ISSN: 1309 – 6575
Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi
Journal of Measurement and Evaluation in Education and Psychology Research Article; 2023; 14(3); 266-286

Analysis of Peer and Self-Assessments Using the Many-facet Rasch Measurement Model and Student Opinions

Seda DEMİR*

Abstract

The aim of this study is to analyze the peer and self-assessments of higher education students' oral presentation skills with the many-facet Rasch measurement model and to determine students' opinions on peer and self-assessment. In the study, the convergent parallel method, one of the mixed-method research approaches, was used. The study group consisted of 11 university students studying at a state university in the 2022-2023 academic year. The FACETS program was used to analyze the data. The three facets identified in the study were the assessee (11 students), the assessor (11 students), and the items (16 items). Therefore, 11 participants scored (peer and self-assessment) on a 16-item assessment form. In addition, students' opinions on peer and self-assessment were obtained through three open-ended interview questions prepared by the researcher. According to the results of the study, it was determined that there was a statistically significant difference between the students in terms of their oral presentation skills, between the assessors in terms of their strictness/generosity in scoring, and between the criteria (items) in terms of the level of difficulty in realization. In addition, the participant opinions obtained from each interview question were analyzed through themes and sub-themes formed according to the general thoughts on peer and self-assessment, experiences, and whether the participants considered themselves as a reliable rater or not. In terms of practice, it can be suggested to provide detailed and enlightening information to students before peer and/or self-assessment in the classroom environment, and to give quick feedback to those who have not done the assessment appropriately. In addition, the reasons for the biases identified in peer and self-assessments in the current study can be investigated in future studies.

Keywords: Peer assessment, Self-assessment, Many-facet Rasch measurement model, Oral presentation skills

Introduction

Effective assessment of the educational process can be considered as one of the basic requirements that contribute to the discovery and development of students' true potential. In this context, it can be said that alternative (performance-based) assessment methods that support a student-centered education approach offer the opportunity to assess students' different aspects, learning styles or abilities and thus provide a more comprehensive learning process. In this process, students are generally expected to be able to apply knowledge to real-world situations. Peer and self-assessment, alternative methods, are among the most widely researched assessment methods in the literature that encourage students' active participation in assessment processes and develop their self-confidence (Falchikov & Goldfinch, 2000). According to Cheong et al. (2023) peer and self-assessment are processes in which students judge the quality of their peers or their own work. Peer assessment involves students assessing each other's work and providing feedback (Evans et al., 1993), while self-assessment allows students to observe and assess their own learning processes (Boud & Falchikov, 1989). Therefore, self-assessment is known to be closely related to reflection (Yan & Brown, 2017) and during peer assessment, students can benefit from both giving and receiving feedback (Hoo et al., 2021; Liu & Carless, 2006; Lundstrom & Baker, 2009). Peer and self-assessment practices help students identify their own strengths and weaknesses and create motivation for lifelong learning (Hanrahan & Isaacs, 2001; Panadero et al., 2023; Sande & Godino-Llorente, 2014).

Various studies in the literature indicate that peer assessment provides cognitive, affective, pedagogical, and metacognitive benefits to students (Butler & Winne, 1995; Nicol & Macfarlane-Dick, 2006; Orluwene & Ekim, 2020; Tseng & Tsai, 2007; Zhan et al., 2023). According to Brown (2004), if the

* Assist. Prof. Dr., Tokat Gaziosmanpasa University, Faculty of Education, Tokat-Turkey, seddadmr@gmail.com, ORCID ID: 0000-0003-4230-5593

To cite this article:
Demir, S. (2023). Analysis of peer and self-assessments using the many-facet rasch measurement model and student opinions. *Journal of Measurement and Evaluation in Education and Psychology, 14*(3), 266-286. https://doi.org/10.21031/epod.1344196

Received: 16.08.2023
Accepted: 27.09.2023

students selected for peer assessment are adequately informed and the assessment process is planned correctly, this approach can save considerable time for the teachers and the process, give students a chance to learn in depth, and help support students' higher-order cognitive activities. In their study, Crisp and Ward (2008) stated that peer assessment is a method of assessment that offers constant feedback while also improving students' academic achievement, class participation, and motivation. Panadero et al. (2023) examined the relationships between peer assessment and intrapersonal and interpersonal factors through a systematic review on peer assessment. According to the study, there are six intrapersonal factors including motivation, self-efficacy, emotions, trust in the self as an assessor, fairness, and comfort; and five interpersonal factors including social connections, trust in the other as an assessor, psychological safety, value diversity/congruence, and interdependence in peer assessment. In Boud and Falchikov's (1989) study, it was determined that students with high achievement gave themselves lower grades in self-assessment, while students with low achievement gave themselves higher grades. In this direction, Hanrahan and Isaacs (2001) emphasized in their study that students should be informed and trained in detail about how to assess self-assessment. Yan et al. (2021) examined the effect of self-assessment on academic performance in a meta-analysis study. The results of the study show that the overall effect of self-assessment was significant. Topping et al. (2000) compared the reliability of peer assessment and self-assessment results and found that peer assessment was a more reliable method than self-assessment. Gürlen et al. (2019) examined the reliability coefficients of teacher, self- and peer evaluations of primary school students with the help of generalizability theory. The results of the study revealed that the variance component estimated for the student main effect was the largest component of the total variance. According to the literature, it is possible to say that peer and self-assessment practices help students develop high-level skills such as taking ownership of their own learning and abilities to think critically, creatively, and analytically, solve problems, present information clearly, and conduct research. In addition, they aim to involve students in the assessment process and support student learning rather than grading (Pantiwati & Husamah, 2017; Stefani, 1994). In addition, in the study conducted by Cheong et al. (2023) with undergraduate students, peer and self-assessment were used together for the academic writing task and it was investigated how self-assessment complemented peer assessment. As a result of the study, self-assessment has been found to complement peer assessment in five ways: it guides students to make corrections when peer assessment is incomplete; when students have access to peer assessment, self-assessment effectively supports peer assessment; even when a student has access to quality peer assessment, self-assessment complements peer assessment because of the different reflections in the two processes; self-assessment can support peer assessment on issues related to social emotional burdens; and self-assessment also complements peer assessment in that it benefits high and low-achieving students. Therefore, it can be said that complementing peer assessment with self-assessment is an effective solution to overcome possible problems that may be encountered in the peer assessment process, such as students' limited ability to provide feedback and non-objective assessment.

In addition to peer and self-assessment, presentations are frequently preferred, especially in higher education, in order to ensure students' active participation in the course. In measuring presentation skills, the use of rubrics provides an objective assessment process and offers the opportunity to give more qualified and meaningful feedback. According to Fete et al. (2017), meaningful feedback enables students to be more responsible for their behaviors while ensuring their personal growth and development. In addition, students who know the scoring criteria produce better-quality work (Liu et al., 2001; Lu & Law, 2011). However, scoring may not always be based solely on performance. Various sources of variance (factors/facets) may be involved in scoring , which may negatively affect the validity of scoring (Prieto & Nieto, 2014). As stated in the study conducted by Gu (2020), there are some problems in peer assessment, such as students' hesitation to criticize their peers and students' doubts about each other's ability to make correct decisions. In addition, there are also studies showing that students score themselves lower or higher than they should be in self-assessments (Mumpuni et al., 2022; Semerci, 2011a). Therefore, it is necessary to examine in depth whether students make objective evaluations or not. For this reason, in the current study, peer and self-assessments of students' oral presentation skills were analyzed using a many-facet Rasch measurement model in which assessor characteristics were added as facets to the measurement model.

The Rasch measurement model (Linacre, 1993) is utilized to objectively calculate the precise intervals between options in tests, scales, and rubrics. This method aids in determining the interval unit with greater precision and accuracy (Elhan & Atakurt, 2005). In the many-facet Rasch measurement model, there is no facet limitation and it is suitable for multiple scoring (Eckes, 2005). With this model, the facets (such as assessor, assessee, and items) that may affect the predictions for the latent trait measured are considered. Semerci (2011a) analyzed faculty member, peer, and self-assessments within the framework of doctoral qualifications with the Rasch measurement model and determined the differences observed in student performances, jury strictness/generosity, and the difficulty/ease of the tasks expected to be performed. Similarly, Köse et al. (2016) analyzed rater, criterion, and presentation skills using peer assessments of student presentations with the many-facet Rasch measurement model. As a result, it was exemplified that the many-facet Rasch measurement model is an alternative measurement model that can be used to determine student performance. Mumpuni et al. (2022), in their study aiming to analyze how peer assessment takes place, concluded that students have the ability to make peer assessments and make their assessments objectively.

When the related literature is examined, it is seen that there are qualitative studies on peer assessment and/or self-assessment, or quantitative studies designed according to the many-facet Rasch measurement model. However, depending on the rubric used, the interview questions or the study group, the results of the studies differ from each other, and the need for new research arises.

As a result, the aim of this study was to analyze the peer and self-assessments of the students taking the Teaching Probability and Statistics course on oral presentation skills with the many-facet Rasch measurement model and to determine the students' views on peer and self-assessment.

In line with this overall objective;

i. General analysis of opinions on oral presentation skills in the Teaching Probability and Statistics course,

ii. Analysis of assessors' rigor/generosity,

iii. Task difficulty analysis of oral presentation skills,

iv. Assessor bias analysis,

v. Qualitative data obtained from interviews with all students participating in the study will be analyzed.

It is thought that the current study will provide students and educators with ideas about the objective use of peer and self-assessments, which are alternative assessment methods. In general, the accuracy of an assessment is directly related to the validity of the previous assessment. Accordingly, the fact that the scoring of oral presentation skills contains bias errors will directly affect the validity negatively. In addition, students' views on peer and self-assessment also enriched the study in terms of qualitative data. Therefore, this study, designed as a mixed research, is thought to be important in terms of its contribution to related literature.

Methods

This study employs the convergent parallel method, one of the mixed method research approaches, to collect and analyze both quantitative and qualitative data simultaneously (Creswell, 2014). In this regard, the peer and self-assessments of higher education students' oral presentation skills with the many-facet Rasch measurement model and students' opinions on peer and self-assessment were merged for a more complete understanding.

Study Group

In the Rasch measurement model, there is no assumption that sample statistics generalize to the population (Linacre, 1993). Therefore, the study group was determined for the research. The same study

group took part in both quantitative and qualitative parts of the research. The study group consisted of 11 university students (four males and seven females) who took the Teaching Probability and Statistics course at a state university in the 2022-2023 academic year. The convenience sampling method was used for the selection of participants. The participants were selected on a purely voluntary basis among the students who made oral presentations within the course.

Data Collection

In the study, all students enrolled in the Teaching Probability and Statistics course were first told about the goals of peer and self-assessment and what they should pay attention to in their assessments with the use of the peer and self-assessment guide prepared by the researcher. Then, 11 students who volunteered to participate in the study were identified. The participants both made oral presentations and then self-assessed and made peer assessments by listening to other oral presentations. In order to score the students' oral presentation skills, the "Oral Presentation Skills Peer Assessment Form" developed by the researcher and the "Oral Presentation Skills Self-Assessment Form" consisting of the same items were used. The items used to evaluate students' oral presentation skills are presented in Table 1.

Table 1

Items Used to Evaluate Oral Presentation Skills

Heading	Skills
Form of presentation	1. The subject is emphasized with main lines.
	2. Fluent language is used.
	3. The tone of voice is used correctly.
Content	4. The ideas put forward on the subject are supported by solid evidence.
	5. The examples given on the subject are interesting and original.
	6. There are no contradictory explanations about the subject.
Understanding the Subject and Participation in Discussions	7. The subject is fully understood.
	8. Sufficient technical information is given.
	9. The subject is presented in a convincing way.
	10. An overall evaluation including important points has been made.
	11. Thoughts have been expressed clearly.
	12. Questions and comments have been successfully answered.
	13. The more complex parts of the subject have been sufficiently emphasized.
Communication Skills and Time Management	14. Good communication has been established with the audience.
	15. The listeners who asked questions or made comments were not interrupted.
	16. Time was used efficiently and there were no problems in time management.

As seen in Table 1, the peer and self-assessment forms consisted of 16 items. These items-range from 1 (very inadequate) to 5 (very adequate) and scored on a five-point Likert scale. In addition, students' views on peer and self-assessment were obtained through three open-ended interview questions prepared by the researcher.

Data Analysis

In the quantitative part of the study, students' peer and self-assessments were analyzed using the many-facet Rasch measurement model. In this model, multiple sources of variability (ability, item, rater, situation, task, etc.) can be analyzed simultaneously and independently (Mulqueen et al., 2000; Sudweeks et al., 2005). In addition, the analysis results obtained from the sample are not intended to be generalized to the population (Linacre, 1993). In the analysis of the data, the FACETS program developed by Linacre (1993, 2023), which deals with three facets as ability, item/measure/task, and rater

as a general use, was used. The three facets identified in the current study were the assessee (11 students), the assessor (11 students), and the items (16 items). Therefore, 11 participants scored (peer and self-assessment) on the 16-item assessment form and a total of 1936 (11x11x16) data were obtained. 11 students both scored as assessors and were scored as assessee. In this context, Assessee 1 and Assessor A, Assessee 2 and Assessor B, Assessee 3 and Assessor C, Assessee 4 and Assessor D, Assessee 5 and Assessor E, Assessee 6 and Assessor F, Assessee 7 and Assessor G, Assessee 8 and Assessor H, Assessee 9 and Assessor I, Assessee 10 and Assessor J, and Assessee 11 and Assessor K are codes representing the same student. The interpretation of peer or self-assessments was made by considering these codes. In addition, each item was coded according to the order in which it appeared in the form, for example, Item 1, Item 2. With the many-facet Rasch measurement model, the study explored factors such as assessors' fairness, bias, the ease or difficulty of criteria, and identified which students had stronger oral presentation skills based on the established criteria.

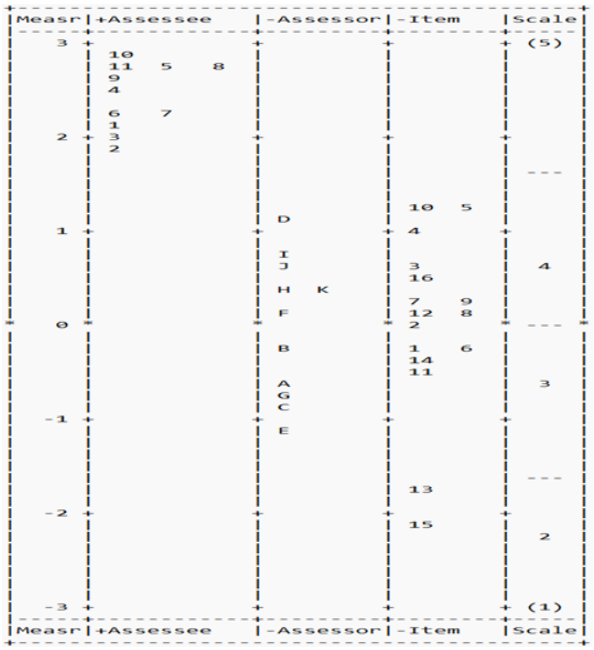
In the qualitative part of the study, themes and sub-themes were formed by content analysis of 11 students' responses to three open-ended questions to determine their views on peer and self-assessment. The opinions of the participant students were given in the form of quotations. For the quotations, the codes representing the assessors (Assessor A, Assessor B, Assessor C, Assessor D, Assessor E, Assessor F, Assessor G, Assessor H, Assessor I, Assessor J, and Assessor K) were used to represent the same students. For the reliability of the study, firstly, the participants' responses to each open-ended question were combined in a single document. Then, another expert was consulted for the codes and themes determined based on these responses. In addition, the confirmation of the findings obtained from the current study with the participants can be considered as evidence for the internal validity of the study, and the fact that the findings are compatible with the literature can be regarded as evidence for the external validity of the study.

Results

In the analysis of oral presentation skills with the many-facet Rasch measurement model, three facets (assessee, assessor, and items) were used. The Wright Map containing general information about these facets is presented in Figure 1.

Figure 1

Calibration Map of the Distribution of Assessee, Assessor and Items



When Figure 1 is examined, it is seen that the oral presentation students are ranked on the same logit scale according to their ability level (assessee), the strictness/generosity of the raters (assessor) and the difficulty level of the tasks (items). In this distribution, the assessee facet is ranked from the best oral presentation performance to the lowest, the assessor facet is ranked from the strictest rater to the most generous, and the item facet is ranked from the most difficult task to the easiest, from top to bottom. Accordingly, in terms of the oral presentation, Assessee 10 had the best performance, while Assessee 2 had the lowest performance. However, Assessor D gave the strictest assessment and Assessor E gave the most generous assessment. Based on the data obtained, it can be said that Assessor D, who gave the strictest score, realized a moderately good oral presentation, while Assessor E, who gave the most generous score, realized the second best oral presentation. In addition, Item 5 and Item 10 were determined as the most difficult items (the most difficult criterion/task to perform), while Item 15 was determined as the easiest item (the easiest criterion/task to perform).

The detailed measurement report on the oral presentation skills of 11 undergraduate students (assessees) who took the Teaching Probability and Statistics course is presented in Table 2.

Table 2

Oral Presentation Skills Measurement Report of the Assessees

Assessee	Logit	Standart Error	Infit MnSq	Outfit MnSq
10	2.87	0.17	1.07	1.18
8	2.79	0.16	0.97	0.82
5	2.69	0.16	1.03	1.22
11	2.69	0.16	0.93	0.76
9	2.58	0.15	1.09	1.09
4	2..45	0.14	1.08	0.86
7	2.31	0.14	0.78	0.69
6	2.27	0.14	0.89	0.93
1	2.15	0.13	1.11	1.54
3	1.98	0.13	1.08	1.10
2	1.88	0.12	1.03	1.15
RMSE= 0.14	sd= 0.28	Separation= 1.96	Strata= 2.94	Reliability= 0.79
chi-square= 56.2	df= 10	*p*= 0.00		

When Table 2 is examined, it is seen that the Root Mean Square Standard Error- RMSE of the measurement values is calculated as 0.14 and the standard deviation is calculated as 0.28. In addition, the separation index was calculated as 1.96 and the strata value as 2.94. The discrimination index shows the ability of the measurement tool in Rasch analysis to distinguish participants with different ability levels (Linacre, 1994). The higher the discrimination index, the better the measurement tool is understood to be (Mumpuni et al., 2022). The strata value calculated as approximately three indicates that there are three groups of students in terms of oral presentation skills. The reliability coefficient obtained from the analysis shows that the students assesses in terms of oral presentation skills are ranked with 0.79 confidence. In addition, according to the chi-square test results ($\chi 2$ =56.2, df=10, p=0.00), the null hypothesis was rejected. Therefore, it was determined that there was a statistically significant difference between the students in terms of oral presentation skills.

The general order of the students assessed in terms of oral presentation skills from the best performer to the lowest performer is as follows: 10, 8, 5, 11, 9, 4, 7, 6, 1, 3, 2. Here, Assessees 5 and 11 have the same performance. In self-assessment, this order is: Assessee 10, 1, 5, 8, 9, 3, 7, 2, 6, 4, 11, and Assessee 1, 5, 8, 9 perform similarly. On the other hand, in peer assessment, Assessee 10, 8, 5, 11, 6, 9, 4, 7, 1, 3, 2, and Assessee 4, 7 have the same performance. When these rankings are analyzed collectively, it can be said that it is a remarkable finding that the position of Assessee 10 did not change.

In Rasch analysis, the fit and misfit values of the facets, which indicate the degree of fit between the data and the model, are also calculated. The out-of-fit statistic, which is more sensitive to unexpected extreme values compared to the in-fit statistic, is equal to the mean squares of the residuals between the observed data and the expected values (Randall and Engelhard, 2009). A fit statistic of 1 indicates that the variance between the data is greater than expected; a fit statistic of less than 1 indicates that the variance between the data is less than expected. The range of 0.5 to 1.5 for fit statistics is the range of values considered appropriate as an indicator of accurate and effective measurements (Turner, 2003; Wright & Linacre, 1994). Accordingly, it can be concluded that the model fits the data obtained from all the assessed data.

The detailed measurement report for the 11 assessors who scored oral presentation skills is presented in Table 3.

Table 3

Assessors' Strictness/Generosity Measurement Report

Assessor	Logit	Standart Error	Infit MnSq	Outfit MnSq
D	1.18	0.11	0.93	0.85
I	0.80	0.11	1.01	0.89
J	0.66	0.12	0.96	0.87
H	0.40	0.13	1.06	0.81
K	0.39	0.13	1.22	1.58
F	0.16	0.14	1.12	1.11
B	-0.26	0.16	0.73	1.41
A	-0.62	0.19	0.79	0.80
G	-0.69	0.19	1.53	1.68
C	-0.94	0.21	0.82	0.77
E	-1.08	0.23	0.72	0.56
RMSE= 0.16	sd= 0.71	Separation= 4.41	Strata= 6.21	Reliability= 0.95
chi-square= 235.5	df= 10	*p*= 0.00	Inter-rater exact agreements= 60.1%	

According to Table 3, with a measurement value of 1.18, it is seen that Assessor D is the strictest, and with a measurement value of -1.08, Assessor E is the most generous in scoring. Therefore, the general order of the assessors is from the most strict to the most generous in terms of scoring oral presentation skills is Assessor D, I, J, H, K, F, B, A, G, C, and E. In addition, as seen in Table 3, the standard error of the measurement values was calculated as 0.16, the standard deviation as 0.71, the separation index as 4.41 and the strata value as 6.21. The strata value calculated as approximately six indicates that there are six groups of assessors in terms of strictness/generosity in scoring oral presentation skills. The reliability coefficient of 0.95 obtained from the Rasch analysis shows that the students who were assessed in terms of their strictness/generosity were ranked with very high reliability. Moreover, according to the results of the chi-square test ($\chi 2$ =235.5, df=10, p=0.00), the null hypothesis was rejected. Therefore, there was a statistically significant difference between the assessors regarding their strictness/generosity in scoring.

When the congruent and incongruent values of the facets in the Rasch analysis are analyzed, it is seen that only the incongruent value of Assessor K and Assessor G is outside the recommended value range (0.5 to 1.5 range). Accordingly, it can be said that there are some inconsistencies in the scoring of Assessor K and Assessor G. Finally, according to Table 3, the absolute inter-rater agreement value was calculated as 60.1%.

A detailed measurement report on the criteria/tasks (items) in the form used to assess students' oral presentation skills is presented in Table 4.

Table 4

Measurement Report of the Items Used to Assess Oral Presentation Skills

Item	Logit	Standart Error	Infit MnSq	Outfit MnSq
5	1.25	0.13	0.81	0.83
10	1.21	0.13	1.19	1.09
4	1.05	0.13	0.81	0.79
3	0.61	0.15	0.97	1.12
16	0.56	0.15	1.57	1.26
7	0.24	0.17	0.79	1.02
9	0.21	0.17	0.85	0.78
8	0.07	0.18	0.74	0.68
12	0.07	0.18	0.89	0.74
2	0.00	0.18	0.99	1.14
6	-0.24	0.20	1.81	2.34
1	-0.28	0.20	0.74	0.57
14	-0.37	0.21	1.05	0.92
11	-0.50	0.22	1.01	0.70
13	-1.76	0.38	1.03	1.52
15	-2.11	0.45	0.98	0.97
RMSE= 0.22	sd= 0.88	Separation= 4.00	Strata= 5.67	Reliability= 0.94
chi-square= 243.5	df= 15	*p*= 0.00		

When the item measurement report in Table 4 is examined, according to the measurement values obtained, it is seen that the most difficult criterion (the criterion with the lowest rate of high score) is Item 5: "The examples given on the subject are interesting and original." with a measurement value of 1.25, followed by Item 10: "A general assessment including the important points of the subject was made." The easiest criterion (with the highest rate of high scores) was Item 15: "Listeners who asked questions or made comments were not interrupted." with a measurement value of -2.11. This was followed by Item 13: "The more complex parts of the topic were sufficiently emphasized." A visual of these results is given in Figure 1. In addition, as seen in Table 4, the standard error of the measurement values was calculated as 0.22, the standard deviation as 0.88, the separation index as 4.00 and the strata value as 5.67. In addition, the calculated reliability value is quite high at 0.94. The significant results of the chi-square test ($\chi 2$= 243.5, df=15, p=0.00) indicate a statistically significant difference between the difficulty levels of the criteria. When the fit statistics for the criteria are examined, it is observed that all criteria except the sixth criterion are between acceptable values within and outside the acceptable fit. Accordingly, it can be said that only the sixth criterion is an obstacle to data-model fit.

With the help of the many-facet Rasch analysis, unexpected responses obtained with the measurement tool can also be identified. Unexpected responses show which rater scored the response of which individual in an unexpected way. In addition, it provides information (such as training of raters and revision of items) for determining the sources of decreased reliability and planning the measurement process more reliably (Güler, 2014; Nakamura, 2002).

In the current study, a sample of unexpected responses between the assessee, the assessor, and the item is presented in Table 5.

Table 5

Unexpected Responses between Assessee, Assessor, and Item

Sequence	Score	Expected	StRes	Assessor	Asseessee	Item
102	2	4.9	-8.2	1	G	6
454	2	4.9	-7.7	3	G	6
278	2	4.8	-7.2	2	G	6
111	4	5.0	-6.9	1	G	15
1613	4	5.0	-6.7	10	B	13
733	4	5.0	-6.1	5	B	13
1437	4	5.0	-5.7	9	B	13
879	4	5.0	-5.2	5	K	15
909	4	5.0	-4.9	6	B	13
1622	4	5.0	-4.3	10	C	6
1581	4	4.9	-4.1	9	K	13
175	4	4.9	-3.9	1	K	15
974	3	4.8	-3.9	6	F	14
1298	4	4.9	-3.9	8	E	2
1490	3	4.8	-3.7	9	F	2
527	4	4.9	-3.6	3	K	15
1590	4	4.9	-3.6	10	A	6
112	3	4.7	-3.5	1	G	16
694	3	4.7	-3.5	4	K	6
351	4	4.9	-3.4	2	K	15
1053	4	4.9	-3.4	6	K	13
1474	4	4.9	-3.4	9	E	2
1884	3	4.7	-3.4	11	H	12
1625	4	4.9	-3.3	10	C	9
1754	2	4.4	-3.3	10	K	10
173	4	4.9	-3.2	1	K	13
174	3	4.7	-3.2	1	K	14
775	4	4.9	-3.2	5	E	7
962	3	4.7	-3.2	6	F	2
1667	3	4.7	-3.2	10	F	3
542	4	4.9	-3.1	4	A	14
1271	4	4.9	-3.1	8	C	7
1262	4	4.9	-3.0	8	B	14

When the standardized StRes values given in Table 5 are examined, it is seen that all of them have a minus (-) sign. Accordingly, it can be said that all of the unexpected data resulted from the fact that some students gave lower than expected scores to other students. It is seen that the most unexpected data stemmed from the score given by Assessor G to Item 6 for Assessee 1. Here, while the expected value for Item 6: "There were no contradictory explanations about the topic." was 4.9, Assessor G gave 2 points to Assessee 1 for this item and the standardized StRes value was calculated as -8.2. In addition, the first four most unexpected data belong to Assessor G; all were scored below the expected value. As

a remarkable finding from the study, this means that Assessor G performed worse than expected. It is also seen that the most recurrent rater in terms of giving unexpected scores was Assessor K, and the top three items with the highest recurrence of bias were Item 7, Item 6, and Item 15, respectively.

When the data presented in Table 5 related to self-assessment are analyzed, it is seen that Assessor F gave himself lower scores than expected for Item 14: "Good communication with the audience was established." (three points were given while the expected score was 4.8) and Item 2: "Fluent language was used" (three points were given while the expected score was 4.7). Similarly, Assessor E gave herself a lower than expected score for Item 7: "The topic was fully understood" (four points were given when the expected score was 4.9).

The bias analysis of self- and peer-assessors is presented in Table 6.

Table 6

Assessor and Assessee Interaction Bias Report

Observed Score	Expected Score	Obs-Exp Average	Bias	Standart Error	*z* Score	Infit MnSq	Outfit MnSq	Assessee	Assessor
67	73.69	-0.42	-0.97	0.34	-2.84	1.3	1.2	6	F
63	70.61	-0.48	-0.89	0.32	-2.80	0.8	0.7	8	D
60	67.32	-0.46	-0.77	0.31	-2.46	0.9	0.9	3	I
73	77.09	-0.26	-1.01	0.42	-2.41	0.4	0.4	7	G
63	69.02	-0.38	-0.68	0.32	-2.12	0.8	0.7	9	D
60	66.40	-0.40	-0.66	0.31	-2.12	0.9	0.9	2	I
73	76.63	-0.23	-0.85	0.42	-2.02	2.3	3.7	1	G
76	69.72	0.39	1.16	0.53	2.19	1.2	1.0	1	J
78	71.74	0.39	1.63	0.73	2.24	1.0	0.9	9	I
80	67.99	0.75	3.58<	1.43	2.51	0.0	0.0	4	D
chi-square= 156.7		df= 121		*p*= 0.02					

The fact that the z scores given in Table 6 are outside the commonly accepted range of -2 to +2 points to interaction bias between assessors and assessees. Assessor F gave a significantly (p<0.05) rigid scoring by giving himself 67 points when he should have given himself approximately 74 points in his self-assessment. Similarly, Assessor G made a significantly rigid peer assessment for Assessee 7 and Assessee 1. In addition, it is seen that Assessor D and Assessor I gave lower scores to some students than expected in their peer assessments and made a significantly strict scoring, while they gave higher scores to some students and made a significantly generous scoring. Assessor J gave a significantly generous peer assessment for Assessee 1.

In addition to the analyses conducted with the many-facet Rasch measurement model, the participants' responses to three questions regarding their views on peer and self-assessment were also analyzed and themes and sub-themes were formed.

1. The themes and sub-themes determined in line with the answers to the question "What are your general thoughts about the peer/self-assessment practice you participated in?" are presented in Table 7 and Table 8, respectively.

Table 7

General Thoughts about the Peer Assessment Practice

Themes	Sub Themes
Benefits of Peer Assessment	Gaining a critical perspective
	Increasing awareness of responsibility
	Being respectful for different ideas
	Increasing motivation
	Gaining different perspectives
	Focusing on the learning process without worrying about grades
	Developing reasoning skills
	Developing empathy skills
	Recognizing professional values
	Gaining an objective perspective
	Providing students with the drive to be better
	Improving academic performance
	Developing reflective thinking skills
	Gaining awareness of assessment
	Supporting future development
Characteristics of the Assessment Process	Performance based
	Objectivity
	Process and product oriented
	Based on criteria
Learning Process	Increasing teacher-student coordination
	Ensuring effective participation in the lesson
	Providing a student-centered learning environment
	Creating work discipline
	Providing feedback
	Sharing responsibility for learning
Problems in the Peer Assessment Process	Time consuming
	Complexity
	Performing assessments in line with prejudices
	Lack of experience
	Increasing the level of anxiety

Some sample responses reflecting the participants' general thoughts about the peer assessment practice are given below.

Assessor B: "While doing peer assessment, I had the opportunity to assess the process as well as the product. I think that with this practice, the course was carried out in teacher-student coordination and student-centered. Although at first I found peer assessment complex and difficult due to my lack of experience and my prejudices against some of my friends, I realized that the assessments I made improved my ability to empathize and reason over time.

Assessor C: "While doing peer assessment, it is useful to know that it is important to make an assessment. In other words, it has many benefits both for ourselves and for our friends we assess. From our own point of view, we see that it develops critical thinking. For our friends, we see that it is important for them to see their shortcomings and good sides."

Assessor H: "Peer assessment makes the lesson environment more productive by making the lesson more active and attentive. I think peer assessment should be done for every lesson. The only negative aspect I can say is that the peer assessment process is a bit laborious and time-consuming. Other than that, I think it is a good assessment that should be done."

Assessor K: "I think this practice is useful for us because we make presentations by taking into consideration which criteria our friends who listen to the presentation may pay attention to while

assessing. I also think that we cannot be fully objective in scoring the individuals with whom we are in closer contact and this is the disadvantage of the application."

Table 8

General Thoughts on Self-Assessment Practices

Themes	Sub Themes
Benefits of Self-Assessment	Developing self-awareness
	Recognizing mistakes/deficiencies
	Recognizing strengths and weaknesses
	Developing self-criticism
	Improving oral communication skills
	Improving presentation performance
	Improving metacognitive thinking strategy use
	Developing creativity skills
	Improving decision-making skills
	Creating a perception of success
	Feeling valued
	Contribution to lifelong learning
	Tracking the development process
	Increasing self-confidence
	Developing multiple perspectives
	Providing personal development
	Reinforcing learning
	Creating cognitive awareness
	Developing awareness of democracy
Learning Process	Encouraging active participation in the lesson
	Taking responsibility for own learning
	Increasing the efficiency of the course
	Providing professional development
	Developing metacognition about their own learning
Problems in the Self-Assessment Process	Not assessing their own performance objectively
	Being overly critical and scoring rigidly
	Being too generous in scoring
	Not being conscious enough
	Loss of self-confidence
	Reluctance to learn

Some sample responses reflecting the participants' general thoughts about the self-assessment practice are given below.

Assessor A: "I think that self-assessment is a study developed for us to notice our mistakes or shortcomings. I believe that self-assessment will shed light on our future studies and enable us to take care not to make the same mistakes again and to continue our studies in this direction. Thanks to self-assessment, we have developed a metacognitive perspective on our learning and performance by taking responsibility for our own learning."

Assessor B: "I can honestly say that this practice leaves the person alone with himself/herself. And in this way, the person wants to be more honest with himself/herself and makes his/her assessment accordingly. Therefore, I can say that I found this practice useful. The biggest difficulty I had while trying to make an objective self-assessment was trying not to be more optimistic or pessimistic towards myself than I should be."

Assessor D: "The self-assessment practice made a great contribution to my ability to look at myself objectively and criticize myself. It enabled me to discover myself and see my strengths and weaknesses.

It contributed to gaining a realistic perspective and being impartial. I also think that self-assessment is very important not only in lessons but also in every aspect of life."

Assessor J: "Thanks to the self-assessment, I had the opportunity to realize where I was lacking and what I could do to improve myself. Self-assessment will help me perform better in other presentations by improving myself."

2. The themes and sub-themes determined in line with the responses to the question "What are your experiences with peer/self-assessment?" are presented in Table 9 and Table 10, respectively.

Table 9

Experiences with Peer Assessment

Themes	Sub Themes
Positive Experiences	Developing an empathic approach
	Gaining a critical perspective
	Listening to lesson effectively
	Analyzing the lesson process
	Objective thinking
	Gaining high-level cognitive skills
	Improving social relations
	Progression of competencies
	Improving communication skills
	Fair assessment
	Development of presentation skills
	Development of teaching skills
	Identifying misconceptions
	Developing a sense of responsibility
	Providing permanent learning
	Handling the process holistically and analytically
	Developing research skills
	Interacting with the environment
	Providing feedback
Awareness-Building Experiences	Importance of criteria-based assessment
	Recognizing the importance of making assessments independent from personal feelings and thoughts
	Identifying knowledge gaps
	Impact of peer assessment on social relationships
	Importance of fair/objective assessment

Some sample responses reflecting the participants' experiences in peer assessment are given below.

Assessor A: I realized that peer assessment is a difficult task, especially because we are at similar ages and when it comes to the negative aspects of your friends whom you like very much, whom you are sincere with, it is more difficult to point out these aspects. I gained a more critical perspective. I closed the deficiencies in myself by seeing the deficiencies of my friends. I made an effort to be fair and since I tried to assess from an objective point of view, my learning developed in parallel with this. I based my peer assessment on certain criteria. I learned that such assessments are very necessary. Finally, I realized that peer assessment is not as easy as it seems.

Assessor B: I have developed critical thinking skills and gained experience by assessing the work of my peers. I had never listened to someone's oral presentation before and reached a conclusion or seen the shortcomings of this person and thought about how to overcome these shortcomings while I was explaining. Peer assessment provided me with the opportunity to be objective and to analyze a person or myself from an objective point of view, thus forming the basis of my future experiences. The notes I

took during the lesson for assessment purposes and focusing on my friend who made the presentation made me listen to the lesson more carefully and made the lesson more productive, making my learning more permanent.

Assessor F: The peer assessment practice reminded me that my responsibility for the lesson continues. I think it contributed to my development in terms of objective assessment. In addition, since we need to have knowledge on the subject presented while performing these assessments, it directed me to listen to the presentation more effectively. While doing peer assessment, I tried to look at both positive and negative aspects at the same time. Although I avoided making comparisons between individuals, I realized that at first I filled out the form a little bit influenced by the presentation of the previous presenter.

Assessor I: I realized that in order to analyze the process correctly in peer assessment, the presentation should be listened to carefully. I saw that the assessments of almost all presenters were close to each other when they were listened to carelessly. I think that the presentation should be listened to with focus and calm mind to catch the details.

Table 10

Experiences with Self-Assessment

Themes	Sub Themes
Experiences Supporting Development	Questioning the level of self-efficacy
	Being open to development
	Developing planning skills
	Taking responsibility for own learning
	Developing the ability to make observations
	Striving for perfection
	Contribution to organizing the learning environment
	Providing academic development
	Developing affective skills
	Improving time management
	Discovering different learning methods
	Increasing attention level
	Creating active learning environment
	Providing in-depth learning
	Developing self-regulation strategies
	Creating a desire to learn
	Recognizing aspects open for improvement
	Gaining experience in the learning process
	Making original inferences
	Preparing instructional content
	Developing creative thinking skills
Various Educational Experiences	Complexity of self-assessment
	Revealing one's potential
	The difficulty of conducting objective scoring
	Necessity of process management
	Need for assessment away from comparisons
	Cognitive adaptation to the process

Some sample responses reflecting participants' experiences of self-assessment are given below.

Assessor C: The self-assessment practice contributed positively to my learning process by helping me discover myself and recognize my strengths and weaknesses. It paved the way for me to objectively and realistically assess my own performance and development throughout my life. In addition, self-assessment reminded me that my responsibility for the course continues even after I finish my presentation. If I make progress in my next presentation in terms of the issues I observe in myself and need to work on, the self-assessment practice will have made a concrete contribution to my learning.

Assessor G: "To be honest, assessing myself was more difficult than assessing someone else, but it was also useful for me to see my mistakes. I realized that when I was doing self-assessment, I was doing it by comparing myself with my other friends. Instead of assessing myself, I saw that I was ranking myself from the most successful to the least successful. When

I realized this, I did my self-assessment from the beginning. In the meantime, I approached myself with the same tolerance as I did when assessing my other friends."

Assessor H: "I think I made a good presentation, but there may be shortcomings. I think that self-assessment improved my research skills and contributed to my permanent learning. Examining the process holistically and analytically and working in a planned way before the presentation helped me to cope with my excitement. While doing self-assessment, I realized that one can give feedback even to oneself, and that while we see ourselves positively at certain points, we have mistakes at certain points. I believe that this application is suitable for eliminating these mistakes."

Assessor I: "I realized that self-assessment is actually a difficult task and that one can improve oneself according to some criteria while considering oneself adequate. I think that self-assessment enables us to manage time more easily before or during the presentation and improves self-regulation skills after the presentation. I have experienced different learning methods. I think that my creative thinking skills have improved thanks to the research and studies I have done in order to make a more effective presentation. Self-assessment has enabled me to improve my self-control, knowledge, understanding and skills and to gain the experience of looking at myself objectively even in different areas. It also gave me the experience of understanding each other in the relationship with my fellow listeners, respecting different opinions, etc."

3. Would you describe yourself as a reliable assessor when doing peer/self-assessment? Why? The themes and sub-themes determined in line with the answers given to the question are presented in Table 11 and Table 12, respectively.

Table 11

Whether the Participant Considers Him/herself Reliable in Peer Assessment

Themes	Sub Themes
Characteristics of a Self-Reliable Assessor	Compliance with the principle of impartiality
	Making assessments in line with objective criteria
	Having professional experience
	Considering only the performance
	Performing rational assessment
	The ability to utilize prior knowledge
	Mastering alternative assessment techniques
	The ability to think critically
	Having a collaborative perspective
	Being respectful for the person being assessed
	Having a constructive attitude
	Having ability to make comparisons

Table 11

Whether the Participant Considers Him/herself Reliable in Peer Assessment (Continued)

Themes	Sub Themes
Characteristics of a Partially ReliableAssessor	Not being sure about their assessments
	Lack of self-confidence
	Thinking of missing something due to inattention
	Thinking that assessments may need correction
	Having a competitive perspective
	Influenced by group dynamics
	Feeling incompetent for assessment
	Seeing oneself as inadequate for assessment
	Inability to act impartially

Some sample responses reflecting the participants' views on whether they consider themselves reliable in peer assessment are given below.

Assessor D: I define myself as a reliable assessor because I have always looked at people and situations objectively. I have not hesitated to emphasize my friends' shortcomings or strengths.

Assessor E: Yes. I consider myself to be a reliable assessor because I think I was objective in assessing even the people I was closest to. I tried to be very careful and attentive during the assessments.

Assessor F: Yes, I define myself as a reliable assessor because I listened to everyone's presentations in the group that week in line with the criteria in the scale and reflected my own views transparently in the practice by critically and analytically filtering my mind.

Assessor G: I don't think I'm completely reliable, but I would say I'm mostly reliable because I haven't done a lot of negative assessments, I'm not sure about the assessments I've done because I'm not fully qualified to assess.

Assessor J: I define it partially because as I listened to my friends, I looked at their performances in the presentation and revised the assessment scale of those whom I thought I was unfair in my previous assessments and corrected the places where I needed to make corrections. However, I may not have answered the assessment scale completely correctly for the places I missed or could not listen to, so I think I am a partially reliable assessor.

Assessor K: Of course. I listened carefully to my friends who made presentations and scored them after assessing whether the given criteria were met or not.

Table 12

Whether the Participant Considers Him/herself Reliable in Self-Assessment

Themes	Sub Themes
Requirements for Reliable Self-Assessment	Avoiding overly generous scoring
	Objectivity
	Being open to criticism
	Transparency
	Acting independently from prejudices
	Integrity
	Empathic thinking skills
	Having belief in benefits of fair assessment
	Being constructive
	Being realistic
	Having introspective skills

Table 12

Whether the Participant Considers Him/herself Reliable in Self-Assessment (Continued)

Themes	Sub Themes
Factors Affecting Self-AssessmentNegatively	Ignoration of deficiencies
	Being more tolerant/generous with oneself
	Experiencing cognitive contradiction
	Perfectionism
	Past experiences
	Lack of goal-oriented assessment
	Defensive attitude

Some sample responses reflecting the participants' views on whether they consider themselves reliable in their self-assessment are given below.

Assessor D: I define myself as reliable. Because I looked at the events objectively in my assessment. I judged myself impartially. I did not include contradictory statements.

Assessor E: Yes, I do. Because when I assessed myself, I assessed myself by taking into account my deficiencies.

Assessor F: Yes, I define myself as a reliable assessor. While sharing my personal views, I transparently conveyed what I experienced during the practices. I tried to concretize my views with additional explanations and examples I gave for clarity.

Assessor G: I don't think I am very reliable, people tend to consider themselves as perfect, I believe that people who look at me from the outside can be more objective.

Assessor J: Yes, because I think I assess myself as transparently as possible and I think I am a reliable assessor because what is important for me is to recognize my deficiencies and mistakes.

Assessor K: Yes, I see myself as a good assessor because I commented on my own performance objectively.

Discussion

In this study, the results of peer and self-assessment of 11 students' oral presentation skills in an undergraduate course using a 16-item rubric were analyzed using the many-facet Rasch measurement model. In addition, the opinions of the students participating in the study regarding peer and self-assessment were also determined simultaneously. In the current study, first of all, the data calibration map was examined to obtain general information about the relationship between the facets (assessee, assessor, and items) used in the many-facet Rasch measurement model (Nakamura, 2000) and it was seen that all facets were sorted on the same logit ruler.

The results of the study showed that there were statistically significant differences between the students' oral presentation skills, the assessors' strictness/generosity in scoring, and the criteria's (items') level of difficulty in realization. In support of this finding, in many studies in the literature (Baştürk, 2008; Baştürk, 2010; Köse et al., 2016; Mumpuni et al., 2022; Semerci, 2011a; Semerci, 2011b; Semerci et al., 2013; Uyanık et al., 2019; Yüzüak et al., 2015), it was determined that different rater characteristics created statistically significant differences between raters.

According to the oral presentation skills measurement report, the compliance statistics were among the desired values. According to the overall, peer and self-assessments, it was observed that the rankings from the best-performing student to the lowest-performing student changed in general. However, it is noteworthy that the ranking of the top-performing Assessee 10 remained the same in both peer and self-assessment. It can be interpreted that this situation indicates that the reliability of the ranking of the Assessee 10 is higher.

According to the strictness/generosity measurement report of the assessors, it was found that Assessor D was the most strict and Assessor E was the most generous in scoring. In addition, the non-compliance value of Assessor K and Assessor G was outside the desired value range. This can be interpreted as some inconsistencies in scoring of Assessor K and Assessor G. This problem can be solved by giving extra training to Assessor K and Assessor G on peer and self-assessment. In support of this finding, in most of the studies in the literature using the many-facet Rasch model (Atılgan, 2005; Baştürk, 2008, Baştürk, 2010; Semerci, 2011a, Semerci, 2011b, Akın & Baştürk, 2012; Semerci et al. 2013; Uyanık et al. 2019; Yüzüak et al. 2015), it was stated that the raters can sometimes be objective and sometimes biased.

According to the measurement report of the items used to assess oral presentation skills, the most difficult criterion is item 5: "The examples given on the topic are interesting and original." The easiest criterion is item 15: "Listeners who asked questions or made comments were not interrupted." which can be handled under the heading of communication skills and time management. In addition, the agreement statistics for Item 6: "There were no contradictory explanations about the topic." were outside the desired value range. Therefore, it can be interpreted that this item with double negativity is not a suitable item for measuring oral presentation skills. The reason for this situation may be that the item contains double negativity, both conceptual (contradictory explanation statement) and structural (not done statement).

When the unexpected responses between the assessee, assessor and item were analyzed, it was determined that Assessor G and Assessor K gave lower scores than expected in peer assessment and showed a poor performance. When the unexpected responses were analyzed in terms of self-assessment, it was seen that Assessor F and Assessor E gave themselves lower scores than expected in some items. In support of these results, when the assessee and assessor interaction bias report was examined, it was seen that Assessor F made a significantly strict self-assessment and Assessor G made a significantly strict peer assessment for some assessees. In this case, how Assessor G, Assessor K, Assessor F and Assessor E made sense of the items and how they scored them can be investigated and feedback can be given on how to make appropriate peer and self-assessment. Thus, these unexpected situations can be eliminated. However, it is seen that the first three items with the highest recurrence of bias are Item 7: "The topic was fully understood.", Item 6 and Item 15, respectively. Therefore, it can be said that the assessment forms can be further improved by reviewing and revising these items. Based on the results obtained, it can be said that examining unexpected responses is very useful in improving peer and self-assessment practices.

In the light of the results obtained from the quantitative part of the current study, which was designed as a mixed research, it can be interpreted that the many-facet Rasch measurement model provides very useful information in measurement studies where there is more than one rater and the facets determined will be examined in detail. In the qualitative part of the study, the participants' responses to three open-ended questions were analyzed to determine their views on peer and self-assessment. Regarding the first question, the participants' general thoughts about the peer assessment practice were grouped under four themes: Benefits of Peer Assessment, Characteristics of the Assessment Process, Learning Process, Problems Experienced in the Peer Assessment Process. For their general thoughts on self-assessment, three themes were identified as Benefits of Self-Assessment, Learning Process, and Problems Experienced in the Self-Assessment Process. In the second question, in which the opinions of the participants about their experiences were taken, the experiences for peer assessment were grouped under two themes as Positive Experiences and Awareness-Building Experiences, and the experiences for self-assessment were grouped under two themes as Experiences Supporting Development and Various Educational Experiences. In the third question, which asked whether the participants defined themselves as a reliable assessor, two themes were identified for peer assessment: Characteristics of a Self-Reliable Assessor and Characteristics of a Partially Reliable Assessor, and for self-assessment: Requirements for Reliable Self-Assessment and Negative Factors Affecting Self-Assessment. When the participant opinions obtained from the third question were compared with the results of the many-facet Rasch measurement model, it was seen that qualitative and quantitative partially supported each other. As a result of the analysis conducted with the many-facet Rasch measurement model, Assessor D, who was

determined as the strictest rater, Assessor E, who was determined as the most generous rater, Assessor K, who was seen to score more strictly than expected in peer assessment, and Assessor F, who was seen to score more strictly than expected in self-assessment, stated in the interview that they considered themselves as a reliable assessor, which contradicts these findings. In addition, Assessor G, who was found to have some inconsistencies in his scoring according to quantitative data, stated that he did not consider himself as a fully reliable rater in both peer and self-assessment. Therefore, it can be said that quantitative and qualitative data support each other for Assessor G.

As a result, the many-facet Rasch measurement model highlights through the designated facets, which assessors perform the bias, its source, and direction. In addition, with this study, it was tried to develop suggestions that can be effective in minimizing the errors that may be encountered in the scoring process and minimizing these errors. Participants' views are related to general thoughts and experiences about peer and self-assessment and awareness of bias in scoring. There may be many different reasons for the biases observed in peer and self- assessments. The reasons for the identified biases can be investigated in future studies. In terms of practice, it can be suggested to give detailed and enlightening information to the students before the peer and/or self-assessment in the classroom environment and to give quick feedback to those who have not done the assessment appropriately. Thus, possible biases can be minimized and students' assessment skills and indirectly the teaching process can be improved. It is recommended to employ peer assessment as an impartial instrument for assessing student performances in teaching and learning practices. It is suggested that more applications and experimental investigations related to peer assessment should be conducted in the future.

Declarations

Conflict of Interest: No potential conflict of interest was reported by the authors.

Ethical Approval: This study was approved by the Ethical Committee of Tokat Gaziosmanpasa University Social and Human Sciences Researches dated 13.06.2023 and numbered E-46052777-100-307436.

References

Akın, Ö. ve Baştürk, R. (2012). The evaluation of the basic skills in violin training by many facet Rasch model. *Pamukkale University Journal of Education, 31*(1), 175-187. https://dergipark.org.tr/en/pub/pauefd/issue/11112/132860

Atılgan, H. (2005). Analysis of Special Ability sellection examination for music education department using many-facets Racsh measurement (İnönü University case). *Eurasian Journal of Educational Measurement, 0*(20), 62-73. https://web.s.ebscohost.com/ehost/pdfviewer/pdfviewer?vid=0&sid=1c78c0f3-93e0-43b5-9499-026878d2fcb1%40redis

Baştürk, R. (2008). Applying the many facet Rasch model to evaluate PowerPoint presentation performance in higher education. *Assessment and Evaluation in Higher Education, 33*(4), 431-444. https://doi.org/10.1080/02602930701562775

Baştürk, R. (2010). Bilimsel araştırma ödevlerinin çok yüzeyli Rasch ölçme modeli ile değerlendirilmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, 1*(1), 51-57. https://dergipark.org.tr/en/pub/epod/issue/5808/77254

Boud, D., & Falchikov, N. (1989). Quantitative studies of student self-assessment in higher education: a critical analysis of findings. *Higher Education, 18*(5), 529-549. https://doi.org/10.1007/BF00138746

Brown, S. (2004). Assessment for learning. *Learning and Teaching in Higher Education, 5*(1), 81-89. https://eprints.glos.ac.uk/3607/

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65*(3), 245-281. https://doi.org/10.3102/00346543065003245

Cheong, C. M., Luo, N., Zhu, X., Lu, Q., & Wei, W. (2023). Self-assessment complements peer assessment for undergraduate students in an academic writing task. *Assessment & Evaluation in Higher Education, 48*(1), 135-148. https://doi.org/10.1080/02602938.2022.2069225

Creswell, J. W. (2014). *A concise introduction to mixed methods research*. SAGE.

Crisp, V., & Ward, C. (2008). The development of a formative scenario-based computer assisted assessment tool in psychology for teachers: The PePCAA project. *Computers & Education, 50*(4), 1509-1526. https://doi.org/10.1016/j.compedu.2007.02.004

Eckes, T. (2005). Examining rater effects in TestDaF writing and speaking performance assessments: A many-facet Rasch analysis. *Language Assessment Quarterly: An International Journal, 2*(3), 197-221. https://doi.org/10.1207/s15434311laq0203_2

Elhan, A. H., & Atakurt, Y. (2005). Why is it necessary to use Rasch analysis when evaluating measures? *Ankara Üniversitesi Tıp Fakültesi Mecmuası, 58*(1), 47-50. https://doi.org/10.1501/Tipfak_0000000134

Evans, A. T., Mcnutt, R. A., Fletcher, S. W., & Fletcher, R. H. (1993). The characteristics of peer reviewers who produce good-quality reviews. *Journal of General Internal Medicine, 8*(8), 422-428. https://doi.org/10.1007/bf02599618

Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research, 70*(3), 287-322. https://doi.org/10.3102/00346543070003287

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), 1-10. https://doi.org/10.5688/ajpe81468

Gu, C. (2020). Student Peer Assessment. *Review of Educational Theory, 3*(2), 74-78. https://doi.org/10.30564/ret.v3i2.1762

Güler, N. (2014). Analysis of open-ended statistics questions with many facet Rasch model. *Eurasian Journal of Educational Research, 55*, 73-90. http://dx.doi.org/10.14689/ejer.2014.55.5

Gürlen, E. , Boztunç Öztürk, N. & Eminoğlu, E. (2019). Investigation of the reliability of teachers, self and peer assessments at primary school level with Generalizability Theory. *Journal of Measurement and Evaluation in Education and Psychology, 10*(4), 406-421. http://dx.doi.org/10.21031/epod.583891

Hanrahan, S. J., & Isaacs, G. (2001). Assessing self- and peer-assessment: The students' views. *Higher Education Research & Development, 20*(1), 53-70. https://doi.org/10.1080/07294360123776

Hoo, H. T., Deneen, C., & Boud, D. (2022). Developing student feedback literacy through self and peer assessment interventions. *Assessment & Evaluation in Higher Education, 47*(3), 444-457. https://doi.org/10.1080/02602938.2021.1925871

Köse, İ. A., Usta, H. G., & Yandı A. (2016). Evaluation of presentation skills by using many facets Rasch model. *Bolu Abant İzzet Baysal University Journal of Faculty of Education, 16*(4), 1853-1864. https://dergipark.org.tr/en/pub/aibuefd/issue/28550/304600

Linacre, J. M. (1993). Rasch-based generalizability theory. *Rasch Measurement Transactions, 7*(1), 283-284. https://www.rasch.org/rmt/rmt71h.htm

Linacre, J. M. (1994). *Many-facet rasch model* (2nd ed.). Mesa Press.

Linacre, J. M. (2023). A user's guide to FACETS: Rasch-model computer programs (Program manual 3.85. 1). Chicago, IL. https://www.winsteps.com/a/Facets-Manual.pdf

Liu, N. F., and D. Carless. 2006. "Peer feedback: The learning element of peer assessment." *Teaching in Higher Education 11*(3), 279-290. https://doi.org/10.1080/13562510600680582

Liu, E. Z.-F., Lin, S. S. J., Chiu, C.-H., & Yuan, S.-M. (2001). Web-based peer review: the learner as both adapter and reviewer. *IEEE Transactions on Education, 44*(3), 246-251. https://doi.org/10.1109/13.940995

Lu, J., & Law, N. (2011). Online peer assessment: effects of cognitive and affective feedback. *Instructional Science, 40*(2), 257-275. https://doi.org/10.1007/s11251-011-9177-2

Lundstrom, K., and W. Baker. 2009. "To give is better than to receive: The benefits of peer review to the reviewer's own writing." *Journal of Second Language Writing 18*(1), 30-43. https://doi.org/10.1016/j.jslw.2008.06.002

Mulqueen C., Baker D. & Dismukes R.K. (2000) *Using multifacet Rasch analysis to examine the effectiveness of rater training*. SIOP.

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

Nakamura, Y. (2000). Many-facet Rasch based analysis of communicative language testing results. *Journal of Communication Students, 12*, 3-13. https://eric.ed.gov/?id=ED449678

Nakamura, N. (2002). Teacher assessment and peer assessment in practice. *Educational Studies, 44*, 204-215. https://files.eric.ed.gov/fulltext/ED464483.pdf

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education, 31*(2), 199-218. https://doi.org/10.1080/03075070600572090

Orluwene, G. W., & Ekim, D. K. (2020). Promoting self-regulated learning through self-and peer-assessment techniques among secondary school students. *International Journal of Arts and Commerce, 9*(4), 1-16. https://ijac.org.uk/articles/9.4.1.1-16.pdf

Panadero, E., Alqassab, M., Fernández Ruiz, J., & Ocampo, J. C. (2023). A systematic review on peer assessment: intrapersonal and interpersonal factors. *Assessment & Evaluation in Higher Education*, 1-23. https://doi.org/10.1080/02602938.2023.2164884

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

Prieto, G., & Nieto, E. (2014). Analysis of rater severity on written expression exam using many faceted Rasch measurement. *Psicologica, 35*(2), 385-397. https://www.redalyc.org/pdf/169/16931314011.pdf

Randall, J. & Engelhard, G. Jr. (2009). Examining teacher grades using Rasch measurement theory. *Journal of Educational Measurement, 46*(1), 1-18. https://doi.org/10.1111/j.1745-3984.2009.01066.x

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. https://oa.upm.es/35804/

Semerci, Ç. (2011a). Doktora yeterlikler çerçevesinde öğretim üyesi, akran ve öz değerlendirmelerin Rasch ölçme modeliyle analizi. *Journal of Measurement and Evaluation in Education and Psychology, 2*(2), 164-171. https://dergipark.org.tr/en/pub/epod/issue/5804/77226

Semerci, Ç. (2011b). Analyzing microteaching applications with many-facet Rasch measurement model. *Education and Science, 36* (161), 14-25. http://egitimvebilim.ted.org.tr/index.php/EB/article/view/145

Semerci, Ç., Semerci, N., & Duman, B. (2013). Analysis of seminar presentation performances of postgraduate students with many-facet Rasch model. *The Journal of Sakarya Unıversity Education Faculty, 0*(25), 7-22. https://dergipark.org.tr/tr/pub/sakaefd/issue/11221/133978

Stefani, L. A. J. (1994). Peer, self and tutor assessment: Relative reliabilities. *Studies in Higher Education, 19*(1), 69-75. https://doi.org/10.1080/03075079412331382153

Sudweeks, R. R., Reeve, S., & Bradshaw, W. S. (2004). A comparison of generalizability theory and many-facet Rasch measurement in an analysis of college sophomore writing. *Assessing Writing, 9*(3), 239-261. https://doi.org/10.1016/j.asw.2004.11.001

Topping, K. J., Smith, E. F., Swanson, I., & Elliot, A. (2000). Formative Peer Assessment of Academic Writing Between Postgraduate Students. *Assessment & Evaluation in Higher Education, 25*(2), 149-169. https://doi.org/10.1080/713611428

Tseng, S.-C., & Tsai, C.-C. (2007). On-line peer assessment and the role of the peer feedback: A study of high school computer course. *Computers & Education, 49*(4), 1161-1174. https://doi.org/10.1016/j.compedu.2006.01.007

Turner, J. (2003). *Examining on art portfolio assessment using a many facet Rasch measurement model* [Unpublished dissertation]. Boston College.

Uyanık, G. K., Güler, N., Teker, G. T., & Demir, S. (2019). The analysis of elementary science education course activities through many-facet Rasch model. *Kastamonu Education Journal, 27*(1), 139-150. https://doi.org/10.24106/kefdergi.2417

Wright, B.D., & Linacre, J.M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions, 8*(3), 370-382. https://www.rasch.org/rmt/rmt83b.htm

Yan, Z., & Brown, G. T. (2017). A cyclical self-assessment process: Towards a model of how students engage in self-assessment. *Assessment & Evaluation in Higher Education, 42*(8), 1247-1262. https://doi.org/10.1080/02602938.2016.1260091

Yan, Z., Wang, X., Boud, D., & Lao, H. (2023). The effect of self-assessment on academic performance and the role of explicitness: a meta-analysis. *Assessment & Evaluation in Higher Education, 48*(1), 1-15. https://doi.org/10.1080/02602938.2021.2012644

Yüzüak, A. V., Yüzüak, B., & Kaptan, F. (2015). A many-facet Rasch measurement approach to analyze peer and teacher assessment for authentic assessment task. *Journal of Measurement and Evaluation in Education and Psychology, 6*(1), 1-11. https://doi.org/10.21031/epod.57425

Zhan, Y., Yan, Z., Wan, Z. H., Wang, X., Zeng, Y., Yang, M., & Yang, L. (2023). Effects of online peer assessment on higher-order thinking: A meta-analysis. *British Journal of Educational Technology, 54*, 817-835. https://doi.org/10.1111/bjet.13310