

## STUDENT-BASED ASSESSMENT VS. TEACHER ASSESSMENT: IS THERE ANY CONSISTENCY?

**Setiawan Adi Prasetyo**

*Universitas Muhammadiyah Malang*

[adisetiawan057@gmail.com](mailto:adisetiawan057@gmail.com)

**Aninda Nidhommil Hima**

*Universitas Muhammadiyah Malang*

[aninda@umm.ac.id](mailto:aninda@umm.ac.id)

**Abstract:** In Second Language Writing class, the use of student-based assessment such as self and peer assessments has been increasingly promoted to help the students create good written products. Nevertheless, the inclusion of the student-based assessment result in determining the students' final score still becomes a controversial matter since some studies prove that the student-based assessment and teacher assessment are not in agreement and tend to be questionable. To bridge the gap, this study aims at investigating the level of consistency among self-assessment, peer assessment, and teacher assessment in a writing course. Through a correlational research design, the current study involved 21 students who took a paragraph writing course and experienced in self and peer assessing. To collect the data, the students were asked to write an expository paragraph that was assessed through self, peer and teacher assessments. The data were then analyzed by using the Spearman Rank Order Correlation to answer whether or not there is consistency among self-assessment, peer assessment, and instructor assessment. The results reveal that the students provided the same score in self, peer and teacher assessment. However, the correlations in both self and peer-assessment, as well as self and teacher assessment, are not considered statistically significant. The significant difference occurs in the correlation between peer and teacher assessments result only. From this finding, it can be recommended to the teacher to include the result of peer assessment in determining the students' final grade.

**Keywords:** Self-assessment, Peer Assessment, Teacher Assessment

### BACKGROUND

In a few years back, the use of student-based assessment that are self-(Andrade et al., 2010; Kirby & Downs, 2007; Porto, 2001) and peer assessment (Tsai, Lin, & Yuan, 2002; Vu & Dall'Alba, 2007; Panadero, Romero, & Strijbos, 2013) in the classrooms has gained to foster the student-centered learning especially in writing class.

Self-assessment is a form of assessment which facilitates the students to evaluate the quality of their work (Andrade, Du, and Mycek, 2010). In other words, through self-assessing the students can judge their work by themselves.

Self-assessment is believed can provide some potential benefits. Self-assessment can shape students' confidence, boost up the students' effort, and facilitate the students' awareness in improving their performance and competence (Blue, 1994). It also can support and promote the effective teaching and learning process, help the students in establishing learning goals, and facilitate the students' need in the classroom (Oscarson 1989 cited in Esfandiari and Myford 2013). Although some experts claim that self-assessment can be

beneficial for the students, the limitation on its use is also found. Some students tend to overestimate their capability and performance, this can make their score better (Linblom-Ylaine, 2006). Filter (2012) and Leach (2012) concluded that self-assessment would not bring reasonable result if the students do not get very much prepared and well-trained.

Another type of student-based assessment is peer assessment. Peer assessment is a way of measurement where the students give their judgments both in process and outcomes to their peers' work (van Gennip, Segers and Tillema, 2009). It also can be used as a reference for the students to determine, ponder, and categorize the quality, value, and level of the performance of their equal level students (Topping, 2009).

Davies (2002) states that peer assessment is also assumed can boost the students' responsibility in the learning process. It is considered to bring a good impact on the students' professional life, especially in having some important life skills such as autonomy, judgments, responsibility (Kilic, 2016). He also added that peer assessment can help students find a chance to the rational thinking through the given feedback. When the students receive and give feedback, they will obtain additional knowledge, for instance, how to analyze and solve a complex problem; Jung, 2016).

Peer assessment also draws some weaknesses (Esfandiari & Myford, 2013). As peer assessment is done by the students, it spends much time and will be workload for students and teacher. Adequate time is important for training, and preparing since the students are the beginners (Topping, 2009). The peer assessor will feel uncomfortable when he or she assesses their classmates because they are worrying they will hurt their friend (Vu and Dall'Alba, 2007). It may cause unrealistic award that the student gives for assessment. The last disadvantage is an ethical problem that faced students. Some students will appraise their peers by dealing with different background and achievement. As a result, there will be friction between each other's, also they hurt each other.

In the development of the study, the combination of self and peer assessment is interesting to be analyzed. Abolfazli and Sadeghi (2013) compared self and peer assessment and found that peer assessment can significantly increase the students' learning. Further, Topping (2003) cited in Dominguez *et al.*, (2016) added compared to peer assessment, the validity of Self-Assessment inclined to be lower and more variable. On the other hand, Peer Assessment and Self-Assessment has shown the effect on the learning process. Chen (2010), through online self and peer assessment, he proved that the student-based assessment enabled the students actively participate in the learning process, such as giving feedback, self-reflection, and sharing the idea. Another study Kilic (2016) mentioned that the students were accustomed to do self-assessment after finishing their work because they have the experience to evaluate their task through the provided criteria.

Although there has been increasing research interests about the use of student-based assessment in the classrooms, the investigations toward the student-based assessment compared to teacher-assessment are still highlighted in the literature (Esfandiari and Myford, 2013; Panadero *et al.*, 2013; Alias, Masek and Salleh 2015; Kilic 2016). Instructor assessment or Teacher assessment is the method that is used most to determine the level of students' skill and knowledge in the educational aspect (Chang *et al.*, 2012). The studies in the area of the agreement level and the level of the severity among those types of assessment still become controversial. Some studies prove that the student-based assessment and teacher assessment show inconsistency and are not inconclusive (Alias *et al.*, 2015; Panadero *et al.*, 2013). The result of these studies shows that the score for self and peer assessment differ from teacher assessment where the score from self and peer is much higher rather than teacher assessment.

In contrast, Kilic (2016) mentioned that there is a consistency between self and teacher assessment, although there is a significant difference between peer assessment and self-teacher assessment in writing. The result revealed that the means of peer assessment is higher than self and teacher assessment. This result is in line with Magin & Helmore (2001); Rudy, Fejfar, Griffith & Wilson (2001) cited in Kilic (2016) that found the students tend to have a higher score in peer assessment than teacher assessment. Chen (2010) also found out that teacher and peer assessment results were not in agreement as well.

On the other hand, the consistency among self-, peer-, and teacher assessment have been found (Sadler & Good, 2006; Sung, Chang, Chiou, & Hou, 2005; Bouzidi & Jaillet, 2009). They stated that these differences may be caused by the assessment procedure, the levels of students, and the scoring rubric that is used. Consequently, the teachers never take the result of the student-based assessment into account to be considered in determining the students' final grade. The pros and cons in the result of investigating the level of agreement among the score in the self, peer, and teacher- assessment in the literature indicate that this scope still needs to be investigated to provide more empirical data in the literature. Therefore, the current study aims to examine the level of consistency among self-, peer-, and teacher assessment in writing class.

## METHOD

This current study focuses on investigating the use of self, peer-, and teacher assessment in which correlational research design was used to explore the consistency among them.

### Participants

A class of paragraph writing course consisting of 21 students (13 females and 8 males) was selected for the study since they have experience in doing self and peer assessment.

### Data Collection

In this current study, the researcher used a test and scoring rubrics for self and peer assessments as the instruments to collect the data. This test was used to collect the students' writing about an expository text that was assessed by using the scoring rubrics for self and peer assessment. The analytical scoring rubric was used to get more consistent and reliable data as it provides many details. By using this type of rubric, the raters can easily diagnose the weakness and strength of the writing product. In this case, the raters had been trained previously in order they understand how to use the rubric.

### Research Procedure

This research was done in three meetings. In the beginning, the students were given a test to make an expository text individually. In the next meeting, the lecturer distributed the scoring rubric and explained how to use it before they were asked to do self-assessment. After they finished self-assessing, the students were given the scoring rubric for doing peer assessment and trained how to use it as well. Then, they were asked to do peer assessment toward their friends' writing. Lastly, in the third meeting, the teacher assessment was done by the lecturer.

Tabel. 2.1 Research Procedure

Meeting	Agenda
1	Asking the students to write an expository text
2	Doing self- Assessment & peer assessment
3	Doing teacher assessment

### Data Analysis

For the data analysis, the researcher uses SPSS 21 to analyze the consistency (the correlation) among self-, peer-, and instructor assessment. Before measuring the consistency of those three different methods of assessment, the normality of data is important to check to decide an appropriate tool. Shapiro-wilk test was the first way to check the normality of the result of three assessments. After checking the normality of the data, the Spearman's Rank Order Correlation was applied to know the consistency among three assessments since the data was non-parametric. Spearman's correlation coefficient is a statistical measurement of the strength of a monotonic relationship between paired data. In a sample, it is symbolized by  $r_s$  and is by design constrained as  $-1 \leq r_s \leq 1$ . To conclude, the result of statistical

calculation about the consistency of those three different results of the assessment are compared to the following guide for the absolute value of  $r_s$ .

Table 2.2. The Guide for the Absolute Value of  $r_s$ .

No	Range	Descriptions
1.	.00 - .19	“very weak”
2.	.20 - .39	“weak”
3.	.40 - .59	“moderate”
4.	.60 - .79	“strong”
5.	.80 - 1.0	“very strong”

In addition to the coefficient correlation, *the* p-value of the data is also checked by comparing it to the significance level. If *the* p-value is lower than 0,05 H0 is rejected while if the p-value is greater than 0,05 H0 is accepted.

## FINDINGS AND DISCUSSION

In this chapter, the researcher provides the findings and discussion about the statistical result of the focus. The researcher will elaborate on the focus of the study such as the result of an examination between self-, peer-, and teacher assessment.

### Findings

The data gained from the assessment form were analyzed by using SPSS 17. In this study, the researcher applied Shapiro-wilk test to ensure the data met the normality. In this Shapiro-wilk test, the data is normal if sig > 0,05. From those result, the researcher concludes that these results were not normal (non parametric). The result did not meet the normality because its sig score is under 0,05. This may be caused there are scores in those assessments techniques which are the outlier. The results of the test are as follows:

#### *The Data Normality*

##### Self-Assessment Shapiro-wilk test

In Shapiro-wilk test for self-assessment result, the significance was found 0,04 which was smaller than the critical level of significance (sig<0,05). From this result, it could be concluded that self-assessment data included nonparametric. The result can also be concluded from Kolmogorov-Smirnov, the significance was found 0,030. It indicated that the critical level of significance is higher than the significance of self-assessment (sig<0,05). This supported Shapiro-wilk test to prove that self-assessment included to nonparametric.

Table 3.1 Shapiro-wilk test for Self-Assessment

	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistic	Df	Sig.	Statistic	Df	Sig.
self_assessment	.199	21	.030	.846	21	.004

The data from peer assessment was similar in which the significance was found as 0.007. This numeral was smaller than the critical level of significance (sig<0,05). Therefore, it could be concluded that the data from peer assessment was nonparametric.

Table 3.2 Shapiro-wilk Test for Peer Assessment

	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistic	Df	Sig.	Statistic	Df	Sig.
peer_assessment	.163	21	.048	.862	21	.007

The last analysis of normality was not found in teacher assessment as well. This result was similar to peer assessment in which the significance was smaller than the critical level of significance ( $\text{sig} < 0,05$ ). The normality was not found in here.

Table 3.3 Shapiro-wilk Test for Teacher Assessment

	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistic	Df	Sig.	Statistic	Df	Sig.
instructor_assessment	.220	21	.009	.847	21	.004

Based on the result of Shapiro-wilk test, the scores of self-, peer-, and instructor assessments were included nonparametric. To answer whether or not those three assessments are correlated one to another, Spearman’s Rank Order Correlation is applied.

The Consistency among Self, Peer, and Teacher Assessment

The coefficient of Spearman’s Rank Order Correlation reveals the score of  $r_s = 0.39482$ , meaning that the students' score in the self and peer assessment have a positive weak correlation. The increase the score in the self-assessment process, the score from peer-assessing is increasing as well. However, the p-value (2-tailed) = 0.07652. It indicated that by normal standards, the association between the two variables would not be considered statistically significant.

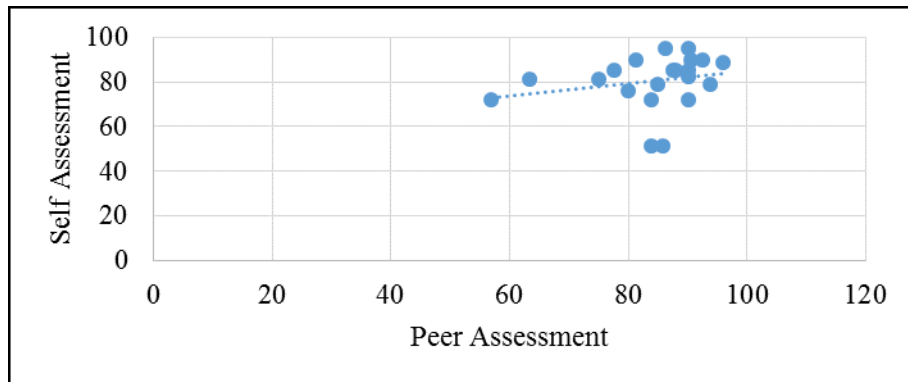


Figure 3.1. The Correlation between the Self and Peer Assessment Result

A very strong correlation was found between self-assessment score and teacher assessment. It was proved by the coefficient correlation in which  $r_s = 0.08656$ . However, the same as when the self-assessment compared to peer assessment, the p-value showed is 0.7091. It is concluded that by normal standards, the association between the two variables is not statistically significant.

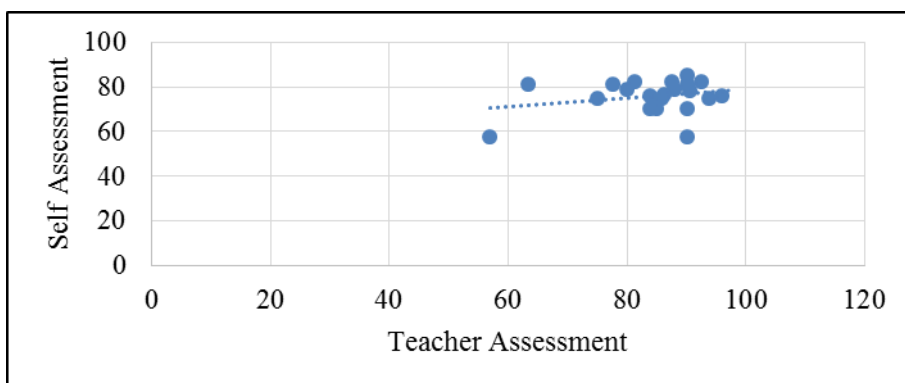


Figure 3.2 the Correlation between the Self and Teacher Assessment Result

The next comparison was examined between peer and teacher assessment. The analysis is based on p-value and a critical level of significance. Based on Spearman Rank Order calculations, there was a positive moderate correlation between the paired variables that were shown by  $r_s = 0.54656$ . Different from the previous comparisons, the association between the students' score resulted from peer and teacher assessment was considered statistically significant as  $p$  (2-tailed) = 0.01036.

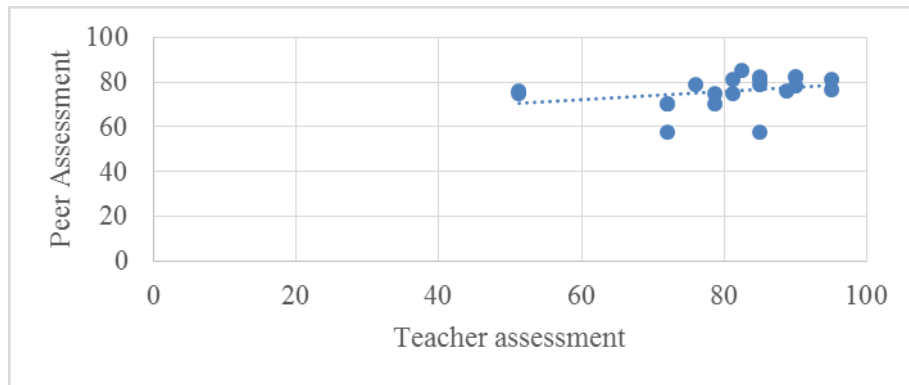


Figure 3.3 The Correlation between the Peer and Teacher Assessment Result

## Discussions

From this result, it can be concluded that there is consistency between self and peer assessment. This result is in line with Sung et al (2005) that found self-assessment is in agreement with peer assessment. Supported by Alias et al (2015), the students assess their product similar to their peer assessment, and it is much higher than the assessment given by their teacher. However, these result is different from Abolfazli & Sadeghi, (2013); Sung et al (2005) in which peer assessment was performing higher than an assessment that is done by the students itself.

The next finding is about the strong correlation between self and teacher assessment. This result is supported by Sung et al. (2005) showed there is a high level of consistency between self and teacher assessment. On the other hand, this finding was not considered different significantly. de Grez, Valcke and Roozen (2012); Dominguez et al (2016) stated that the score which is given by students tends to be much higher than given by the teacher. In this study also revealed, the students who have critical and detailed comments might help the others in revising their work. Thus makes the students' assessment is higher than teacher assessment (Tsai et al., 2002). Also, Chen (2010) added, the cause of self-assessment is higher than teacher assessment predicted that the students tend to revise their product after they receive feedback and correction from their peers.

In investigating peer and instructor assessment association, there was statistical significance if those paired variables have a moderate correlation. This conclusion was not similar to Chen (2010) in which there is no consistency between students and instructor assessment. Students here include self and peer assessment. In this study, the inconsistency exists because the student's tendency to revise their work after they are receiving peer feedback from their friends.

The result regarding the correlation among self-, peer-, and instructor assessment showed that the consistency exists in those kinds of assessment, however, only in peer and teacher assessment the significant difference is shown. Sung et al. (2005), the consistency in high level found among three types of assessment may occur because there is a different level of students, assessment environment, and assessor training.

## CONCLUSION

The conclusion of this current research has been obtained from the finding and discussion. The result based on statistical analysis has revealed that the consistency exists among self, peer and teacher assessment. It happens because the students were trained before doing the assessment. This training is important to be conducted in order the students understand how to use the instrument which is scoring rubrics. However, the correlations in both self and peer-assessment, as well as self and teacher assessment, are not considered statistically significant. The significant difference occurs in the correlation between peer and teacher assessment result only. From this finding, it can be recommended to the teacher to include the result of peer assessment in determining the students' final grade.

The further researchers are expected to do deep investigation in the area of the implementation student-based and teacher-based assessment in the classrooms. Since this study only uses writing class to be investigated, further researchers are suggested to conduct the research in a different subject, such as speaking, listening and reading by involving a larger number of participants. These will be the gaps for the next researchers.

## ACKNOWLEDGMENT

The authors express their gratitude to Directorate of Research and Community Service Center, Universitas Muhammadiyah Malang for the funding support.

## REFERENCES

- Abolfazli, Z., & Sadeghi, K. (2013). The effect of assessment type ( self vs . peer ) on Iranian university EFL s course achievement. *Procedia - Social and Behavioral Sciences*, 70, 1552–1564. <https://doi.org/10.1016/j.sbspro.2013.01.223>
- Alias, M., Masek, A., & Salleh, H. H. M. (2015). Self, Peer and Teacher Assessments in Problem Based Learning: Are They in Agreements? *Procedia - Social and Behavioral Sciences*, 204(November 2014), 309–317. <https://doi.org/10.1016/j.sbspro.2015.08.157>
- Andrade, H. L., Du, Y., & Mycek, K. (2010). Rubric-referenced self-assessment and middle school students' writing. *Assessment in Education: Principles, Policy and Practice*, 17(2), 199–214. <https://doi.org/10.1080/09695941003696172>
- Blue, G. M. (1994). Self-Assessment of Foreign Language Skills: Does It Work? *CLE Working Papers*, 3, 18–35.
- Chang, C. C., Tseng, K. H., & Lou, S. J. (2012). A comparative analysis of the consistency and difference among teacher-assessment, student self-assessment and peer-assessment in a Web-based portfolio assessment environment for high school students. *Computers and Education*, 58(1), 303–320. <https://doi.org/10.1016/j.compedu.2011.08.005>
- Chen, C. (2010). Computers & Education The implementation and evaluation of a mobile self- and peer-assessment system. *Computers & Education*, 55(1), 229–236. <https://doi.org/10.1016/j.compedu.2010.01.008>
- Davies, P. (2002). Using student reflective self-assessment for awarding degree classifications. *Innovations in Education and Teaching International*, 39(4), 307–319. <https://doi.org/10.1080/13558000210161034>
- De Grez, L., Valcke, M., & Roozen, I. (2012). How effective are self- and peer assessment of oral presentation skills compared with teachers' assessments? *Active Learning in Higher Education*, 13(2), 129–142. <https://doi.org/10.1177/1469787412441284>
- Esfandiari, R., & Myford, C. M. (2013). Severity differences among self-assessors, peer-assessors, and teacher assessors rating EFL essays. *Assessing Writing*, 18(2), 111–131. <https://doi.org/10.1016/j.asw.2012.12.002>
- Jaime, A., Ana, S., & Blanco, M. (2016). Computers in Human Behavior A comparative analysis of the consistency and difference among online self- , peer- , external- and instructor-assessments : The competitive effect. 60, 112–120. <https://doi.org/10.1016/j.chb.2016.02.061>

- Kilic, D. (2016). An Examination of Using Self-, Peer-, and Teacher-Assessment in Higher Education: A Case Study in Teacher Education. *Higher Education Studies*, 6(1), 136. <https://doi.org/10.5539/hes.v6n1p136>
- Kirby, N. F., & Downs, C. T. (2007). Self-assessment and the disadvantaged student: Potential for encouraging self-regulated learning? *Assessment and Evaluation in Higher Education*, 32(4), 475–494. <https://doi.org/10.1080/02602930600896464>
- Leach, L. (2012). Assessment & Evaluation in Higher Education Optional self-assessment : some tensions and dilemmas. *Assessment & Evaluation in Higher Education*, 37(2), 137–141. <https://doi.org/10.1080/02602938.2010.515013>
- Linblom-Ylänne. (2006). Self-, peer- and teacher-assessment of student essays. *Active Learning in Higher Education*, 7(1), 51–62. <https://doi.org/10.1177/1469787406061148>
- Panadero, E., Romero, M., & Strijbos, J. W. (2013). The impact of a rubric and friendship on peer assessment: Effects on construct validity, performance, and perceptions of fairness and comfort. *Studies in Educational Evaluation*, 39(4), 195–203. <https://doi.org/10.1016/j.stueduc.2013.10.005>
- Porto, M. (2001). Cooperative writing response groups and self-evaluation. *ELT Journal*, 55(January), 38–46.
- Sadler, P., & Good, E. (2006). The impact of self-and peer-grading on student learning. *Educational Assessment*, 11(1), 37–41. [https://doi.org/10.1207/s15326977ea1101\\_1](https://doi.org/10.1207/s15326977ea1101_1)
- Sung, Y. T., Chang, K. E., Chiou, S. K., & Hou, H. T. (2005). The design and application of a web-based self- and peer-assessment system. *Computers and Education*, 45(2), 187–202. <https://doi.org/10.1016/j.compedu.2004.07.002>
- Topping, K. J. (2009). Peer assessment. *Theory into Practice*, 48(1), 20–27. <https://doi.org/10.1080/00405840802577569>
- Tsai, C., Lin, S. S. J., & Yuan, S. (2002). Developing science activities through a networked peer assessment system. *COMPUTERS AND EDUCATION*, 38, 241–252.
- Van Gennip, N. A. E., Segers, M. S. R., & Tillema, H. H. (2009). Peer assessment for learning from a social perspective: The influence of interpersonal variables and structural features. *Educational Research Review*, 4(1), 41–54. <https://doi.org/10.1016/j.edurev.2008.11.002>
- Vu, T. T., & Dall’Alba, G. (2007). Students’ experience of peer assessment in a professional course. *Assessment & Evaluation in Higher Education*, 32(5), 541–556. <https://doi.org/10.1080/02602930601116896>