.65	0.62	0.10
N	-3.00	0.18	1.02	0.99	0.47
G	-3.09	0.17	1.70	1.76	0.44
O	-3.33	0.17	1.40	1.41	0.31
A	-4.55	0.17	0.98	0.98	0.44
K	-5.02	0.18	1.15	1.19	0.35
C	-6.40	0.29	1.27	1.20	0.24
Q	-6.86	0.30	1.03	0.29	0.76

Table 3 shows the pattern for the quality of the assessors. The score in the S.E model suggested the assessors' levels of accuracy. Good assessors should have the logit value <0.5, so it can be concluded that all the assessors gave scores carefully. The assessors with the infit and outfit Mnsq values close to 1 were the most ideal. It can be concluded that the ideal assessor was Assessor L. A positive correlation score indicated that the assessor's quality was good. Meanwhile, Assessor G and J had a certain note. Assessor G was an underfit (high mean-squares), mean-square > 1.5, which means his scoring was too unpredictable. Meanwhile, Assessor J was an overfit (low mean-squares), mean-square < 0.5, which means his scoring was too predictable.

The assessors' reliability was very good (0.99), which means good assessors' consistency. Sahin, Teker, & Güler (2016) also confirmed the function of peer assessment as a reliable assessment for distinguishing students' performances. Topping (2009) stated that peer assessment has higher levels of reliability and validity than the teacher's assessment. Peer assessment is an effective technique to assess students of all age levels and assess various competencies. This technique can provide feedback for

students to find out and improve their strengths and weaknesses. They also develop their metacognitive skills.

Table 4
Assessors' ability in peer assessment

Reliability	0.99
Separation	9.46
Strata	12.95
significance chi-squared	0.40
Exact agreements	47.3%
Expected agreements	46.8%

The separation value was 9.46, with the strata value of 12.95. The greater the separation value, the better the instrument was. It was because the instrument could identify respondents and item groups. The strata value was 12.95 (rounded to 13), which means there were 13 groups of respondents. Random (normal) chi-square measured a random sample from the normal distribution and got the p-value of 0.40, so the hypothesis was not rejected.

The agreement showed percentages of the ratings by agreed assessors (M. Linacre, 2012). The exact agreement (ratings by an assessor were agreed upon by other assessors) should by same or slightly higher than the expected agreement. Notes bellow Table 1 showed the exact agreement value was 47.3%, and the expected agreement was 46.8%. These small differences indicated a good assessors' quality. The assessors gave scores independently and not just copycatting. The percentage of the agreement showed the agreed value between the assessors on specific items. This result corresponds to the data in Figure 1.

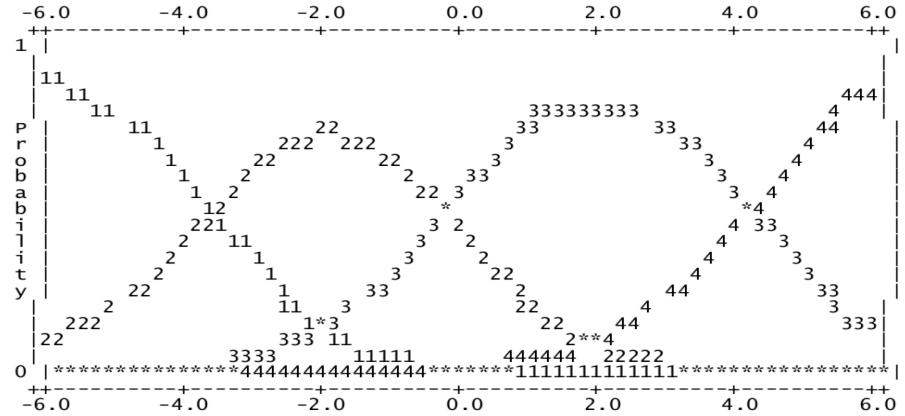

Figure 1
Probability curves of assessor's distribution in peer assessment

Figure 1 indicated four separate peaks. It pinpointed that the assessor understood the differences between rating scores with four scales (1, 2, 3, and 4). The assessees got four

points if they fulfilled all criteria. However, if all criteria were not shown, the assessees got one point. The results suggested that students understand how to use the instruments.

Wright Map

The Wright Map described the distribution of respondents and items' difficulty levels on the same scales. It was the advantage of using the Rasch Model. The Wright map visualized both the location of the items and the respondent in the measurement dimensions (Sumintono & Widhiarso, 2015). The Wright Map was shown in Figure 2.

As displayed in Figure 2, there are three categories: the assessees, assessors, and items. Each has a different grouping rank. In the assessee column, the student with the best performance was at the top, and the lowest was at the bottom. Assessee 6 and Assessee 9 had the best performances and Assessee 17 had the lowest one. The top part was the most difficult item to be assessed in the item column, while the bottom was the easiest one. The most difficult item to be assessed was Item3 (questioning skills). Item10, Item2, Item6, and Item9 were at the same level. In the assessor column, the top was Assessor E, which means that E had the strictest criteria to give a good score compared to other students. On the other hand, the Assessor Q was the most generous one.

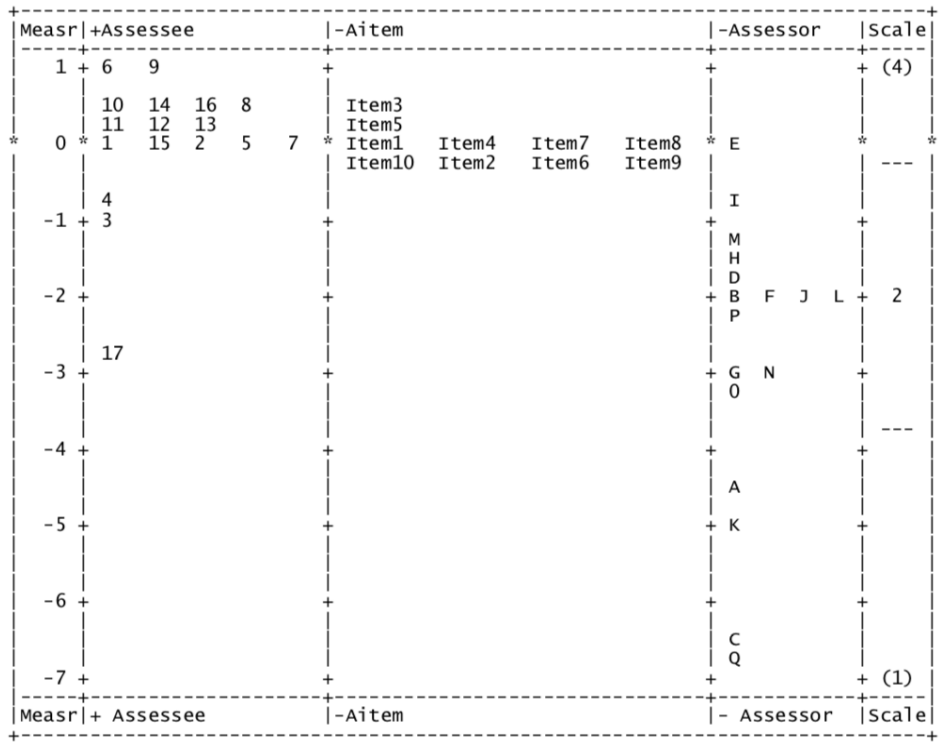

Figure 2
Wright Map Distribution of Assessee, Assessor, and Item on the Same Scales

Unexpected Responses

Unexpected responses are responses with standardized residuals equal to or exceeding the amount specified. In Facet-Rasch Model, it is viewed as a Table 5.

Table 5
Sample of unexpected responses among the assessee, assessor, and item

Sequence	Score	Expected	StRes	Assessor	Assessee	Item
370	3	4.0	-5.5	C	6	Item10
889	1	3.1	-4.8	G	3	Item9
380	3	4.0	-4.5	C	8	Item10
425	3	3.9	-3.5	C	14	Item5
439	3	3.9	-3.5	C	15	Item9
440	3	3.9	-3.5	C	15	Item10
940	2	3.5	-3.0	G	9	Item10
2562	2	3.5	-3.0	Q	17	Item2
2564	2	3.5	-3.0	Q	17	Item4
2566	2	3.5	-3.0	Q	17	Item6
688	1	2.6	-2.9	E	15	Item8
2568	2	3.5	-2.9	Q	17	Item8
2201	2	3.3	-2.7	O	15	Item1
2202	2	3.3	-2.7	O	15	Item2
915	2	3.2	-2.6	G	7	Item5
2203	2	3.2	-2.6	O	15	Item3
1731	2	3.1	-2.5	L	12	Item1
165	4	3.0	2.4	B	2	Item5

As shown in Table 5, a sample of 100 out of 2890 data were outside the estimate value (0.03%). A more precise calculation for Assessor C on Item10 showed the standardized residual (StRes) value was -5.5. It was the most unexpected among these data. The minus (-) sign means C did worse than expected. Some students gave lower or higher scores to other students. For example, C gave a three-point for Assessee 6, even though the expected value for Item10 was four (skill to give assignment).

Some unexpected responses were also found when the self-assessments were conducted. Table 4 portrayed that students gave lower scores for themselves. For example, Q gave two points for himself (Item2, explaining), even though the expected score was 3.5. Some students gave high scores for themselves. For example, B gave four points for I5, higher than the expected score of three.

It indicated that the list of unexpected responses was very useful to improve the implementation of peer assessment. The teachers could quickly evaluate the students who did not perform the appropriate assessments. They might confirm to their students by direct discussion then improve the students' skills. The instruments could be improved by investigating which items were subjected to the most judgment bias. The teacher can improve the learning strategy by looking at student data and unexpected items.

Students' Ability to Conduct Peer Assessment

The results showed that students could conduct peer assessments. The assessor measurement report showed that the assessors could judge independently so that the results were accountable. This was shown by Patchan, Schunn, & Clark (2018) that

undergraduate students could provide ratings and feedback. The assessed students (assessees) said that the feedback and ratings improved the draft lab report. Their research explained that assessors not only can provide feedback but also critics, solutions, and localized comments. These findings suggested that useful commentaries can have a significant effect on peer assessment and the consistency of students' ratings. The implementation of peer assessment also avoids multiple perspectives on the assessment result given by the teacher (Sluijsmans, Brand-Gruwel, van Merriënboer, & Bastiaens, 2002).

According to Murillo-Zamorano & Montanero (2018), feedback from peer assessment can improve oral presentation skills by 10%, while feedback from teachers can only increase it by 5%. In this study, combined feedback and ratings by peers were expected to improve the students' skills. Follow-up re-tests were necessary to find out how much the peer feedback and ratings can improve the pre-service teachers' basic teaching skills.

The Wright Map projected that Assessor E is the most parsimonious one. Assessor E gave brief feedback to the assessees, for example, "good"; "should be improved"; and "should be better prepared." However, Assessor E only had a single unexpected data. This means that Assessor E had high consistency in giving the ratings. Assessor E had an exact observed agreement of 36.7% and agreement expected of 39.8%. The difference between the two was less than 10%, so it can be concluded that the Assessor E was good. However, did Assessor E got bad ratings and comments from others?

Assessor E was coded as Assesse 5. Table 4 describes that Assesse 5 was in the middle position. It means that the Assesse 5's performance was good. Assesse 5 got various comments, such as: "Apperception was going too far into the topic, it should be limited"; "Class conditions are not conducive so that students become passive"; "Overall, the learning process was good, but can be improved to be more interesting"; "Good"; "Interesting and mind-opening lessons about the importance of family planning"; and some other similar comments. It can be seen that students were responsible for providing logic ratings and feedback, not backbiting each other.

Accordingly, a corresponding result was explained by Double, McGrane, & Hopfenbeck (2020). The peer assessment did not affect the grading result significantly when it included the comments (critics) from the assessors or when it did not. The comments (writing component) on various types of peer assessment, such as a combination between written qualitative feedback and numerical grade, likely do not affect the independence of the assessment.

In addition, Piaget's theory on cognitive development in adolescence should be considered. According to this theory, the undergraduate students have passed the fourth stage (formal operational). They can make decisions based on real experience and think more abstractly, idealistically, and logically (Santrock, 2008). They can also distinguish the skilled students and unskilled one and determine the levels of ratings on the questionnaire (see Figure 1). Thus, students gave their grades independently and objectively.

The use of online assessment forms also has an effect. Li et al. (2019) claimed that computer-aided peer assessment showed a more significant effect and quality than paper-based peer assessment. Chen (2016) stated that technology could provide many advantages for peer assessment, including flexibility, efficiency, and accessibility. Students are already accustomed to their gadgets and prefer to type rather than write them on paper. Assessment using the online form makes their task easier. Many factors should be considered when preparing a peer assessment to maximize the opportunities for the students to improve their skills. One of the factors is assessor training.

In this study, the minor unexpected responses occurred because of different understandings of the criteria for assessment. The solution was through conducting assessor training. Assessor training was the most significant factor in data bias. Assessor training improves the assessment quality to 9.18%, and it was significant if this was tested alone or in combination with other factors (Li et al., 2019). Sluijsmans et al. (2002) revealed that students' assessment skills could be trained, which positively affect the quality of feedback. When peer assessment would be conducted, the teacher should inform students at the beginning of the lesson. The teacher should ensure fairness and honesty in the assessment process (Baker, 2008).

The quality in peer assessment can be engaged by developing clear criteria that are relevant to the learning outcomes, providing multiple assessors for each work, using the anonymity of assessors, giving training in assessment, and frequency (more than one peer assessment session) (Wride, 2017). Clear criteria and using multiple assessors can improve validity and reliability of student assessments, as well as the anonymously of assessor and assessee to reduce student discomfort (Li et al., 2019). It can also be repeated multiple times (frequency) and trained before the assessment. It helps students develop their experience and build in an element of inter-subjectivity, which makes the marking more objective. Patchan et al. (2018) explained that if peer assessment is compulsory, students might feel more accountable, which makes them responsible for the assessment they provide to their peers.

The Benefit of Peer Assessment for Students

The design and structure of the peer assessment invite students to learn what criteria should be assessed (Vickerman, 2009). Through peer assessment, students become more active in the learning process because they are involved in the assessment process and understanding the learning achievement criteria (Kearney, 2013). Students also give more positive perceptions because they could know their performance, strength, and weakness from peers' feedbacks (Ratminingsih et al., 2017). Papinczak, Young, & Groves (2007) found peer assessment had a positive impact in building student responsibility. It also develops collaborative learning to make a better lesson plan and performance.

Peer assessment is also related to peer feedback, cognitive and metacognitive strategies, and learning achievements (Liu & Lin, 2007). The use of cognitive and metacognitive strategies on peer feedback has a significant relationship with the students' learning achievement. Peer assessment with peer feedback can build students' cognitive and

metacognitive strategies, improving students' learning achievement. The student's involvement in peer assessment affects their metacognitive awareness. It could help them consider how to develop self-concept on what and when to do learning (Pantiwati & Husamah, 2017). They become more attentive in their work and know what to do to reach a certain achievement and improvement (Bloxham & West, 2004; Fete et al., 2017). Peer assessments can improve deep analysis (Joordens, Desa, & Paré, 2009) and evaluative and critical thinking skills (Sluijsmans et al., 2002). Sande & Godino-Llorente (2014) clarified that students' problem-solving skills in peer assessment are better than students who participated in self-assessment.

Non-cognitive benefits can be seen in the positive impacts on students' teamwork, social skills, and learning motivations (Karami & Rezaei, 2015), which make them more attentive during group work, and increase their accountability and responsibility (K. J. Topping, 2009). Self and peer assessments can help students overcome social and life problems such as stereotyping and peer pressures (Harrison, O'hara, & McNamara, 2015). Early students training to assess their work by themselves and others and provide and receive feedback can train them to think independently and assessing based on the assessment objectives. It also trains them to discriminate phenomena or events.

The initial implementation of the peer assessment may not be as good as expected. However, if it is applied continuously, students will get used to it and produce better data. Therefore, teachers have to design the proper peer assessment to maximize its benefits. According to Brookhart (2015), teachers need to choose the right type of performance to be assessed, so the students better understand the learning objectives.

CONCLUSION

Slight differences between exact agreements and expected agreements showed that students could conduct the peer assessment. The percentage agreement indicated the agreement value between the assessors. The exact agreement value was 47.3% and had a good independence limit (20%-80%). Assessors portrayed good understanding on how to use the assessment instrument with four-scale ratings. This claim was supported by students' cognitive abilities, the ease use of online media, and the easy-to-understand instruments.

Peer assessments can provide cognitive and non-cognitive benefits. The use of peer assessment obviously benefits both the teachers and students. The results may vary depending on the design and implementation. It is recommended to use peer assessment as an objective tool to evaluate student performances in teaching and learning processes. However, to minimize the different understanding of marking criteria, the grading rubric should be arranged as clear as possible and the initial assessor briefing need to be held before the assessment process. Also, it is suggested that peer assessment needs more applications and experimental research in the future.

REFERENCES

Alzaid, J. M. (2017). The effect of peer assessment on the evaluation process of students. *International Education Studies*, *10*(6), 159.

https://doi.org/10.5539/ies.v10n6p159

Baker, D. F. (2008). Peer assessment in small groups: A comparison of methods. *Journal of Management Education*, *32*(2), 183–209. Retrieved from https://doi.org/10.1177/1052562907310489

Bloxham, S., & West, A. (2004). Understanding the rules of the game: Marking peer assessment as a medium for developing students' conceptions of assessment. *Assessment and Evaluation in Higher Education*, *29*(6), 721–733. https://doi.org/10.1080/0260293042000227254

Bostock, S. (2000). Student peer assessment Stephen Bostock. *Studies in Higher Education*, 1–4. Retrieved from http://www.keele.ac.uk/depts/cs/Stephen_Bostock/docs/caa-ktn.htm

Brookhart, S. M. (2015). Performance assessment. In *SAGE Handbook of Research on Classroom Assessment*. https://doi.org/10.4135/9781452218649.n18

Chen, T. (2016). Technology-supported peer feedback in ESL/EFL writing classes: a research synthesis. *Computer Assisted Language Learning*, *29*(2), 365–397. https://doi.org/10.1080/09588221.2014.960942

Creswell, J. W. (2009). Research design. In *SAGE Publications* (3rd ed.). https://doi.org/10.1080/14675980902922143

Double, K. S., McGrane, J. A., & Hopfenbeck, T. N. (2020). The Impact of Peer Assessment on Academic Performance: A Meta-analysis of Control Group Studies. *Educational Psychology Review*, *32*(2), 481–509. https://doi.org/10.1007/s10648-019-09510-3

Fete, M. G., Haight, R. C., Clapp, P., & McCollum, M. (2017). Peer evaluation instrument development, administration, and assessment in a team-based learning curriculum. *American Journal of Pharmaceutical Education*, *81*(4). https://doi.org/10.5688/ajpe81468

Gu, C. (2020). Student Peer Assessment. *Review of Educational Theory*, *3*(2), 74. https://doi.org/10.30564/ret.v3i2.1762

Harrison, K., O'hara, J., & McNamara, G. (2015). Re-thinking assessment: Self- and peer-assessment as drivers of self-direction in learning. *Egitim Arastirmalari - Eurasian Journal of Educational Research*, (60), 75–88. https://doi.org/10.14689/ejer.2015.60.5

Izgar, G., & Akturk, A. O. (2018). A Mixed Method Research on Peer Assessment. *International Journal of Evaluation and Research in Education (IJERE)*, *7*(2), 118. https://doi.org/10.11591/ijere.v7i2.12770

Joordens, S., Desa, S., & Paré, D. (2009). The pedagogical anatomy of peer-assessment: Dissecting a peerScholar assignment. *IMSCI 2009 - 3rd International Multi-Conference on Society, Cybernetics and Informatics, Proceedings*, *2*(5), 263–267. Retrieved from http://www.iiisci.org/journal/CV$/sci/pdfs/XE123VF.pdf

Karami, A., & Rezaei, A. (2015). An Overview of Peer-Assessment: The Benefits and Importance. *Journal for the Study of English Linguistics*, *3*(1), 93. https://doi.org/10.5296/jsel.v3i1.7889

Kartono. (2011). Efektivitas Penilaian Diri dan Teman Sejawat untuk Penilaian Formatif dan Sumatif pada Pembelajaran Mata Kulian Analis Kompleks. *Prosiding Seminar Nasional Matematika*, 49–59. Retrieved from https://publikasiilmiah.ums.ac.id/bitstream/handle/11617/573/MAK-KARTONO-%2849-59%29.pdf?sequence=1&isAllowed=y

Kearney, S. (2013). Improving engagement: The use of "Authentic self-and peer-assessment for learning" to enhance the student learning experience. *Assessment and Evaluation in Higher Education*, *38*(7), 875–891. https://doi.org/10.1080/02602938.2012.751963

Li, H., Xiong, Y., Hunter, C. V., Guo, X., & Tywoniw, R. (2019). Does peer assessment promote student learning? A meta-analysis. *Assessment and Evaluation in Higher Education*, *45*(2), 193–211. https://doi.org/10.1080/02602938.2019.1620679

Linacre, J. M. (1999). Investigating rating scale category utility. *Journal of Outcome Measurement*, *3*(2), 103–122. Retrieved from http://jampress.org/jom_v3n2.pdf

Linacre, M. (2012). *Facets Tutorial 3*. 1–29. Retrieved from https://www.winsteps.com/a/ftutorial3.pdf

Liu, E. Z. F., & Lin, S. S. J. (2007). Relationship between peer feedback, cognitive and metacognitive strategies and achievement in networked peer assessment: Colloquium. *British Journal of Educational Technology*, *38*(6), 1122–1125. https://doi.org/10.1111/j.1467-8535.2007.00702.x

Maseko, J., Luneta, K., & Long, C. (2019). Towards validation of a rational number instrument: An application of rasch measurement theory. *Pythagoras*, *40*(1), 1012–2346. https://doi.org/10.4102/PYTHAGORAS.V40I1.441

Murillo-Zamorano, L. R., & Montanero, M. (2018). Oral presentations in higher education: a comparison of the impact of peer and teacher feedback. *Assessment and Evaluation in Higher Education*, *43*(1), 138–150. https://doi.org/10.1080/02602938.2017.1303032

Orsmond, P., Merry, S., & Reiling, K. (2002). The use of exemplars and formative feedback when using student derived marking Criteria in peer and Self-assessment. *Assessment and Evaluation in Higher Education*, *27*(4), 309–323. https://doi.org/10.1080/0260293022000001337

Pantiwati, Y., & Husamah. (2017). Self and peer assessments in active learning model to increase metacognitive awareness and cognitive abilities. *International Journal of Instruction*, *10*(4), 185–202. https://doi.org/10.12973/iji.2017.10411a

Papinczak, T., Young, L., & Groves, M. (2007). Peer Sssessment in Problem-Based Learning: A qualitative study. *Advances in Health Sciences Education*, *12*(2), 169–186.

https://doi.org/10.1007/s10459-005-5046-6

Patchan, M. M., Schunn, C. D., & Clark, R. J. (2018). Accountability in peer assessment: examining the effects of reviewing grades on peer ratings and peer feedback. *Studies in Higher Education*, *43*(12), 2263–2278. https://doi.org/10.1080/03075079.2017.1320374

Purchase, H. C. (2000). Learning about interface design through peer assessment. *Assessment and Evaluation in Higher Education*, *25*(4), 341–352. https://doi.org/10.1080/026029300449245

Ratminingsih, N. M., Artini, L. P., & Padmadewi, N. N. (2017). Incorporating self and peer assessment in reflective teaching practices. *International Journal of Instruction*, *10*(4), 165–184. https://doi.org/10.12973/iji.2017.10410a

Rust, C. (2002). The Impact of Assessment on Student Learning: How Can the Research Literature Practically Help to Inform the Development of Departmental Assessment Strategies and Learner-Centred Assessment Practices? *Active Learning in Higher Education*, *3*(2), 145–158. https://doi.org/10.1177/1469787402003002004

Rust, C., Price, M., & O'Donovan, B. (2003). Improving Students' Learning by Developing their Understanding of Assessment Criteria and Processes. *Assessment & Evaluation in Higher Education*, *28*(2), 148–164. https://doi.org/10.1080/0260293032000045509

Sahin, M. G., Teker, G. T., & Güler, N. (2016). An Analysis of Peer Assessment through Many Facet Rasch Model. *Journal of Education and Practice*, *7*(32), 172–181.

Sahin, S. (2008). An application of peer assessment in higher education. *Turkish Online Journal of Educational Technology*, *7*(2), 5–10.

San Martín, E., & Rolin, J. M. (2013). Identification of parametric Rasch-type models. *Journal of Statistical Planning and Inference*, *143*(1), 116–130. https://doi.org/https://doi.org/10.1016/j.jspi.2012.06.014

Sande, J. C. G., & Godino-Llorente, J. I. (2014). Peer assessment and self-assessment: Effective learning tools in higher education. *International Journal of Engineering Education*, *30*(3), 711–721.

Santrock, J. W. (2008). *Educational Psychology*. New York: McGraw-Hill.

Sluijsmans, D. M. A., Brand-Gruwel, S., van Merriënboer, J. J. G., & Bastiaens, T. J. (2002). The training of peer assessment skills to promote the development of reflection skills in teacher education. *Studies in Educational Evaluation*, *29*(1), 23–42. https://doi.org/10.1016/S0191-491X(03)90003-4

Stefani, L. A. J. (1998). Assessment in partnership with learners. *Assessment and Evaluation in Higher Education*, *23*(4), 339–350. https://doi.org/10.1080/0260293980230402

Sumintono, B. (2018). *Rasch Model Measurements as Tools in Assesment for Learning*.

173(Icei 2017), 38–42. https://doi.org/10.2991/icei-17.2018.11

Sumintono, B., & Widhiarso, W. (2014). *Aplikasi Model RASCH untuk Penelitian Ilmu-ilmu Sosial* (1st ed.; B. Trim, ed.). Cimahi: Penerbit Trim Komunikata.

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan RASCH pada Assessment Pendidikan* (B. Trim, ed.). Cimahi: Penerbit Trim Komunikata.

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, *68*(3), 249–276. https://doi.org/10.3102/00346543068003249

Topping, K. J. (2009). Peer assessment. *Theory into Practice*, *48*(1), 20–27. https://doi.org/10.1080/00405840802577569

Van Zile-Tamsen, C. (2017). Using Rasch Analysis to Inform Rating Scale Development. *Research in Higher Education*, *58*(8), 922–933. https://doi.org/10.1007/s11162-017-9448-0

Vickerman, P. (2009). Student perspectives on formative peer assessment: An attempt to deepen learning? *Assessment and Evaluation in Higher Education*, *34*(2), 221–230. https://doi.org/10.1080/02602930801955986

Wei, S., Liu, X., & Jia, Y. (2014). Instrument of Students ' Understanding of Models in. *International Journal of Science & Mathematics Education*, *12*(5), 1067–1082.

Wride, M. (2017). Guide to peer-assessment. In *Academic Practice and eLearning (CAPSL) Resources*. Retrieved from https://www.tcd.ie/CAPSL/assets/pdf/Academic Practice Resources/Guide to Student Peer Assessment.pdf