Peer Assessment Analysis via the Many-Facet Rasch Model

Burak AYÇİÇEK*

Abstract: Peer assessment has gained increasing attention in educational research as a dynamic approach. In this regard, this paper aims to analyze peer assessments of university students' drama skills using the many-facet Rasch model. In this study, quantitative research method with a descriptive design was used. The study group comprised 10 university students enrolled at a state university during the 2024-2025 academic year. Data were collected using the "Drama Skills Peer Assessment Form," which includes 32 items scored on a five-point Likert scale. The analysis was conducted using the FACETS program. Three facets were determined as student, assessor, and item. All participants conducted peer assessments for each other on the assessment form and 2880 (10x9x32) data were obtained. The findings revealed significant statistical differences in students' drama skills, assessors' strictness/generosity in scoring, and the levels of difficulty for the item. The study highlighted the need for further examination of the factors influencing peer assessment process.

Keywords: Peer assessment, drama skills, many-facet Rasch model, higher education, performance analysis.

Çok Yönlü Rasch Modeli ile Akran Değerlendirme Analizi

Öz: Akran değerlendirme, dinamik bir yaklaşım olarak eğitim araştırmalarında giderek daha fazla ilgi görmektedir. Bu bağlamda, bu çalışma üniversite öğrencilerinin drama becerilerine yönelik akran değerlendirmelerini çok yönlü Rasch modelini kullanarak analiz etmeyi amaçlamaktadır. Bu çalışmada betimsel desenli nicel araştırma yöntemi kullanılmıştır. Çalışma grubu, 2024-2025 akademik yılında bir devlet üniversitesinde öğrenim gören 10 üniversite öğrencisinden oluşmaktadır. Veriler, beşli Likert ölçeğine göre puanlanan 32 madde içeren "Drama Becerileri Akran Değerlendirme Formu" kullanılarak toplanmıştır. Analizler FACETS programı kullanılarak gerçekleştirilmiştir. Öğrenci, değerlendirici ve madde olmak üzere üç boyut belirlenmiştir. Tüm katılımcılar değerlendirme formu üzerinden birbirleri için akran değerlendirmeleri yapmış ve 2880 (10x9x32) veri elde edilmiştir. Bulgular, öğrencilerin drama becerilerinde, değerlendiricilerin puanlamadaki katılık/cömertliklerinde ve maddelerin zorluk düzeylerinde anlamlı istatistiksel farklılıklar olduğunu ortaya koymuştur. Çalışma, akran değerlendirme sürecini etkileyen faktörlerin daha fazla incelenmesi gerektiğini göstermiştir.

Anahtar Kelimeler: Akran değerlendirme, drama becerileri, çok yönlü Rasch modeli, yükseköğretim, performans analizi.

*Sorumlu yazar Doç. Dr., Tokat Gaziosmanpaşa Üniversitesi, Eğitim Fakültesi, Tokat-Türkiye, ORCID:0000-0001-8950-2207, e-posta: burak.aycicek@gop.edu.tr

Introduction

Foreign language teaching has gained significant global importance, emphasizing the necessity of incorporating activities such as drama, engaging tasks, and interactive activities to enhance the effectiveness of the learning process. Individuals explore and acquire knowledge about their environment through sensory experiences including observing, touching, smelling, hearing, and tasting (Sengül & Sünbül, 2015). Among educators, it is widely acknowledged that successful instructional activities cannot be achieved without interaction (Yao, 2017). Hence, an essential responsibility of foreign language teachers is to create an active and engaging learning environment (Yao & Shao, 2024), which not only enhances the quality of teaching but also improves learning processes and outcomes (Hagenauer et al., 2023).

Student engagement is considered as an important factor and supported by a strong commitment to its enhancement (Bradley et al., 2015; Lubicz-Nawrocka & Bunting, 2019; Mercer & Dornyei, 2020). Moreover, student engagement, recognized as effective presentation skills, strong communication abilities, and practical application of knowledge (Bradley et al., 2015; Heffernan et al., 2010; Jin, 2000) constitute fundamental components of successful educational practices. Building on the critical role of engagement and interaction in foreign language learning, drama, as an effective method for enabling students to articulate their thoughts and emotions (Isyar & Akay, 2017), emerges as a powerful pedagogical tool.

Barrera et al. (2021) argued that enhancing the effectiveness of student learning across all educational levels represents a critical challenge of the 21st century. In addressing this challenge, drama provides a dynamic and interactive approach to support student learning. Pinciotti (1993) defined drama as a structured learning activity facilitated by a leader, providing participants with the opportunity to imagine, enact, and reflect on human experiences. Fleming (2006) argues that drama texts support a learner-centered approach to classroom practice, allowing learners to perceive, think, act, and interact actively during the learning process, rather than remaining passive recipients of knowledge.

Drama techniques are valuable tools for fostering active student engagement in the learning process by incorporating excitement, fun, and creativity into the classroom, encouraging collaboration, and immersing learners in real-life discourse models that they internalize through imitation and performance (Bessadet, 2022). Beyond promoting engagement, drama functions as an active methodology that integrates cognitive and affective dimensions, cultivating playfulness, positive emotions, adaptability, fluency, communicative competence, and a supportive environment that strengthens the student-teacher relationship (Belliveau & Kim, 2013; Garaigordobil & Berrueco, 2011; Hammond, 2015; Maley & Duff, 2005). In addition, it offers learners opportunities to reflect on their thoughts, emotions, and understanding in relation to others within a collaborative group context (Adams & Owens, 2016). Drama and drama-based activities facilitate the development of competencies in each learner (Asimidou et al., 2021). At this point, drama might play a significant role in peer assessment by fostering mutual understanding, empathy, and constructive feedback among students.

Peer assessment as a formative assessment practice, has gained substantial recognition in educational research (Badea & Popescu, 2022; Double et al., 2020; Topping, 1998). Peer assessment is broadly defined as a process in which individuals evaluate the level or quality of their peers' learning outcomes, typically among those of similar status (Reinholz, 2016). It involves students assessing and providing feedback on one another's learning tasks (Topping, 1998) and encompasses a range of activities such as evaluation, reviewing, and grading (Adachi et al., 2018). As a pedagogical strategy, it encourages students to offer feedback aligned with pre-determined criteria, which fosters engagement and reflective learning (Topping, 1998). Peer assessment in higher education can provide learning outcomes similar to those from teacher assessment. It also helps students learn more effectively and improve their overall academic performance (Jongsma et al., 2023; Noroozi et al., 2023; Zheng et al., 2019).

Peer assessment has been widely recognized as an effective pedagogical strategy with numerous benefits across various educational contexts. It fosters a sense of belonging within the classroom (Brown et al., 2021; Devisakti & Ramayah, 2023) and enhances student motivation (Deeley & Bovill, 2015). By promoting interaction, peer assessment supports the establishment of a partnership between the assessor and the assessee (Winstone & Carless, 2020), thereby enhancing students' engagement and active participation in class (Conde et al., 2017; Deng et al., 2023).

Peer assessment supports collaborative learning processes (Prins et al., 2005) and facilitates self-regulated learning in various contexts (Bijami et al., 2013; Topping, 2009). It positively impacts students' metacognitive skills, such as problem identification (Topping, 1998), and their ability to make evaluative judgments (Panadero & Broadbent, 2018). Furthermore, peer assessment has demonstrated its effectiveness in improving academic performance across diverse educational levels and subject areas (Double et al., 2020), as well as in developing critical thinking abilities and higher-order thinking skills (Gaynor, 2019; Guelfi et al., 2021; Humberstone et al., 2024; Jiang et al., 2022; Zhan et al., 2023). Beyond cognitive outcomes, peer assessment has been shown to benefit students' learning abilities by fostering

independent learning skills, professional skill development (Smith et al., 2002), enhancing reflective thinking (Gaynor, 2019), and increasing writing motivation (Weng et al., 2023). Additionally, it contributes to metacognition (Kim & Ryu, 2013), and learner cognition (Panadero & Alqassab, 2019). These multifaceted benefits make it a valuable tool for promoting both individual and collaborative learning outcomes in higher education.

Based on the information above, it is seen that drama technique stands out as an effective method in education processes. Although drama-based activities have gained significant interest in higher education institutions, it has not been explored extensively. In this context, there is a clear need for further investigation to assess drama activities based on peer assessment. This study seeks to contribute valuable insights by facilitating peer assessment of drama performances among university students. The results and insights derived from this study may be of interest to broader academic circles in higher education.

As a result, this study aims to investigate peer assessments of students' drama skills through the many-facet Rasch model. To this end, analyses were conducted on the ability analysis related to drama skills, analysis of assessors' scoring tendencies, item/criterion difficulty analysis of drama skills, and analysis of assessors' biases.

Method

Research Model

Since the study aims to determine students' ability levels, scoring tendencies (generosity or strictness), difficulty levels of the items, and biases of the assessors, it uses a quantitative research method with a descriptive design. According to Creswell (2014), descriptive research is an approach in which the researcher presents the current situation as it is and analyzes the situations without any manipulation.

Study Group

The study group comprised 10 university students, including three males and seven females. They were third-year students in the Department of English Language Teaching at a state university during the 2024-2025 academic year. The participants were selected voluntarily from students who had taken part in drama activities in an English course.

Data Collection Process

In the first stage of the study, detailed information about the objectives and process of peer assessment and sample applications were given to all students taking the course. Then, 10 students were determined and these students were separated into three groups comprising three, three, and four members, respectively. All 10 students participating in the study performed drama activities and made assessments for all their peers. The "Drama Skills Peer Assessment Form" constructed by the researcher was used to score the drama skills of the students. The peer assessment form consists of 32 items, each rated on a five-point Likert scale from "very inadequate" (1) to "very adequate" (5). This study was carried out with the approval of Tokat Gaziosmanpaşa University Ethics Council for the Research in Social and Human Sciences dated 14.01.2025 and numbered 01.56.

Data Analysis

In this study, peer assessments conducted by students were assessed through the many-facet Rasch model which allows simultaneous and independent analysis of multiple sources of variability, such as ability, items, and raters (Sudweeks et al., 2004). FACETS programme developed by Linacre (1993) was utilized for data analysis. In the study, a total of three facets were determined as student (10 assessee), assessor (9 assessors) and item (32 items). Therefore, 10 participants performed a drama activity and all participants conducted peer assessments for each other on a 32-item assessment form. In total, 2880 (10x9x32) data were obtained. 10 participants both performed the drama activity as a student and scored their peers as an assessor. In this context, the same code number represents the same person in the coding made for the student and the assessor (For example, Student 1 and Assessor 1 are the same person).

Findings

Wright Map representing three facets (student, assessor and item) is shown in Figure 1.

Figure 1.

Wright Map of Students, Assessors and Items

```
+-----------------------------------------------------------------------+
|Measr|+Student|-Assessor                 |-Item                |Scale|
|-----+--------+--------------------------+---------------------+-----|
|  2 +        +          +                 +                    + (5) |
|    |        |          |                 |                    | --- |
|    |        |          |                 |                    |     |
|    |        |          |                 |                    |     |
|    |        |          |                 |                    |     |
|    |        |          |                 |                    |     |
|    |        |          |                 |                    |     |
|    |        |          |                 | 32                 |     |
|  1 +        +          +                 +                    +  4  |
|    |        |          |                 | 20                 |     |
|    |        |          |                 | 12     9           |     |
|    |        |          |                 | 8    23    16      |     |
|    |     4  |          |                 | 5    26    28    4 |     |
|    |    10  |          |                 | 31                 | --- |
|    |     2    5 |      |                 |                    |     |
|    |        |          | 2    4    5    8 | 18    17          |     |
|    |     6    9 |   6  |                 | 27                 |     |
|  * 0 *   7    8 |  * 10    7    9         * 21                 *     *
|    |     3  |          |                 | 19    29    25     |  3  |
|    |        |          |                 | 24     7           |     |
|    |     1  |          |   3             | 22                 |     |
|    |        |          |   1             | 11     6           |     |
|    |        |          |                 | 10     3           | --- |
|    |        |          |                 | 13                 |     |
|    |        |          |                 | 14    15           |     |
|    |        |          |                 | 2                  |     |
| -1 +        +          +                 + 1                  +  2  |
|    |        |          |                 | 30                 |     |
|    |        |          |                 |                    |     |
|    |        |          |                 |                    |     |
|    |        |          |                 |                    | --- |
|    |        |          |                 |                    |     |
|    |        |          |                 |                    |     |
| -2 +        +          +                 +                    + (1) |
|-----+--------+--------------------------+---------------------+-----|
|Measr|+Student|-Assessor                 |            * = 1|Scale|
+-----------------------------------------------------------------------+
```
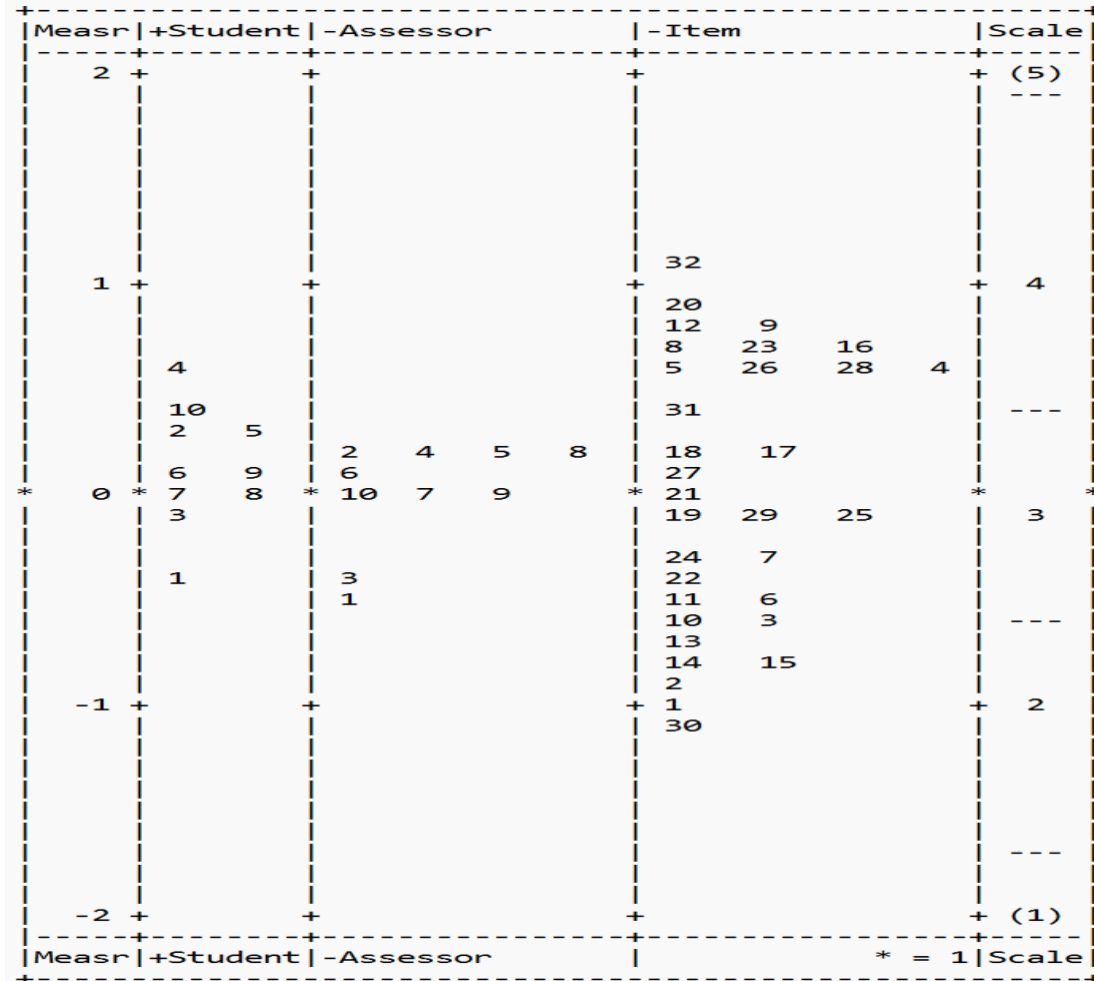
Figure 1 presents the ranking of students participating in drama activities on the same logit scale. As seen, students are ranked by their drama performance, from the highest to the lowest. Similarly, assessors are ranked based on their scoring tendencies, ranging from the strictest to the most generous, and items are ordered from the most difficult to the easiest. Accordingly, Student 4 had the highest performance in terms of the drama activity performed, while Student 1 had the lowest performance. However, Assessor 2, Assessor 4, Assessor 5 and Assessor 8 performed the strictest assessment, while Assessor 1 performed the most generous assessment. In addition, It is evident that the most difficult is Item 32 and the easiest is Item 30. The measurement report detailing the participants' drama skills is provided in Table 1.

Table 1.

Report on Students' Drama Skills

Student	Logit	Standart Error (SE)	Infit MnSq	Outfit MnSq
4	0.62	0.07	1.15	1.10
10	0.36	0.06	0.96	0.99
5	0.32	0.06	0.90	0.96
2	0.27	0.06	0.94	0.96
9	0.10	0.06	1.07	1.09
6	0.09	0.06	1.03	1.03
8	0.04	0.06	1.02	1.07
7	0.00	0.06	1.21	1.20
3	-0.09	0.06	0.95	0.99
1	-0.37	0.06	0.69	0.69
RMSE= 0.06	sd= 0.25	Separation= 4.08	Strata= 5.78	Reliability= 0.94
chi-square= 172.9	df= 9	*p*= 0.00		

Table 1 reveals that the root mean square standard error (RMSE) is 0.06, while the standard deviation is 0.25. Additionally, the seperation index, which reflects the measurement tool's ability to differentiate between varying ability levels (Linacre, 1994) and is considered more effective when higher (Mumpuni et al., 2022), is calculated as 4.08. Furthermore, the strata value is determined as 5.78, indicating approximately six distinct groups of students based on their drama skills. The reliability coefficient demonstrates that the ranking of students' drama skills has a confidence level of 0.94. Based on the chi-square test results, a significant statistical difference is observed among students regarding their drama skills. Thus, the null hypothesis is rejected. The overall ranking of students, from highest to lowest performance in drama skills, is as follows: Student 4, 10, 5, 2, 9, 6, 8, 7, 3, 1.

In the analysis, the compatibility between the data and the model was evaluated using the infit and outfit values of the facets. Fit statistics within the range of 0.5 to 1.5 are considered indicative of precise and reliable measurements (Wright & Linacre, 1994). Based on this criterion, the model shows a good fit with the data derived from all students.

Table 2 provides the measurement report for the assessors.

Table 2.

Assessors' Strictness/Generosity Report

Assessor	Logit	SE	Infit MnSq	Outfit MnSq
4	0.23	0.06	0.94	0.96
5	0.18	0.06	0.97	0.98
8	0.17	0.06	0.75	0.76
2	0.15	0.06	1.17	1.16
6	0.14	0.06	1.02	1.04
7	0.04	0.06	0.80	0.81
9	0.03	0.06	0.84	0.83
10	0.00	0.06	1.07	1.08
3	-0.44	0.06	1.17	1.19
1	-0.50	0.07	1.19	1.26
RMSE= 0.06	sd= 0.24	Separation= 3.88	Strata= 5.50	Reliability= 0.94
chi-square= 148.1	df= 9	*p*= 0.00	Inter-rater exact agreements= 36.3%	

In Table 2, Assessor 4, measured at 0.23, was identified as the strictest. However, Assessor 1 was the most generous, measured at -0.50. The overall ranking of assessors, from the strictest to the most generous, is as follows: Assessor 4, 5, 8, 2, 6, 7, 9, 10, 3, and 1. Additionally, the RMSE value was determined to be 0.06, the standard deviation 0.24, the separation index 3.88, and the strata value 5.50. A reliability coefficient of 0.94 indicates that the assessors were ranked with a very high level of reliability in terms of their strictness or generosity. The chi-square test results led to the rejection

of the null hypothesis, indicating a significant statistical difference among the assessors related to their scoring strictness or generosity. An analysis of infit and outfit values for the assessor facet shows that the data fits with the model for all assessors. However, the inter-rater exact agreement rate was found to be 36.3%.

Table 3 presents a detailed measurement report for the items.

Table 3.

Report of the Items Used in the Assessment of Drama Skills

Item	Logit	SE	Infit MnSq	Outfit MnSq
32	1.11	0.11	1.20	1.22
20	0.92	0.10	0.99	0.98
12	0.83	0.10	1.01	1.01
9	0.76	0.10	0.95	0.94
8	0.75	0.10	0.98	0.98
23	0.72	0.10	0.94	0.94
16	0.65	0.10	0.91	0.92
5	0.64	0.10	0.82	0.83
26	0.61	0.10	1.04	1.04
28	0.58	0.10	1.50	1.50
4	0.57	0.10	0.96	0.96
31	0.41	0.10	1.19	1.22
18	0.19	0.10	0.97	0.96
17	0.12	0.10	0.46	0.46
27	0.04	0.11	1.74	1.73
21	-0.01	0.11	0.55	0.55
19	-0.09	0.11	0.98	0.96
29	-0.10	0.11	0.84	0.88
25	-0.14	0.11	0.35	0.38
7	-0.30	0.11	0.70	0.72
24	-0.30	0.11	1.04	1.03
22	-0.35	0.11	0.86	0.84
11	-0.47	0.11	1.24	1.24
6	-0.54	0.12	0.95	0.96
10	-0.56	0.12	1.03	1.07
3	-0.61	0.12	0.94	0.97
13	-0.74	0.12	0.99	1.03
14	-0.77	0.12	1.01	1.01
15	-0.84	0.13	0.93	0.92
2	-0.94	0.13	1.06	1.08
1	-1.02	0.13	0.82	0.86
30	-1.11	0.13	1.91	2.06
RMSE= 0.11	sd= 0.63	Separation= 5.67	Strata= 7.89	Reliability= 0.97
chi-square= 1005.6	df= 31	*p*= 0.00		

As seen in Table 3, the most difficult item is 32: *"Shows original improvisation skills,"* measured at 1.11. Then, Item 20: *"Adapts seamlessly to unexpected situations"* and Item 12: *"Demonstrates smooth transitions between movements."* follows this. In contrast, the easiest is Item 30: *"Demonstrates fundamental skills with no creativity involved,"* measured at -1.11. Then, Item 1: *"Keeps a consistently loud voice"* and Item 2: *"Maintains a clear voice."* follows this.

The RMSE value was calculated as 0.11, with a standard deviation of 0.63. The strata value (7.89) indicates that the items in the rubric are divided into eight distinct groups based on difficulty levels. Additionally, the reliability value was

found to be very high at 0.97. A significant statistical difference among the levels of difficulty for the items was identified, as evidenced by the results of the chi-square test. According to the fit statistics for the items, except for Item 27: *"Exemplifies collaboration"* and Item 30: *"Demonstrates fundamental skills with no creativity involved"*, all other items are between acceptable the infit and outfit values. Accordingly, it can be said that the 27th and 30th items prevent data-model fit.

A sample of unexpected responses are provided in Table 4.

Table 4.

Unexpected Responses

Sequence	Score	Expected	StRes	Student	Asseessor	Item
610	1	4.0	-3.5	3	2	2
640	5	1.9	3.4	3	2	32
536	1	3.9	-3.3	2	9	24
446	2	4.3	-3.2	2	6	30
414	2	4.3	-3.1	2	5	30
858	5	2.0	3.1	3	10	26
1226	2	4.3	-3.1	5	3	10

The standardized residual (StRes) values in Table 4 reveals that some values are marked with a minus sign, while others have a plus sign. This indicates that unexpected data arise from some assessors scoring lower than expected to certain students, while others scored higher than expected scores. The most significant unexpected data is linked to the score provided by Assessor 2 for Item 2 to Student 3. Specifically, for Item 2: *"Maintains a clear voice,"* the expected score was 4.0; however, Assessor 2 assigned a score of 1, resulting in a StRes value of -3.5. Also, Assessor 2 scored Student 3 above the expected value for Item 32: *"Shows original improvisation skills."* As an other finding of the study, Assessor 10 also scored Student 3 above the expected value for Item 26: *"Maintains professionalism."* Therefore, it was determined that only Student 3 was scored above the expected value. However, it is seen in Table 1 that Student 3 had a low performance in terms of drama skills. Accordingly, it is clear that some items were scored biased in the assessments made for Student 3. However, it is seen that three different assessors scored below the expected value for Student 2, two of which were the scores for Item 30*: "Demonstrates fundamental skills with no creativity involved."*

Table 5 indicates the bias analysis of the students and assessors.

Table 5.

Report on Interaction Bias for Student and Assessor

Observed Score	Expected Score	Obs-Exp Average	Bias	SE	*z* Score	Infit MnSq	Outfit MnSq	Student	Assessor
131	96.81	1.07	1.29	0.22	5.80	1.4	1.3	6	2
130	98.83	0.97	1.17	0.22	5.35	1.9	2.5	7	10
136	112.61	0.73	1.02	0.24	4.29	0.6	0.7	4	2
144	128.86	0.47	0.89	0.28	3.20	2.0	1.5	4	1
130	111.77	0.57	0.74	0.22	3.37	1.2	1.2	4	5
120	102.90	0.53	0.61	0.20	3.06	1.4	1.8	10	4
134	122.12	0.37	0.55	0.23	2.38	1.5	2.0	5	1
123	108.47	0.45	0.54	0.20	2.66	0.9	0.9	10	7
128	116.41	0.36	0.48	0.21	2.23	0.7	0.6	9	1
114	100.06	0.44	0.47	0.19	2.49	1.2	1.3	8	10
124	112.07	0.37	0.46	0.21	2.23	1.3	1.3	7	3
119	107.85	0.35	0.40	0.20	2.05	1.0	1.0	5	9
106	94.60	0.36	0.37	0.18	2.03	1.2	1.2	7	6
102	113.06	-0.35	-0.38	0.18	-2.08	0.5	0.5	4	6
81	93.63	-0.39	-0.40	0.18	-2.24	0.7	0.8	7	8
81	94.18	-0.41	-0.42	0.18	-2.32	1.1	1.1	7	2

104	116.76	-0.40	-0.45	0.18	-2.46	0.9	0.9	4	10
100	113.70	-0.43	-0.46	0.18	-2.58	0.9	0.9	7	1
90	105.03	-0.47	-0.48	0.18	-2.72	0.4	0.4	10	2
86	101.52	-0.48	-0.50	0.18	-2.78	0.7	0.6	6	10
76	93.14	-0.54	-0.55	0.18	-3.02	0.9	1.1	7	5
77	95.33	-0.57	-0.59	0.18	-3.23	1.0	1.1	8	2
98	116.11	-0.57	-0.62	0.18	-3.45	0.7	0.6	4	9
89	108.57	-0.61	-0.64	0.18	-3.57	0.7	0.7	5	10
91	111.22	-0.63	-0.66	0.18	-3.73	1.0	1.0	3	1
77	97.72	-0.65	-0.66	0.18	-3.65	2.3	2.4	9	6
70	91.30	-0.67	-0.70	0.19	-3.73	1.6	2.3	3	2
97	120.68	-0.74	-0.83	0.18	-4.61	0.8	0.8	5	3
chi-square= 370.4		df= 90		p= 0.00					

Z-scores that fall outside the range of -2 to +2 are generally interpreted as interaction bias. In Table 5, Assessor 2 was significantly (z=5.80; p<0.05) generous by giving 131 points to Student 6 when the assessor should have given approximately 97 points, and Assessor 1 was significantly (z=-3.73; p<0.05) strict by giving 91 points to Student 3 when the assessor should have given approximately 111 points to Student 3. As an other finding, the most recurring raters in the student-assessor interaction bias were Assessor 2, Assessor 1, Assessor 10, while the most recurring students were Student 7, Student 4 and Student 5. It is seen that Assessor 2 made a significantly generous peer assessment for Student 4 and Student 6, and a significantly strict peer assessment for Student 7, Student 8 and Student 10. Similarly, Assessor 1 was significantly generous for Student 4, Student 5 and Student 9, and significantly strict for Student 3 and Student 7.

Discussion and Conclusion

In the current study, the peer assessment results of 10 students' drama skills, assessed using a 32-item rubric, were assessed through the many-facet Rasch model. The results of the study indicated that the overall ranking of students' drama skills from the highest to the lowest performance is Student 4, 10, 5, 2, 9, 6, 8, 7, 3, 1. This variability may be influenced by students' perceptions of peer assessment, which are closely tied to their commitment to the process (Patchan et al., 2016). Student role-play, as a key component of drama activities, fosters dynamism in the learning environment (Stevens, 2015) and enhances interest in the subject matter (Poitras et al., 2013), potentially contributing to greater engagement and improved performance. In this regard, it may be wise to mention that students who achieved high performance, such as Student 4 and Student 10, likely demonstrated strong skills in areas like creativity, expression, and engagement during the drama activity. Students who achieved lower performance, such as Student 1 and Student 3, may have faced challenges in meeting the drama activity items. These challenges could include difficulties in creativity, emotional expression, or confidence. In this context, Zhan (2021) argued that students may have negative experiences in peer assesment process.

The overall ranking of assessors from the strictest to the most generous is Assessor 4, 5, 8, 2, 6, 7, 9, 10, 3, and 1. The results suggest that assessor bias may influence the assessment process. Assessor 4, as one of the strictest raters, may reflect higher personal standards, possibly due to their interpretation of the items or a preference for perfection. In contrast, Assessor 1, as the most generous rater, may demonstrate flexibility or a lower standard for recognizing achievement, potentially influenced by empathy, a desire to encourage peers, or differing perceptions of performance expectations. These contrasting approaches raise questions about fairness in peer assessment (Reddy et al., 2021; Wilson et al., 2015).

According to the item measurement values, the most difficult is Item 32: *"Shows original improvisation skills,"* measured at 1.11. It is clear that Item 32 represents an aspect of the drama activity that requires advanced skills, creativity, and higher-order thinking. At this point, Seppanen (2022) concluded that improvisation training significantly enhances preservice teachers' interpersonal communication skills and confidence, suggesting that developing improvisation skills is both challenging and highly beneficial. By fostering metacognitive thinking skills, peer assessment can help students better understand and meet the demands of advanced items (Armengol-Asparó et al., 2022) like Item 32. Then, Item 20: *"Adapts seamlessly to unexpected situations"* and Item 12: *"Demonstrates smooth transitions between movements."* follows this. Item 20, proved challenging, likely due to the unpredictability of such scenarios. Students may lack experience in handling unstructured situations. Item 12 may be difficult due to the physical and technical demands

of ensuring a seamless flow in performance. Achieving smooth transitions requires a deep understanding of movement dynamics, coordination, and practice. In this sense, implementing student-centered learning approaches can address these challenges by allowing students to take an active role in their own skill development (Nicholson, 2009). In contrast, the easiest is Item 30: *"Demonstrates fundamental skills with no creativity involved,"* measured at -1.11. Then, Item 1: *"Keeps a consistently loud voice"* and Item 2: *"Maintains a clear voice."* follows this. These findings may be due to the fact that basic skills are more accessible for students to achieve, as they require less abstract thinking and creativity. It can be said that these skills can be explicitly taught and practiced, making them easier to demonstrate compared to more complex or subjective items.

The strata value calculated as approximately eight show that the items in the rubric are divided into eight groups in terms of their difficulty. It can be interpreted that this finding reflects the analytical rubric used for peer assessment, which consists of eight headings. According to the fit statistics for the items, except for Item 27: *"Exemplifies collaboration"* and Item 30: *"Demonstrates fundamental skills with no creativity involved"*, all other items are between acceptable the infit and outfit values. The findings suggest that these two items may not align well with the overall model, potentially due to differences in how they are interpreted or applied during assessment. In addition to these points, high reliability was calculated for all facets included in the Rasch analysis.

Based on unexpected responses, it is identified that the most unexpected data is due to the score provided by Assessor 2 to Item 2 for Student 3. Furthermore, Assessor 2 scored Student 3 above the expected value for Item 32. This situation can be interpreted as Assessor 2, who was the most repetitive in terms of providing unexpected scores, showed a low performance in terms of peer assessment. As a remarkable finding of the study, Assessor 10 also scored Student 3 above the expected value for Item 26. Therefore, it was determined that only Student 3 was scored above the expected value. However, it is seen that Student 3 had a low performance in terms of drama skills. Accordingly, it can be said that some items were scored biased in the assessments made for Student 3. However, it is seen that three different assessors scored below the expected value for Student 2, two of which were the scores for Item 30. As a result, it can be said that Assessor 2 was the most recurrent rater in assigning unexpected scores, Student 2 and Student 3 were the most recurrent in receiving unexpected scores, and Item 30 was the most recurrent item for bias.

One of the remarkable findings of the study is that the most recurring raters in the student-assessor interaction bias were Assessor 2, Assessor 1, Assessor 10, while the most recurring students were Student 7, Student 4 and Student 5. It is seen that Assessor 2 made a significantly generous peer assessment for Student 4 and Student 6, and a significantly strict peer assessment for Student 7, Student 8 and Student 10. Similarly, Assessor 1 was significantly generous for Student 4, Student 5 and Student 9, and significantly strict for Student 3 and Student 7. Assessor 10 demonstrated a notably generous approach in the peer assessments for Student 7 and Student 8 and a notably strict approach in the peer assessments for Student 4, Student 5 and Student 6. Accordingly, Assessor 10 gave generous scores for his group mates (Student 7 and Student 8), while the assessor gave strict scoring for all members of another group (Student 4, Student 5 and Student 6). These findings suggest potential group-based bias in peer assessment, where assessors may favor peers they work closely with, potentially due to familiarity or loyalty. Conversely, stricter assessments for members of other groups might stem from a lack of collaboration or perceived competition. In this sense, it is vital to emphasize that peer assessment, when effectively implemented, can mitigate such issues by fostering teamwork and collaboration (Reinholz, 2016; Zheng et al., 2019).

Limitations and Recommendations for Further Research

The study was conducted with a relatively small sample of 10 university students from a single state university, which may limit the generalizability of the findings. Future research could expand the sample size and include students from diverse academic and cultural backgrounds to enhance the generalizability of the findings. In the study, students' peer assessments were collected at one point in time. This approach can be insufficient to reflect how their drama skills, scoring tendencies, or understanding of assessment items might change over time. Therefore, longitudinal studies can be conducted to provide deeper insights into the development of students' drama skills over time through repeated peer assessments. In addition, the study relied solely on quantitative analysis. The lack of qualitative data limits the understanding of students' perceptions regarding peer assessment. In this regard, incorporating qualitative methods in future research could enrich the interpretation of the findings. Overall, there is a need for further research to enhance the implementation of peer assessment in drama-based education, particularly to ensure its validity, reliability, and pedagogical effectiveness within higher education settings.

Ethics Committee Approval Information

This study was carried out with the approval of Tokat Gaziosmanpaşa University Ethics Council for the Research in Social and Human Sciences dated 14.01.2025 and numbered 01.56.

Conflict of Interest

No potential conflict of interest was reported by the author.

Financial Support

No financial support was received for this study.

References

Adachi, C., Tai, J., & Dawson, P. (2018). A framework for designing, implementing, communicating and researching peer assessment. *Higher Education Research & Development*, *37*(3), 453-467. https://doi.org/10.1080/07294360.2017.1405913

Adams, J., & Owens, A. (2016). *Creativity, education and democracy: The practices and politics of learning in the arts*. Routledge.

Armengol-Asparó, C., Mercader, C., & Ion, G. (2022). Making peer-feedback more efficient: What conditions of its delivery make the difference? *Higher Education Research & Development*, *41*(2), 226-239. https://doi.org/10.1080/07294360.2020.1840527

Asimidou, A., Lenakakis, A., & Tsiaras, A. (2021). The contribution of drama pedagogy in developing adolescents' self-confidence: A case study. *NJ: Drama Australia Journal*, *45*(1), 45-58. https://doi.org/10.1080/14452294.2021.1978145

Badea, G., & Popescu, E. (2022). A dynamic review allocation approach for peer assessment in technology enhanced learning. *Education and Information Technologies, 27*(9), 13131-13162. https://doi.org/10.1007/s10639-022-11175-5

Barrera, F., Venegas-Muggli, J. I., & Nuñez, O. (2021). The impact of role-playing simulation activities on higher education students' academic results. *Innovations in Education and Teaching International*, *58*(3), 305-315. https://doi.org/10.1080/14703297.2020.1740101

Belliveau, G., & Kim, W. (2013). Drama in L2 learning: A research synthesis. *Scenario 7*(2), 7-27. https://doi.org/10.33178/scenario.7.2.2

Bessadet, L. (2022). Drama-based approach in English language teaching. *Arab World English Journal, 13*(1) 525-533. https://dx.doi.org/10.24093/awej/vol13no1.34

Bijami, M., Kashef, S. H., & Nejad, M. S. (2013). Peer feedback in learning English writing: Advantages and disadvantages. *Journal of Studies in Education, 3*(4), 91. https://doi.org/10.5296/jse. v3i4.431

Bradley, S., Kirby, E., & Madriaga, M. (2015). What students value as inspirational and transformative teaching. *Innovations in Education and Teaching International*, *52*(3), 231-242. https://doi.org/10. 1080/14703297.2014.880363

Brown, T., Rongerude, J., Leonard, B., & Merrick, L. C. (2021). Best practices for online team-based learning: Strengthening teams through formative peer evaluation. *New Directions for Teaching and Learning, 2021,* 53-64. https://doi.org/10.1002/tl.20436

Conde, M. A., Sanchez-Gonzalez, L., Matellan-Olivera, V., & Rodriguez-Lera, F. J. (2017). Application of peer review techniques in engineering education. *International Journal of Engineering Education, 33*(2), 918-926. https://robotica.unileon.es/vmo/pubs/ijeee2017.pdf

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Sage Publications.

Deeley, S. J., & Bovill, C. (2015). Staff student partnership in assessment: enhancing assessment literacy through democratic practices. *Assessment & Evaluation in Higher Education, 42*(3), 463-477. https://doi.org/10.1080/02602938.2015.1126551

Deng, Y., Liu, D., & Feng, D. (2023). Students' perceptions of peer review for assessing digital multimodal composing: The case of a discipline-specific English course. *Assessment & Evaluation in Higher Education*, *48*(8), 1254-1267. https://doi.org/10.1080/02602938.2023.2227358

Devisakti, A., & Ramayah, T. (2021). Sense of belonging and grit in e-learning portal usage in higher education. *Interactive Learning Environments*, *31*(8), 4850-4864. https://doi.org/10.1080/10494820.2021.1983611

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*, 481-509. https://doi.org/10.1007/s10648-019-09510-3

Fleming, M. (2006). Drama and language teaching: the relevance of Wittgenstein's concept of language games. *Humanising Language Teaching, 8*(4), 1-12. http://old.hltmag.co.uk/jul06/mart01.htm

Garaigordobil, M., & Berrueco, L. (2011). Effects of a play program on creative thinking of preschool children. *The Spanish Journal of Psychology, 14*(2), 608-618. https://doi.org/10.5209/rev_SJOP.2011.v14.n2.9

Gaynor, J. W. (2019). Peer review in the classroom: Student perceptions, peer feedback quality and the role of assessment. *Assessment & Evaluation in Higher Education, 45*(5), 758-775. https://doi.org/10. 1080/02602 938. 2019.1697424

Guelfi, M. R., Formiconi, A. R., Vannucci, M., Tofani, L., Shtylla, J., & Masoni, M. (2021). Application of peer review in a university course: Are students good reviewers? *Journal of E-Learning and Knowledge Society, 17*(2), 1-8. https://doi.org/10.20368/1971-8829/1135380

Hagenauer G., Muehlbacher F., & Ivanova, M. (2023)."It's where learning and teaching begins - is this relationship"- insights on the teacher-student relationship at university from the teachers' perspective. *Higher Education, 85*, 819-835. https://doi.org/10.1007/s10734-022-00867-z

Hammond, N. (2015). *Forum theatre for children: Enhancing social, emotional and creative development*. Institute of Education Press.

Heffernan, T., Morrison, M., Sweeney, A., & Jarratt, D. (2010). Personal attributes of effective lecturers: The importance of dynamism, communication, rapport and applied knowledge. *The International Journal of Management Education*, *8*(3), 13-27. https://doi.org/10.3794/ijme.83.275

Humberstone, J. H. B., Zhao, C., Liu, D., & Elizabeth, M. (2024). Stimulating music educator worldview change through a pedagogy of provocation, critical thinking and peer-review. *International Journal of Music Education, 42*(1), 47-67. https://doi.org/10.1177/02557614221149223

Isyar, O., & Akay, C. (2017). The use of drama in education in primary schools from the viewpoint of the classroom teachers: A mixed method research. *Journal of Education & Practice*, *8*(28), 215-230. https://files.eric.ed.gov/fulltext/ED578395.pdf

Jiang, J. P., Hu, J. Y., Zhang, Y. B., & Yin, X. C. (2022). Fostering college students' critical thinking skills through peer assessment in the knowledge building community. *Interactive Learning Environments*, *31*(10), 6480-6496. https://doi.org/10.1080/10494 820. 2022.20399 49

Jin, Z. (2000). The learning experience of students in Middlesex University Business School (MUBS): Why do they enjoy some modules/lectures and dislike others? *The International Journal of Management Education*, *1*(1), 22-36. https://doi.org/10.3794/ijme.11.c

Jongsma, M. V., Scholten, D. J., van Muijlwijk-Koezen, J. E., & Meeter, M. (2023). Online versus offline peer feedback in higher education: A meta-analysis. *Journal of Educational Computing Research*, *61*(2), 329-354. https://doi.org/10.1177/07356331221114181

Kim, M., & Ryu, J. (2013). The development and implementation of a web-based formative peer assessment system for enhancing students' metacognitive awareness and performance in ill-structured tasks. *Educational Technology Research and Development, 61*, 549-561. https://doi.org/10.1007/s11423-012-9266-1

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

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

Lubicz-Nawrocka, T., & Bunting, K. (2019). Student perceptions of teaching excellence: An analysis of student-led teaching award nomination data. *Teaching in Higher Education*, *24*(1), 63-80. https://doi.org/10.1080/13562517.2018.1461620

Maley, A., & Duff, A. (2005). *Drama techniques: A resource book of communication activities for language teachers*. Cambridge University Press. https://doi.org/10.1017/CB09780511733079

Mercer, S., & Dornyei, Z. (2020). *Engaging language learners in contemporary classrooms.* Cambridge University Press.

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

Nicholson, H. (2009). *Theatre and education*. Palgrave Macmillan.

Noroozi, O., Banihashem, S. K., Biemans, H. J. A., Smits, M., Vervoort, M. T. W., & Verbaan, C.-L. (2023). Design, implementation, and evaluation of an online supported peer feedback module to enhance students' argumentative essay quality. *Education and Information Technologies, 28*(10), 12757-12784. https://doi.org/10.1007/s10639-023-11683-y

Panadero, E., & Alqassab, M. (2019). An empirical review of anonymity effects in peer assessment, peer feedback, peer review, peer evaluation and peer grading. *Assessment & Evaluation in Higher Education, 44*(8), 1253-1278. https://doi.org/10.1080/02602938.2019.1600186

Panadero, E., & Broadbent, J. (2018). Developing evaluative judgement: A self-regulated learning perspective. In D. Boud, R. Ajjawi, P. Dawson, & J. Tai (Eds.), *Developing Evaluative Judgement in Higher Education: Assessment for Knowing and Producing Quality Work* (pp. 81-89). Routledge.

Patchan, M. M., Schunn, C. D., & Correnti, R. J. (2016). The nature of feedback: How peer feedback features affect students' implementation rate and quality of revisions. *Journal of Educational Psychology, 108*(8), 1098-1120. https://doi.org/10.1037/edu0000103

Pinciotti, P. (1993). Creative drama and young children: The dramatic learning connection. *Arts Education Policy Review*, *94*(6), 24-28. https://doi.org/10.1080/10632913.1993.9936938

Poitras, J., Stimec, A., & Hill, K. (2013). On teaching: Fostering student engagement in negotiation role plays. *Negotiation Journal*, *29*(4), 439-462. https://doi.org/10.1111/nejo.12036

Prins, F. J., Sluijsmans, D. M. A., Kirschner, P. A., & Strijbos, J. W. (2005). Formative peer assessment in a CSCL environment: A case study. *Assessment and Evaluation in Higher Education, 30*(4), 417-444. https://doi.org/10.1080/02602930500099219

Reddy, K., Harland, T., Wass, R., & Wald, N. (2021). Student peer review as a process of knowledge creation through dialogue. *Higher Education Research & Development*, *40*(4), 825-837. https://doi.org/10.1080/07294360.2020.1781797

Reinholz, D. (2016). The assessment cycle: A model for learning through peer assessment. *Assessment & Evaluation in Higher Education, 41*(2), 301-315. https://doi.org/10.1080/02602938.2015.1008982

Sengül, K., & Sünbül, A. M. (2015). An investigation into comprehension competence and grammar achievements of learners of Turkish as a foreign language in Turkey. *Educational Alternatives*, *13*, 573-580. https://www.scientific-publications.net/get/1000013/1440677452648069.pdf

Seppanen, S. (2022). *What would happen if I said yes? Measuring the immediate and long-term impact of improvisation training on student teachers' subjective, neuroendocrine and psychophysiological responses* [Unpublished doctoral dissertation]. University of Helsinki.

Smith, H., Cooper, A., & Lancaster, L. (2002). Improving the quality of undergraduate peer assessment: A case for student and staff development. *Innovations in Education and Teaching International, 39*(1), 71-81. https://doi.org/10.1080/13558 000110102904

Stevens, R. (2015). Role-play and student engagement: Reflections from the classroom. *Teaching in Higher Education*, *20*(5), 481-492. https://doi.org/10.1080/13562517.2015.1020778

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

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

Weng, F., Ye, S. X., & Xue, W. (2023). The effects of peer feedback on L2 students' writing motivation: An experimental study in China. *The Asia-Pacific Education Researcher, 32*, 473-483. https://doi.org/10.1007/s40299-022-00669-y

Wilson, M. J., Diao, M. M., & Huang, L. (2015). 'I'm not here to learn how to mark someone else's stuff': An investigation of an online peer-to-peer review workshop tool. *Assessment & Evaluation in Higher Education*, *40*(1), 15-32. https://doi.org/10.1080/02602938.2014.881980

Winstone, N., & Carless, D. (2020). *Designing effective feedback processes in higher education: A learning-focused approach*. Routledge.

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

Yao, C. (2017). A case study on the factors affecting Chinese adult students' English acquisition in a blended learning environment. *International Journal of Continuing Engineering Education and Life Long Learning*, *27*(1/2), 22-44. https://doi.org/10.1504/IJCEELL.2017.080993

Yao, C., & Shao, S. (2024). Factors affecting teacher-student interactions in a foreign language teaching class: A language socialisation theory perspective. *Humanities and Social Sciences Communications, 11*, 1358. https://doi.org/10.1057/s41599-024-03900-x

Zhan, Y. (2021). Are they ready? An investigation of university students' difficulties in peer assessment from dual perspectives. *Teaching in Higher Education*, *29*(4), 823-840. https://doi.org/10. 1080/13562517.2021.2021393

Zhan, Y., Yan, Z., Wan, H. Z., 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*(4), 817-835. https://doi.org/10.1111/bjet.13310

Zheng, L., Chen, N. S., Cui, P., & Zhang, X. (2019). A systematic review of technology-supported peer assessment research. *The International Review of Research in Open and Distributed Learning*, *20*(5), 168-191. https://doi.org/10.19173/irrodl.v20i5.4333

Extended Abstract / Genişletilmiş Özet

Giriş

Drama, öğrencilerin öğrenmesini desteklemek için dinamik ve etkileşimli bir yaklaşım sağlar. Pinciotti (1993) dramayı, katılımcılara insan deneyimlerini hayal etme, canlandırma ve yansıtma fırsatı sunan, bir lider tarafından kolaylaştırılan yapılandırılmış bir öğrenme etkinliği olarak tanımlamıştır. Fleming (2006), drama metinlerinin sınıf uygulamalarında öğrenen merkezli bir yaklaşımı desteklediğini, öğrenenlerin pasif bilgi alıcıları olarak kalmak yerine öğrenme sürecinde aktif olarak algılamalarına, düşünmelerine, hareket etmelerine ve etkileşimde bulunmalarına olanak tanıdığını savunmaktadır. Drama, katılımı teşvik etmenin ötesinde, bilişsel ve duyuşsal boyutları bütünleştiren, oyunculuğu, olumlu duyguları, uyum yeteneğini, akıcılığı, iletişimsel yeterliliği ve öğrenci-öğretmen ilişkisini güçlendiren destekleyici bir ortamı geliştiren aktif bir yöntem olarak işlev görür (Belliveau ve Kim, 2013; Garaigordobil ve Berrueco, 2011; Hammond, 2015; Maley ve Duff, 2005).

Drama, öğrenciler arasında karşılıklı anlayışı, empatiyi ve yapıcı geri bildirimi teşvik ederek akran değerlendirme sürecinde önemli bir rol oynar. Akran değerlendirme öğrencilerin birbirlerinin öğrenme görevlerini değerlendirmesini ve geri bildirim vermesini içerir (Topping, 1998) ve değerlendirme, gözden geçirme ve not verme gibi bir dizi faaliyeti kapsar (Adachi vd., 2018). Pedagojik bir strateji olarak, öğrencileri önceden belirlenmiş kriterlere uygun geri bildirim sunmaya teşvik eder, katılımı ve yansıtıcı öğrenmeyi teşvik eder (Topping, 1998). Yükseköğretimde akran değerlendirmesi, öğretmen değerlendirmesine benzer öğrenme çıktıları sağlayabilir. Ayrıca öğrencilerin daha etkili öğrenmelerine ve genel akademik performanslarını geliştirmelerine yardımcı olur (Jongsma vd., 2023; Noroozi vd., 2023; Zheng vd., 2019).

Akran değerlendirme sınıf içinde, aidiyet duygusunu teşvik eder (Brown vd., 2021; Devisakti ve Ramayah, 2023) ve öğrenci motivasyonunu artırır (Deeley ve Bovill, 2015). Ayrıca, etkileşimi teşvik ederek değerlendiren ve değerlendirilen arasında bir ortaklık kurulmasını destekler (Winstone ve Carless, 2020) ve böylece öğrencilerin derse katılımını ve aktif katılımını artırır (Conde vd., 2017; Deng vd., 2023). Akran değerlendirme aynı zamanda işbirlikçi öğrenme süreçlerini destekler (Prins vd., 2005) ve çeşitli bağlamlarda öz-düzenlemeli öğrenmeyi kolaylaştırır (Topping, 2009; Bijami vd., 2013). Öğrencilerin problem tanımlama gibi üstbilişsel becerilerini (Topping, 1998) ve değerlendirici yargılarda bulunma yeteneklerini olumlu yönde etkiler (Panadero ve Broadbent, 2018).

Akran değerlendirmenin, farklı eğitim seviyelerinde ve konu alanlarında akademik performansı artırdığı (Double vd., 2020), eleştirel düşünme yeteneklerini ve üst düzey düşünme becerilerini geliştirmede etkili olduğu belirlenmiştir (Gaynor, 2019; Guelfi vd., 2021; Humberstone vd., 2024; Jiang vd., 2022; Zhan vd., 2023). Bilişsel sonuçların ötesinde, akran değerlendirmenin bağımsız öğrenme becerilerini, mesleki beceri gelişimini teşvik ederek (Smith vd., 2002), yansıtıcı düşünmeyi geliştirerek (Gaynor, 2019) ve yazma motivasyonunu artırarak (Weng vd., 2023) öğrencilerin öğrenme becerilerine fayda sağladığı belirlenmiştir. Bu çok yönlü faydalar, akran değerlendirmeyi yükseköğretimde hem bireysel hem de iş birliğine dayalı öğrenme çıktılarını teşvik etmek için değerli bir araç haline getirmektedir.

Bu çalışmanın amacı, öğrencilerin drama becerilerine ilişkin akran değerlendirmelerini çok yüzeyli Rasch modeli aracılığıyla incelemektir. Bu amaç doğrultusunda, drama becerilerine ilişkin yetenek analizi, değerlendiricilerin puanlama eğilimlerinin analizi, drama becerilerine ilişkin madde güçlük analizi ve değerlendiricilerin yanlılıklarının analizi yapılmıştır.

Yöntem

Çalışmada betimsel bir desene sahip nicel bir araştırma yöntemi kullanılmıştır. Creswell'e (2014) göre betimsel araştırma, araştırmacının mevcut durumu olduğu gibi sunduğu ve herhangi bir manipülasyon yapmadan durumları analiz ettiği bir yaklaşımdır. Çalışma grubunu İngilizce Öğretmenliği Bölümü'nde öğrenim gören 10 üniversite öğrencisi oluşturmaktadır. Öğrencilerin drama becerilerini puanlamak için araştırmacı tarafından oluşturulan "Drama Becerileri Akran Değerlendirme Formu" kullanılmıştır. Akran değerlendirme formu, her biri "çok yetersiz" (1) ile "çok yeterli" (5) arasında beşli Likert ölçeğine göre derecelendirilmiş 32 maddeden oluşmaktadır. Bu çalışmada, öğrenciler tarafından yapılan akran değerlendirmeleri çok yönlü Rasch modeli ile değerlendirilmiştir. Veri analizi için Linacre (1993) tarafından geliştirilen FACETS programı kullanılmıştır. Bu çalışmada öğrenci, değerlendirici ve madde olmak üzere toplam üç boyut belirlenmiştir. Bu nedenle, 10 katılımcı bir drama etkinliği gerçekleştirmiş ve tüm katılımcılar değerlendirme formu üzerinde birbirleri için akran değerlendirmeleri yapmıştır. Toplamda 2880 (10x9x32) veri elde edilmiştir. 10 katılımcı hem öğrenci olarak drama etkinliğini gerçekleştirmiş hem de değerlendirici olarak akranlarını puanlamıştır. Bu bağlamda öğrenci ve değerlendirici için yapılan kodlamalarda aynı kod numarası aynı kişiyi temsil etmektedir (Örneğin, Öğrenci 1 ve

Değerlendirici 1 aynı kişidir). Değerlendiricilerin puanlama eğilimleri (katılık/cömertlik), yanlılıkları, en kolay/en zor ölçüt ve belirlenen maddelere göre hangi öğrencinin drama becerisinin daha yüksek olduğu araştırılmıştır.

Bulgular

Öğrencilerin drama becerilerine ilişkin ölçüm raporunda, ki-kare testi sonuçlarına göre, öğrenciler arasında drama becerileri açısından anlamlı bir istatistiksel farklılık gözlenmektedir (χ^2=172,9; sd=9; p=0,00). Böylece sıfır hipotezi reddedilmiştir. Öğrencilerin drama becerilerindeki en yüksek performanstan en düşük performansa doğru genel sıralaması Öğrenci 4, 10, 5, 2, 9, 6, 8, 7, 3, 1 şeklinde belirlenmiştir.

Değerlendiricilerin katılık/cömertliklerine ilişkin ölçüm raporunda, ki-kare testi sonuçları sıfır hipotezinin reddedilmesine yol açarak değerlendiriciler arasında puanlama katılıkları veya cömertlikleriyle ilgili önemli bir istatistiksel fark olduğunu göstermiştir. (χ^2=148,1; sd=9; p=0,00). Değerlendiricilerin en katıdan en cömert olana doğru genel sıralaması Değerlendirici 4, 5, 8, 2, 6, 7, 9, 10, 3, 1 şeklinde belirlenmiştir.

Maddelere ilişkin ölçüm raporunda, ki-kare testinin sonuçlarına göre kriterlerin zorluk seviyeleri arasında anlamlı bir istatistiksel fark tespit edilmiştir (χ^2=1005,6; sd=31; p=0,00). En zor maddeler 32, 20 ve 12; en kolay maddeler ise 30, 1 ve 2 olarak belirlenmiştir.

Beklenmedik yanıtlar verilerine dayanarak, en beklenmedik verinin Değerlendirici 2 tarafından Öğrenci 3 için Madde 2'ye verilen puan olduğu belirlenmiştir. Elde edilen sonuçlar, Değerlendirici 2'nin beklenmedik puanlar vermede en çok tekrar eden değerlendirici olduğunu, Öğrenci 2 ve Öğrenci 3'ün beklenmedik puanlar almada en çok tekrar eden değerlendiriciler olduğunu ve Madde 30'un en çok tekrar eden madde olduğunu göstermektedir.

Öğrenci ve değerlendiriciler için etkileşim yanlılığına ilişkin ölçüm raporuna göre, öğrenci-değerlendirici etkileşim yanlılığında en çok tekrar edenler; Değerlendirici 2, Değerlendirici 1, Değerlendirici 10, Öğrenci 7, Öğrenci 4 ve Öğrenci 5 olarak belirlenmiştir.

Tartışma ve Sonuç

Araştırma sonucunda, öğrencilerin drama becerileri performanslarında görülen değişkenlik, öğrencilerin akran değerlendirme algılarından etkilenmiş olabilir (Patchan vd., 2016). Öğrencilerin drama becerilerinin en yüksek performanstan en düşük performansa doğru genel sıralamasına göre, Öğrenci 4 ve Öğrenci 10 gibi yüksek performans gösteren öğrenciler muhtemelen drama etkinliği sırasında yaratıcılık ve katılım gibi alanlarda güçlü beceriler sergilemişlerdir. Öğrenci 1 ve Öğrenci 3 gibi daha düşük performans gösteren öğrenciler drama etkinliği kriterlerini karşılamada zorluklarla karşılaşmış olabilir. Bu zorluklar yaratıcılık, duygusal ifade ya da özgüvenle ilgili güçlükleri içerebilir. Bu bağlamda, Zhan (2021) öğrencilerin akran değerlendirme sürecinde olumsuz deneyimler yaşayabileceğini belirtmiştir.

Değerlendiricilerin en katıdan en cömert olana doğru genel sıralamasına göre, en katı değerlendiricilerden biri olan Değerlendirici 4, muhtemelen kriterleri yorumlamaları veya mükemmellik tercihleri nedeniyle daha yüksek kişisel standartları yansıtıyor olabilir. Buna karşılık, en cömert değerlendirici olan Değerlendirici 1, muhtemelen empati, akranlarını teşvik etme arzusu veya performans beklentilerine ilişkin farklı algılardan etkilenerek esneklik göstermiş olabilir. Bu zıt yaklaşımlar, akran değerlendirme sürecinde tarafsızlığa ilişkin kuşkular doğurmaktır (Reddy vd., 2021; Wilson vd., 2015).

Madde ölçüm değerlerine göre, en zor ölçütün Madde 32 (Özgün doğaçlama becerileri gösterir) olduğu belirlenmiştir. Madde 32'nin drama etkinliğinin ileri düzey beceri, yaratıcılık ve üst düzey düşünme gerektiren bir yönünü temsil ettiği açıktır. Bu noktada, Seppanen (2022) doğaçlama eğitiminin öğretmen adaylarının kişilerarası iletişim becerilerini ve özgüvenlerini önemli ölçüde geliştirdiği sonucuna vararak, doğaçlama becerilerini geliştirmenin hem zorlu hem de oldukça faydalı olduğunu öne sürmüştür. Akran değerlendirme, üstbilişsel düşünme becerilerini geliştirerek öğrencilerin Madde 32 gibi ileri düzey maddelerin taleplerini daha iyi anlamalarına ve karşılamalarına yardımcı olabilir (Armengol-Asparó vd., 2022). Bunu Madde 20 (Beklenmedik durumlara sorunsuzca uyum sağlar) ve Madde 12 (Hareketler arasında yumuşak geçişler sergiler) takip etmektedir. Madde 20, muhtemelen bu tür senaryoların öngörülemezliği nedeniyle zorlayıcı olmuştur. Öğrenciler yapılandırılmamış durumlarla başa çıkma konusunda deneyim eksikliği yaşayabilir. Madde 12, performansta kesintisiz bir akış sağlamanın fiziksel ve teknik gereklilikleri nedeniyle zor olabilir. Yumuşak geçişlerin sağlanması, hareket dinamiklerinin derinlemesine anlaşılmasını, koordinasyonu ve pratik yapmayı gerektirir. Bu anlamda, öğrenci merkezli öğrenme yaklaşımlarının uygulanması, öğrencilerin kendi beceri gelişimlerinde aktif rol almalarını sağlayarak bu zorlukların üstesinden gelinebilir (Nicholson, 2009). Buna karşılık, en kolay kriter Madde 30 (Yaratıcılık içermeyen temel becerileri sergiler) olarak belirlenmiştir. Bunu Madde 1 (Sesini sürekli olarak yüksek tutar) ve Madde 2

(Anlaşılır bir ses tonu kullanır) takip etmektedir. Bu sonuçlar, daha az soyut düşünme ve yaratıcılık gerektirdiği için öğrenciler için daha erişilebilir olmasından kaynaklanıyor olabilir.

En beklenmedik verinin Değerlendirici 2 tarafından Öğrenci 3 için Madde 2'ye verilen puan olduğu belirlenmiştir. Ayrıca Değerlendirici 2, Madde 32 için Öğrenci 3'e beklenen değerin üzerinde puan vermiştir. Bu durum, beklenmedik puanlar verme konusunda en fazla tekrara düşen Değerlendirici 2'nin akran değerlendirme konusunda düşük performans gösterdiği şeklinde yorumlanabilir. Araştırmanın dikkat çekici bir bulgusu olarak, Değerlendirici 10 Öğrenci 3'ü Madde 26 için beklenen değerin üzerinde puanlamıştır. Dolayısıyla sadece Öğrenci 3'ün beklenen değerin üzerinde puan aldığı tespit edilmiştir. Ancak Öğrenci 3'ün drama becerileri açısından düşük bir performansa sahip olduğu görülmektedir. Buna göre Öğrenci 3 için yapılan değerlendirmelerde bazı maddelerin yanlı puanlandığı söylenebilir. Bununla birlikte, Öğrenci 2 için üç farklı değerlendiricinin beklenen değerin altında puanlama yaptığı ve bunlardan ikisinin Madde 30'a ait puanlar olduğu görülmektedir. Sonuç olarak, Değerlendirici 2'nin beklenmedik puanlar vermede en çok tekrar eden değerlendirici olduğu, Öğrenci 2 ve Öğrenci 3'ün beklenmedik puanlar almada en çok tekrar eden değerlendiriciler olduğu ve Madde 30'un en çok tekrar eden madde olduğu belirlenmiştir.

Araştırmanın dikkat çekici bulgularından biri, Değerlendirici 10'un Öğrenci 7 ve Öğrenci 8 için akran değerlendirmelerinde oldukça cömert bir yaklaşım sergilerken, Öğrenci 4, Öğrenci 5 ve Öğrenci 6 için akran değerlendirmelerinde oldukça katı bir yaklaşım sergilemesidir. Buna göre, Değerlendirici 10 kendi grup arkadaşları (Öğrenci 7 ve Öğrenci 8) için cömert puanlar verirken, başka bir grubun tüm üyeleri (Öğrenci 4, Öğrenci 5 ve Öğrenci 6) için katı puanlar vermiştir. Bu sonuçlar, grup temelli yanlılığa işaret etmektedir. Buna karşılık, diğer grupların üyelerine yönelik daha katı değerlendirmeler, iş birliği eksikliğinden veya algılanan rekabetten kaynaklanıyor olabilir. Bu anlamda, akran değerlendirmenin etkili bir biçimde uygulanması durumunda, ekip çalışması ve iş birliğini destekleyerek karşılaşılan sorunların azaltılmasına katkı sağlayabileceği ifade edilmektedir (Reinholz, 2016; Zheng vd., 2019).