



Exploring the Relations between Peer Assessment, Self-Assessment, and Academic Achievement in School Science Experiment Course

Riana Nurismawati^{1*}, Lilit Rusyati¹, Hanna Nurul Husna¹, Winda Hidayanti²

¹Science Education, Universitas Pendidikan Indonesia, Bandung, Indonesia

²STEM Education, King's College London, London, United Kingdom

*Correspondence to: nurismawatiriana@upi.edu

Abstract: Assessment has a pivotal role in education to evaluate students' performance, with peer- and self-assessment as alternative tests to enhance the learning process. However, the links between self and peer assessment remain lacking in studies, in addition to academic achievement, particularly in science education. This study explores the correlations among peer assessment, self-assessment, and academic achievement (GPA) for higher education in School Science Experiment Courses. A correlational and comparative research design was conducted involving 96 preservice teachers. Data were analyzed using SPSS, with Spearman's rank correlation for link analyses and the Kruskal-Wallis test with post-hoc comparisons for group differences. For overall scores, results described a strong relationship between peer and academic achievement and a weak correlation between peer- and self-assessment. No significant correlation was shown between self-assessment and GPA. For specific indicators, peer assessment is linked strongly with GPA all across indicators, while peer- and self-assessment correlated in six out of ten indicators, and the correlation between self-assessment and GPA lies in one indicator. Group comparisons revealed significant differences in peer-scores across High, average-, and low-achiever groups but no significant differences in self-scores. This relationship shows other alternative and innovative assessments that improve reflective skills and active learning.

Keywords: academic achievement; correlation; peer-assessment; self-assessment

Recommended citation: Nurismawati, R., Rusyati, L., Husna, H. N., & Hidayanti, W. (2025). Exploring the Relations between Peer Assessment, Self-Assessment, and Academic Achievement in School Science Experiment Course. *Journal of Innovation in Educational and Cultural Research*, 6(3), 657-665.

INTRODUCTION

Assessment is one of the fundamental aspects of education (Heritage & Wylie, 2018; Morris et al., 2021; Ghorbanpour et al., 2021). As Wiliam (2011) stated, assessment is a central process in effective instruction because only through assessment can teachers know whether the learning objective has been fulfilled or not in the learning outcomes. All the learning processes cannot be mentioned as impactful or not if there was no assessment and evaluation after the learning process. Nonetheless, an assessment aims to measure students' abilities and improve the standard (Weeden et al., 2002; Salas-Bustos et al., 2025). Assessment is not only the process of evaluation to explore and reflect on what has been done in the past, but this activity can lead to a continuous improvement of the learning process, either in the preparations or during class, to reach educational long-term and short-term goals. Similar to the definition of summative and formative assessment. Summative assessments are used to evaluate students' outcome after the implementation of the learning process, whereas formative assessments are used to monitor and guide learning and teaching. Summative learning is an evaluation of learning that indirectly affects students' perceptions of success and failure in their learning process (Weeden et al., 2002). Assessment could be utilized to assess students' performance and evaluate the lesson plan's quality and for the educators themselves.

There is a demand for altering the assessment of learning to be focused on assessment for learning (Dythe, 2007) to emphasize the strengths and weaknesses of students' performances (Karami & Rezaei, 2015). Assessment for learning requires the participation of teachers, students' peers, and students themselves (Leenknecht et al., 2021). Evaluation does not focus on the ability of cognitive skills but needs to examine the other abilities, such as skills and performances, and even explore deeper into human behavior attitudes, interpersonal, intrapersonal, emotions, and so on (Brown, 2022). One of the practical applications of Assessment for learning is the activity of student self-assessment (Heritage & Wylie, 2018) and feedback from their peers, such as in Absolum (2011). Shepard (2000) stated that the most effective way is alternative assessment, which the sample is peer-assessment. Peer- and self-assessment in this era pay more attention since it encourages students to learn to analyze and judge their peers and themselves wisely. Both assessments needed to be conducted together since they will learn to compare how they assess themselves and their peers. Moreover, both assessments contribute some merits for students, such as responsibility, critical thinking, and reflective professionalism (Iglesias et al., 2022). Peer assessment implies that students judge their friends' performances, which they have as equal-status learners. This type of assessment provides strength and weakness

identification, metacognitive development, and other professional skills (Topping, 2009). Likewise, regarding self-assessment, Brown and Harris (2013) argue that this typical test supports students in evaluating their work to enhance their performance or products based on the set standards. Peer and self-assessment are valuable tools for providing students feedback on their performance and its improvement (Morris et al., 2021). Both assessments could be summative and formative, referring to a judgment on the outcomes for the former and in-depth feedback for the latter (Stančić, 2021). However, summative and formative assessments have the same goals beyond the grading session: to develop reflective abilities, promote self-regulation in learning, and comprehend learning outcomes and criteria (Boud, 2013; Panadero et al., 2017).

Studies found that peer evaluation completes self-assessment as it boosts students' comprehension of good performance and awareness to reflect on themselves (Reinholz, 2016), especially regarding performance tests. As Topping (2009) mentioned, the products to be evaluated by peers and self-assessment consist of writing, presentations, test performance, or other skills and behaviors. Nonetheless, the link between peer assessment and self-assessment remains questionable and, at the same time, an interesting topic, especially in terms of how both assessments influence one another. Iglesias et al. (2022) depicted a moderate concordance between self-assessment and the lecturer, along with a strong link between judgments from their fellows and the lecturer. Therefore, both assessments are pivotal to Higher Education's student-centered evaluation. Many studies highlighted the connection between evaluation from the teacher and peer assessment or teacher evaluation and self-assessment, but there are currently insufficient studies exploring the relationship between peer- and self-assessments. Moreover, Cirit (2021) found some correlations between self and peer, but they are still inconsistent. This performance test in the course involves a lot of fellows' participation. For instance, during a student's performance, the other peers act like junior high school students. Therefore, exploring the correlation between both assessments, besides examining peer evaluation and self-assessment, is important.

Along with both assessments, academic achievement or GPA was added as another score with more absolute data in the beginning yet rarely explored. While the others compared to teacher assessments such as in Iglesias et al. (2022) and Salehi and Daryabar (2014) on that subject, which still contribute to the bias factors, this GPA gives an unpredictable result that might related or not be related to peer and self-assessment. Academic achievement score is a score that is gained with a long process of learning, with various subjects as well as its evaluation. Those results might relate to the students' quality cognitive skills and performance as university students. Additionally, in teaching activities, due to the lack of knowledge about the concept of formative assessment and insufficient guidelines, the implementation of formative assessment is still low (Rahman et al., 2021). Thus, more research data about formative assessment is needed in instructional settings (Dayal, 2021).

As the importance of formative assessment has been explained, Medland (2016) and Deneen and Boud (2014) described the low alignment between assessment as literature and assessment practices. The reason is not because of the lack of assessment developments since numerous studies are available everywhere (e.g., Akbar et al., (2017); Rusyati et al., (2021)) but because the low assessment literacy while conducting the process of assessment (Medland, 2016). Assessment literacy could be included as the knowledge of students' assessment, the interpretation of learning results, and the important one, which is the implementation of the assessment procedures with continuous improvement. Therefore, the results of correlational and comparison studies of both peer- and self-assessment showed contributions to increase the literacy of the assessment, especially in implementing those alternative assessments and the further analysis post-class.

In this paper, this research would like to explore the relationship between peer assessment, self-assessment, and academic achievements (GPA) of students. This includes not only the overall score of the test but also the deeper analysis for each indicator of peer and self-assessment. Peer and self-assessments are alternative assessments. There is still a lack of studies focusing on formative assessment. Meanwhile, students' GPA data was added for various analyses to understand the findings and more varied. The comparison groups were made based on their academic achievements to investigate further the distribution of peer-assessment and self-assessment scores in three categorical groups. Besides that, this research contributes to peer assessment and self-assessment in higher learning institutions. Both assessments have advantages for the learning and teaching process. Formative tests promote learning (Murphy et al., 2023) and indirectly make learners learn metacognitively (Rhodes, 2016).

METHODS

This quantitative research is a correlational study research design, with an addition of a comparative study for deeper analysis. According to Creswell (2015), correlational research design helps researchers to describe and measure the degree of association between two or more variables. Data were collected from students of Science Education majors conducting School Science Experiment courses in their second year (sophomore year). The sample size is 96 preservice teachers within four different cohorts. The final assessment of this course is that each student should perform as a science teacher who conducts science experiments, and the other students act as middle school students. Thus, the aspects are mostly related to presentation or

teaching skills. After each student performed in front of the class, the other fellows were given certain minutes to fill out peer assessments and self-assessments for the performer. The scores would be utilized to evaluate students and the formal grading process.

The gathered data comprises self-assessment, peer assessment, and students' academic achievement scores (GPA). The evaluation for peers and self-assessment consists of ten aspects, which consist of legibility of the experiment worksheet, the attractiveness of the science experiment, suitability of the science experiment in junior high school, delivery of experiment instructions, mastery of science experiment, triggering student's participation, guiding student's activity, the level of enthusiasm, clarity of voice, attitude or ethics. Peer- and Self-assessment instruments use a Skala-Likert Scale rating of 1-10 points for each indicator. The highest point, 10, means the student gives a perfect score to their peers or themselves. Students' achievement scores were collected from cumulative grades during 4 semesters.

Furthermore, the data were analyzed using SPSS to find the correlation among the three data collections. A normality test was conducted to choose the type of correlation to use. The Shapiro-Wilk test would be chosen since the data samples are small, and this test is the best method for the normality test (Hernandez, 2021). If the data is not normally distributed, non-parametric tests will be used. A correlational test was conducted with two different findings. Firstly, overall scores for each peer assessment, self-assessment, and academic achievements (GPA). Secondly, both peer assessment and self-assessment have 10 different aspects. The author would like to analyze this section's aspects comprehensively.

Group comparison analyses were also conducted to explore peer assessment and self-assessment towards their academic achievement score. Based on the mean and standard deviation calculation, academic achievement scores were divided into three groups: low-achiever group, average-achiever group, and high-achiever group. Then, the Kruskal-Wallis test was used to depict the significant differences across GPA categories on peer assessment and self-assessment. This result can support and complement the related findings.

RESULT AND DISCUSSION

In the last two decades (2001-2022), publications on children's education and artisanal and small-scale mining Total data are 96 samples with three scores for each sample: peer assessment, self-assessment, and academic achievement scores. In the beginning, the result of Normality is presented in Table 1.

Table 1. Normality test for overall score in peer assessment, self-assessment, and academic achievement score.

Shapiro-Wilk			
	Statistic	df	Sig.
Total peer	.947	96	.001
Total self	.945	96	.001
GPA score	.905	96	.000

Based on the result, the data is not normally distributed since the significance level for total peer assessment, total self-assessment, and academic performance score are less than .05. Thus, a non-parametric test, which is Spearman's rank correlation would be used for the correlational test (Table 2).

Table 2. Spearman-Rho test for the correlation of total score peer assessment, self-assessment, and academic achievement

		Total peer	Total self	GPA
Total peer	Correlation Coef.	1.000	.231*	.656**
	Sig. (2-tailed)	.	.024	.000
	N	96	96	96
Total self	Correlation Coef.	.231*	1.000	.152
	Sig. (2-tailed)	.024	.	.138
	N	96	96	96
GPA	Correlation Coef.	.656**	.152	1.000
	Sig. (2-tailed)	.000	.138	.
	N	96	96	96

The overall score shows a significant but weakly positive correlation between peer assessment and self-assessment (correlation coefficient = 0.231). It means that the higher the results from their peers, the higher

the possibility that their self-assessment is, too. The other results find a strong correlation between peer assessment and academic achievement scores, with a correlation coefficient of 0.656 and sig. < 0.05. Meanwhile, there is no relation between self-assessment and academic achievement, with only a correlation coefficient of 0.152 (Sig. > 0.05). Another deeper analysis was gained from each aspect of peer assessment and self-assessment to seek any highly correlated indicators. The details are displayed in Table 3.

Table 3. Spearman-Rho test for the correlation of peer assessment and academic achievement in each indicator

Peer Assessment Indicators	Correlation Coefficient	Significance (p-value)
Legibility of experiment worksheet	.628**	.000
The attractiveness of science experiments	.534**	.000
Suitability of science experiment in Junior High School (JHS)	.380**	.000
Delivery of experiment instructions	.695**	.000
Mastery of science experiment	.728**	.000
Triggering student's participation	.515**	.000
Guiding student's activity	.669**	.000
The level of enthusiasm	.379**	.000
Clarity of voice	.638**	.000
Attitude or ethics	.504**	.000

Table 3 describes the result of the Spearman-Rho correlation test for each indicator of peer assessment in line with their academic achievements. All indicators have a significance level (p-value) < .05, indicating that all indicators strongly correlate with GPA scores.

Table 4. Spearman-Rho test for correlations of peer assessment and self-assessment (only show indicators based on the presence of significance value)

Peer & Self-Assessment Indicators	Correlation Coefficient	Significance (p-value)
The attractiveness of science experiments (2)	.230*	.024
Suitability of science experiment in Junior High School (JHS) (3)	.290**	.004
Delivery of experiment instructions (4)	.204*	.046
Triggering student's participation (6)	.263**	.010
The level of enthusiasm (8)	.375**	.000
Clarity of voice (9)	.265**	.009

Peer-assessment and Self-assessment have the same indicators. Based on the provided data, there are six indicators out of ten whose significance levels are below .05. The result indicated that six indicators of peer assessment have strong correlations with six self-assessment indicators.

Table 5. Spearman-Rho test for the correlation of self-assessment and academic achievement (GPA) (only shows indicators based on the presence of significance value)

Peer & Self-Assessment Indicators	Correlation Coefficient	Significance (p-value)
Mastery of science experiment (5)	.212*	.038

The correlation between self-assessment and academic achievement, which is less than .05, lies in only one indicator. The significance level is quite weak (.038 < .05). The remaining indicators were not strongly correlated with their academic achievement score. In the relationship result among the scores, it was found that peer assessment strongly correlates with their academic achievement while self-assessment does not even correlate with their academic achievement. Thus, students' academic achievement was divided into three groups based on the mean and standard deviation for further analysis. Group distribution is shown in Figure 1.

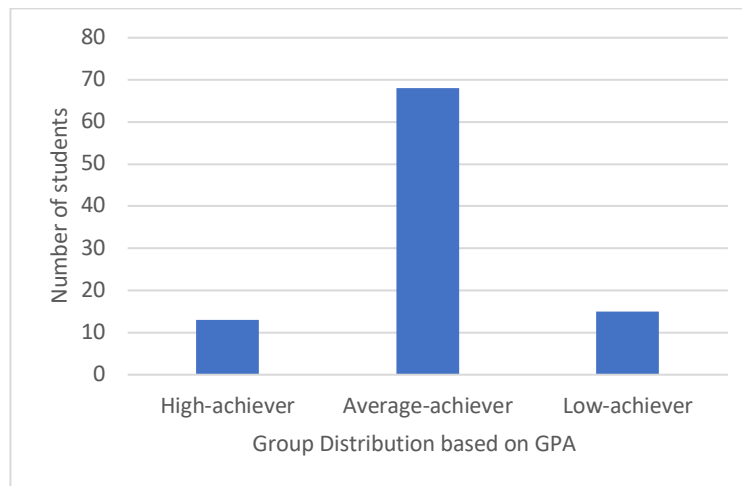


Figure 1. Group Distribution based on Students' Academic Achievement (GPA)

In the line graph, there are three groups: 13 students in the high-achiever group, 68 students in the average-achiever group, and 15 students in the low-achiever group. Based on the calculation of mean score (3.63) and standard deviation = 0.19, then samples in the high-achiever group should have a GPA score of more than 3.82, the low-achiever group should have less than 3.44, and the average-achiever group should have scores between 3.44 and 3.82. After the division of groups, the overall scores for peer assessment and self-assessment were input into a Kruskal-Wallis test with those three categorical data. The result revealed in Table 6

Table 6. Peer assessment and self-assessment among three categories based on independent samples Kruskal-Wallis Test

Null Hypothesis	Sig.	Decision
The distribution of peer assessment is the same across categories of GPA Score	.000	The null hypothesis is rejected (Sig. <.05)
The distribution of self-assessment is the same across categories of GPA Score	.713	The null hypothesis is retained (Sig. >.05)

According to Table 6, the null hypothesis was rejected in peer assessment since the sig value < .05, indicating a significant difference among the GPA categories. On the other hand, the finding shows no significant difference in self-assessment scores among those three groups. This result might indicate that high-achiever students receive higher scores from peers and other groups. However, these results need further analysis before jumping to a conclusion. A post-hoc test must be conducted to identify which specific groups differ significantly. The description of post-hoc analysis can be checked in Table 7.

Table 7. Pairwise Comparisons of Peer-Assessment Scores

Groups	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.
High-Average Groups	24.069	8.430	2.855	.004	.013
High-Low Groups	55.564	10.554	5.265	.000	.000
Average-Low Groups	31.495	7.945	3.964	.000	.000

Following Kruskal-Wallis' Test result, a Dunn's Test, which can use Pairwise comparisons in SPSS, must be conducted to determine how specific groups differ. Pairwise comparison is a statistical method for group comparisons to assess correlations between the pairs of means (Nordstokke & Stelnicki, 2024). The source data of pairwise comparisons in SPSS, which applied Bonferroni correction to avoid Type I Errors, lies on the row Adjusted p-value for multiple comparisons (Adj. Sig.). The table shows that all pairwise comparisons are statistically significant. Low-achiever is significantly different from both the Average-achievers and High-achiever groups. The average group significantly differs from the high-achiever group, although the significance level is slightly weaker than the other comparisons.

The results show some key findings related to peer assessment, self-assessment, and academic achievement. The normality test was conducted as an initial step to begin the analysis. Shapiro-Wilk was chosen compared to Kolmogorov since the sample size is small. The finding shows that the data is not normally distributed for peer assessment, self-assessment, and academic achievement scores (GPA). For the following

step, the Spearman-rho correlation test was used to explore the link among those three variables. [Essam et al. \(2022\)](#) found that Spearman-rho and Pearson correlation tests are good correlation estimators since both have the lowest Root Mean Square Error (RMSE) values compared to Kendall's Tau.

There are three data interpretations of the correlational study in Table 2. The overall score means the total score of 10 indicators in the peer assessment score. The Correlation coefficient for overall peer assessment and academic achievement score is more than .60, indicating a strong relationship between them ([Ebenezer & Atakpa, 2023](#)). Other results revealed a weak correlation between the overall score of peer- and self-assessment scores with the coefficient below .40. Meanwhile, the correlation coefficient between self-assessment and students' academic achievement is below .20, which means there is little or even no relationship between variables. These results are similar, though it is a different point of view from meta-analyses from [Double et al. \(2020\)](#). According to the study, students' performance in school improves more when peer evaluation activities are included than when no assessment is performed. However, the task of self-evaluation is not substantially distinct. Moreover, some other researches support this correlation result, such as from [Oren \(2018\)](#), who states the presence of a moderately high positive correlation between peer-, self-, and teacher assessment in the Science Education II course for preservice teachers. Similarly, [Nurov \(2000\)](#) revealed that in foreign language subjects, there is a weak relationship between students' self-assessment, achievement tests, and teacher assessments. In addition, [Cirit \(2021\)](#) supported the possible correlations among three assessments: peer-, self-, and teacher assessments to evaluate teaching activities on three science topics. The result found no correlations between self- and teacher assessment on those topics, but a correlation appears between peer and self-assessment in several topics. Those correlational researches indicate that peer and teacher assessment is more likely to correlate and, in some cases, has higher scores than the relation between self and teacher assessment. Thus, practicing self- and peer assessment of their performances is needed as a part of the learning process and metacognitive development ([Topping, 2009](#)).

There are many benefits of peer assessment and self-assessment in the learning process, especially for preservice teachers. [Ratminingsih et al. \(2018\)](#) stated that those practices are pivotal in encouraging students to reflect on their performances for their academic learning and future profession. This is similar to [Wanner and Palmer \(2018\)](#) regarding the concern of self- and peer-assessment practices in learning processes that can promote a formative vision of assessment and support a student-centered perspective. Apart from that, [Manisier et al. \(2024\)](#) stated that peer assessment could strengthen the self-belief and motivation of the student since students get confidence from their peers that they are capable of that specific ability. There might be some peer-bias factors in the assessment process, such as peer competition, misunderstandings ([To & Panadero, 2019](#)), and even the initial fellows' perception towards students. [Magin \(2001\)](#) argues that friendship among the other students contributes to the lack of fairness in the peer assessment.

Nevertheless, further research needs to be conducted to determine potential bias factors. The result of a comparison group using the Kruskal-Wallis test in Table 6 revealed that high-achiever groups, based on their GPA, tend to have higher scores than their peers—likewise, the average-achiever groups and low-achiever groups. In Table 7 from Dunn's Test, peer-assessment scores in each group differ significantly, especially between low- and high achievers and low- and average-achiever groups. However, the outcome of comparison groups in self-assessment scores shows the opposite. The distribution of self-assessment scores among high-, average-, and low-achiever groups are similar, which means there is no difference in how they assess themselves in the self-assessment process among all the groups. Some factors might be affecting the result of self-assessment scores, such as overconfidence or even being overly pessimistic about themselves, whereas the result comes from [Sitzmann and Johnson \(2012\)](#) and is similar to [Boud et al. \(2013\)](#). Nonetheless, their self-assessment is similar within those three groups. It indicates that students are not sure enough or perhaps do not know how to assess themselves objectively. [Panadero et al. \(2016\)](#) stated that a more accurate contributor to self-assessment is the students' competencies and experiences added by their skills or comprehensions being self-assessed. Individual characteristics and differences significantly impact the self-assessment ([Panadero et al., 2016](#)). Thus, the ability to assess themselves needs to be trained and supported by teachers to identify their strengths and weaknesses in students' work.

As peer assessment and self-assessment were formed from ten indicators of students' performance, another finding focuses on the correlations within each indicator. Each indicator in peer assessment has strong correlations with academic achievement scores, which implies the way they assess their peers is higher for students who obtain a high academic achievement score even in each indicator. Identification comes from [Langan et al. \(2008\)](#) about the possibilities of biases when assessing their peers, which could be from gender, institutional affinity, and peer-group status. [Matsuno \(2009\)](#) shares that the way students assess their peers is more lenient and for themselves severely which might affect the result of the scores. However, further research needs to be conducted to see the bias factors that might or might not be impactful. It is because students assess their peers based on the performance that they watch directly at that moment. The indicators on the assessment were made to assess their preparation before experimenting, such as legibility of the worksheet,

whether it is well-prepared and understandable or not, the choice of the experiment must be appropriate for Junior High School students, and their mastery of science concepts on that chosen experiment. Besides the preparation, the most important thing is their performance on D-day, which includes the attractiveness of the experiment, the teaching delivery of the instructions, the teacher's ability to trigger students' participation and guide the activity, their enthusiasm for teaching, the clarity of voice and certainly the attitude or ethics while performing in front of the class.

Another finding is the correlation between peer- and self-assessment indicators, appearing in six indicators: the attractiveness of the experiment, its suitability for middle school students, the way how the performer teaches in the class and triggers the students to be active, the performers' enthusiasm, and their voices. All those indicators are correlated since those aspects are feasible for students and the performer to experience directly. In addition, the students were instructed to assess their peers right after the performance; thus, their memories of the experience were still clear to recall and without any discussion with others. One interesting finding is that only one out of ten indicators link self-assessment and academic achievement, which is students' mastery of science experiments. In other words, their peers could not identify their concept mastery, while only those who can evaluate themselves and their judgment about the mastery correlated with their academic achievement. Learning outcomes and self-assessment are positively correlated. Panadero et al. (2017) reported that self-regulation can be enhanced through their self-assessment practice, and Banarjee and Kumar (2014) revealed that students' academic achievement is positively linked to their self-regulation attitudes.

Bland and Gareis (2018) said that an innovative assessment is required for performance tests. This research aligns with innovative and alternative assessment, whereas peer and self-assessment are tools to measure student performance scores. This research can be a sample of conducting self-assessment and peer-assessment as innovative assessments in performance-based tests. The study's findings indicate that the quality of the peer assessment correlates strongly with their GPA. This finding can be used as consideration on future research plans for researchers and as references in the classroom for teachers or educators. This correlational research in formative assessment reveals how self-assessment and peer evaluation were conducted and how they also impacted students' test scores. Butler (2018) stated that a precise self-assessment mechanism needs to be conducted and analyzed further. In addition, the results of correlational and comparison studies of both peer- and self-assessment showed contributions to improve assessment literacy, particularly in implementing those alternative assessments and the further analysis after the class. Besides that, the following research can focus more on only one assessment to make it more intense. The research about formative assessment and its relation to that is highly recommended to be more intensely analyzed to meet the best outcomes and to apply in learning evaluation, learning process, and even in upcoming research.

CONCLUSION

Based on the results for overall score, peer assessment and students' academic achievement are strongly correlated, peer evaluation and self-assessment are weakly correlated, and there is no significant relationship between self-assessment and academic achievement. To be more precise, the relationship between peer assessment and academic achievement appears in all indicators. In addition, self-assessment and students' academic achievement only appear in 1 of 10 indicators. Another finding is group comparisons. It revealed a significant distinction in peer-assessment scores across all GPA-based groups: high-achiever, average-achiever, and low-achiever groups. The self-assessment scores, however, showed no significant differences. Further research is recommended to investigate self-assessment concerning self-regulation, self-perception, reflective skills, and potential biases in peer assessment.

REFERENCES

- Absolum, M. (2011). *Clarity in the Classroom: Using Formative Assessment for Building Learning-Focused Relationships*. Portage & Main Press.
- Akbar, M. N., Firman, H., & Rusyati, L. (2017). Developing Science Virtual Test to Measure Students' Critical Thinking on Living Things and Environmental Sustainability Theme. *Journal of Physics: Conference Series*, 812(1), 12106.
- Banarjee, P., & Kumar, K. (2014). A Study on Self-Regulated Learning and Academic Achievement Among the Science Graduate Students. *International Journal of Multidisciplinary Approach and Studies*, 1(6), 329–342.
- Bland, L. M., & Gareis, C. R. (2018). Performance Assessments: A Review of Definitions, Quality Characteristics, and Outcomes Associated with Their Use in K-12 Schools. *Teacher Educators' Journal*, 11, 52–69.

- Boud, D. (2013). *Enhancing Learning Through Self-Assessment*. Routledge.
- Boud, D., Lawson, R., & Thompson, D. G. (2013). Does Student Engagement in Self-Assessment Calibrate Their Judgement Over Time? *Assessment & Evaluation in Higher Education*, 38(8), 941–956.
- Brown, G. T. L. (2022). The Past, Present and Future of Educational Assessment: A Transdisciplinary Perspective. *Frontiers in Education*, 7.
- Brown, G. T. L., & Harris, L. R. (2013). Student Self-Assessment. In *SAGE Handbook of Research on Classroom Assessment* (pp. 367–393). Sage.
- Butler, Y. G. (2018). The Role of Context in Young Learners' Processes for Responding to Self-Assessment Items. *The Modern Language Journal*, 102(1), 242–261.
- Cirit, D. K. (2021). An Analysis of Self-, Peer-, and Teacher- Assessment Within the Scope of Classroom Teaching Activities. *Shanlax International Journal of Education*, 9(S2-Sep), 150–163.
- Creswell, J. W. (2015). *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*. Pearson.
- Dayal, H. (2021). How Teachers Use Formative Assessment Strategies During Teaching: Evidence from the Classroom. *Australian Journal of Teacher Education*, 46(7), 1–21.
- Deneen, C., & Boud, D. (2014). Patterns of Resistance in Managing Assessment Change. *Assessment & Evaluation in Higher Education*, 39(5), 577–591.
- Double, K. S., McGrane, J. A., & Hopfenbeck, T. N. (2020). The Impact of Peer Assessment on Academic Performance: A Meta-Analysis of Control Group Studies. *Educational Psychology Review*, 32(2), 481–509.
- Dysthe, O. (2007). The Challenges of Assessment in a New Learning Culture. In *Balancing Dilemmas in Assessment and Learning in Contemporary Education* (pp. 15–28). Routledge.
- Ebenezer, E., & Atakpa, A. O. (2023). Assessing the Correlational Statistics: The Direction, Weight and the Relevance in Educational Research. *Intercontinental Journal of Education, Science and Technology*, 7(1).
- Essam, F., El, H., & Ali, S. R. H. (2022). A Comparison of the Pearson, Spearman Rank and Kendall Tau Correlation Coefficients Using Quantitative Variables. *Asian Journal of Probability and Statistics*, 36–48.
- Ghorbanpour, E., Abbasian, G. R., & Mohseni, A. (2021). Assessment Alternatives in Developing L2 Listening Ability: Assessment FOR, OF, AS Learning or Integration? *International Journal of Language Testing*, 11(1).
- Heritage, M., & Wylie, C. (2018). Reaping the Benefits of Assessment for Learning: Achievement, Identity, and Equity. *ZDM*, 50(4), 729–741.
- Hernandez, H. (2021). Testing for Normality: What is the Best Method. *ForsChem Research Reports*, 6(5), 1–38.
- Iglesias, M. C. P., Vidal-Puga, J., & Pino Juste, M. R. (2022). The Role of Self and Peer Assessment in Higher Education. *Studies in Higher Education*, 47(3), 683–692.
- Karami, A., & Rezaei, A. (2015). An Overview of Peer-Assessment: The Benefits and Importance. *Journal for the Study of English Linguistics*, 3(1), 93–100.
- Langan, A. M., Shuker, D. M., Cullen, W. R., Penney, D., Preziosi, R. F., & Wheeler, C. P. (2008). Relationships Between Student Characteristics and Self-, Peer and Tutor Evaluations of Oral Presentations. *Assessment and Evaluation in Higher Education*, 33(2), 179–190.
- Leenknecht, M., Wijnia, L., Köhlen, M., Fryer, L., Rikers, R., & Loyens, S. (2021). Formative Assessment as Practice: The Role of Students' Motivation. *Assessment & Evaluation in Higher Education*, 46(2), 236–255.
- Magin, D. (2001). Reciprocity as a Source of Bias in Multiple Peer Assessment of Group Work. *Studies in Higher Education*, 26(1), 53–63.
- Manisier, K., Glackin, M., & Harrison, C. (2024). Not Missing Their Targets: Exploring the Impact of Mock Examination Feedback Practices. *Impact: Journal of the Chartered College of Teaching*.
- Matsuno, S. (2009). Self-, Peer-, and Teacher-Assessments in Japanese University EFL Writing Classrooms. *Language Testing*, 26(1), 75–100.
- Medland, E. (2016). Assessment in Higher Education: Drivers, Barriers and Directions for Change in the UK. *Assessment & Evaluation in Higher Education*, 41(1), 81–96.
- Morris, R., Perry, T., & Wardle, L. (2021). Formative Assessment and Feedback for Learning in Higher Education: A Systematic Review. *Review of Education*, 9(3), e3292.

- Murphy, D. H., Little, J. L., & Bjork, E. L. (2023). The Value of Using Tests in Education as Tools for Learning—Not Just for Assessment. *Educational Psychology Review*, 35(3), 89.
- Nordstokke, D., & Stelnicki, A. M. (2024). Pairwise Comparisons. In *Encyclopedia of Quality of Life and Well-Being Research* (pp. 4935–4936). Springer.
- Nurov, A. (2000). *Self-Assessment of Foreign Language Achievement: The Relationship Between Students' Self-Assessment, Teachers' Estimates and Achievement Test* [Master's thesis, Bilkent University].
- Oren, F. S. (2018). Self, Peer and Teacher Assessments: What is the Level of Relationship Between Them? *European Journal of Education Studies*.
- Panadero, E., Brown, G. T. L., & Strijbos, J.-W. (2016). The Future of Student Self-Assessment: A Review of Known Unknowns and Potential Directions. *Educational Psychology Review*, 28, 803–830.
- Panadero, E., Jonsson, A., & Botella, J. (2017). Effects of Self-Assessment on Self-Regulated Learning and Self-Efficacy: Four Meta-Analyses. *Educational Research Review*, 22, 74–98.
- Rahman, K. A., Hasan, M. K., Namaziandost, E., & Ibna Seraj, P. M. (2021). Implementing a Formative Assessment Model at the Secondary Schools: Attitudes and Challenges. *Language Testing in Asia*, 11(1), 18.
- Ratminingsih, N. M., Marhaeni, A., & Vigayanti, L. P. D. (2018). Self-Assessment: The Effect on Students' Independence and Writing Competence. *International Journal of Instruction*, 11(3), 277–290.
- Reinholz, D. (2016). The Assessment Cycle: A Model for Learning Through Peer Assessment. *Assessment & Evaluation in Higher Education*, 41(2), 301–315.
- Rhodes, M. G. (2016). Judgments of Learning: Methods, Data, and Theory. In *The Oxford Handbook of Metamemory* (Vol. 1, pp. 65–80). Oxford University Press.
- Rusyati, L., Rustaman, N. Y., Widodo, A., & Ha, M. (2021). Development of Questionnaire Instrument to Assess Students' Transformative Competencies in Science Learning. *Journal of Physics: Conference Series*, 2098(1).
- Salas-Bustos, D. A., Coral-Padilla, S. J., Bustos-Lozano, H. L., & Belén, M. (2025). The Role of Formative Assessment in Higher Education: Strategies to Improve Learning and Knowledge Retention. *Journal of Higher Education Theory and Practice*, 23, 5441–5455.
- Salehi, M., & Daryabar, B. (204). Self-and Peer Assessment of Oral Presentations: Investigating Correlations and Attitudes. *English for Specific Purposes World*, 15(42), 1–12.
- Shepard, L. A. (2000). The Role of Assessment in a Learning Culture. *Educational Researcher*, 29(7), 4–14.
- Sitzmann, T., & Johnson, S. K. (2012). When is Ignorance Bliss? The Effects of Inaccurate Self-Assessments of Knowledge on Learning and Attrition. *Organizational Behavior and Human Decision Processes*, 117(1), 192–207.
- Stančić, M. (2021). Peer Assessment as a Learning and Self-Assessment Tool: A Look Inside the Black Box. *Assessment & Evaluation in Higher Education*, 46(6), 852–864.
- To, J., & Panadero, E. (2019). Peer Assessment Effects on the Self-Assessment Process of First-Year Undergraduates. *Assessment & Evaluation in Higher Education*, 44(6), 920–932.
- Topping, K. J. (2009). Peer Assessment. *Theory into Practice*, 48(1), 20–27.
- Wanner, T., & Palmer, E. (2018). Formative Self-and Peer Assessment for Improved Student Learning: The Crucial Factors of Design, Teacher Participation and Feedback. *Assessment & Evaluation in Higher Education*, 43(7), 1032–1047.
- Weeden, P., Winter, J., & Broadfoot, P. (2002). *Assessment*. Psychology Press.
- William, D. (2011). What is Assessment for Learning? *Studies in Educational Evaluation*, 37(1), 3–14.