THE VALIDITY AND RELIABILITY OF SELF-ASSESSMENT ON READING COMPREHENSION AT JUNIOR HIGH SCHOOL

Irfan Susiyana Putra
Universitas Islam Malang, Jawa Timur, Indonesia
Email: Irfansinkrom@gmail.com

ABSTRACT

This study aims to investigate whether self-assessment is valid and reliable for assessing reading comprehension. Participants in this study were 18 male students and 20 female students from a junior high school. This research is quantitative research designed with mixed methods, causal-comparative and correlative research. Independent t test was used to analyse validity, and correlation analysis was used to test reliability. Researchers used two instruments in the data collection process, (1) a self-assessment questionnaire and (2) a reading comprehension test. This research concludes that 1) self-assessment is valid because male and female students do not carry out self-assessment differently. This means that gender is not a construct irrelevance factor that affects the validity of self-assessment. 2) Self-assessment is reliable or consistent because there is a positive correlation between self-assessment scores and reading comprehension tests. Two statistical analyses were used to examine the data collected. First, independent t-tests were used to investigate the influence of gender on self-assessment scores. From data analysis, this study revealed that there were no significant differences between male and female students in self-assessment. This means that gender is not proven to be an irrelevant factor to the construct that affects validity. Thus, the self-assessment of reading comprehension in junior high school is valid.

KEYWORDS	*Validity, Reliability, Self-Assessment.*

INTRODUCTION

Language assessment is a continually evolving area. In the past, assessments were primarily summative, meaning they evaluated the results after a learning process. Subsequently, formative assessment emerged to evaluate the learning process itself. The most traditional tool for assessment was a test. Today, the field has expanded to include various forms of alternative assessments, such as portfolios, journals, projects, observations, and more. Previously, only teachers served as assessors; however, now students can also assess themselves and their peers through self-

assessment and peer-assessment. The emergence of these two assessments is based on the awareness that students need to recognize and evaluate their abilities to employ autonomous learning and cooperative learning (Brown, 2003).

Among the two types of assessments, self-assessment is the one most often discussed. Opting for self-assessment as an alternative evaluation method in learning comes with both benefits and drawbacks. One of the advantages of self-assessment is related to student learning independence, where students can reflect on their own learning, find out the extent of their learning results, become a benchmark for the results of students' hard work, and encourage them to improve their abilities (Lesmana & Rokhyati, 2020; Lianisa, 2014). Self-assessment is also useful in evaluating, improving student performance and motivation in learning (Vasile-iadou & Karadimitriou, 2021). In addition, it turns out to help students improve their critical thinking (Sharma et al., 2016).

Conversely, the drawbacks of self-assessment commonly highlighted in research include the issue of subjectivity. Students potentially underestimate or overestimate their actual grades (brown, 2004). Self-assessment is sometimes unbeneficial because some students do not know what self-assessment is exactly for (Lianisa, 2014). Contrariwise, if students are told that this self-assessment replaces a test, they are more likely to manipulate their grades. (Ross, 2006).

In light of these drawbacks, questions arise regarding the validity and reliability of self-assessment. Numerous researchers have explored this issue. Matsuno (2009) who studied validity and reliability of self-assessment in writing skill. He revealed that self-assessment is not valid because students who have high ability students underestimated themselves. Click here to enter text.Langan et al. (2008) also revealed that gender also differentiated self-assessment. male students are over confident and scores themselves higher than their actual grades. Besides ability level and gender, ages (Ross, 2006); learners' attitude, subjectivity, intention, and self-efficacy (Yan et al., 2019).

These factors can then be considered construct-irrelevant factors that may render self-assessment invalid. However, these factors do not inherently invalidate all self-assessments or make them unreliable. Several studies have successfully demonstrated the validity and reliability of self-assessment. First, (Shameem, 1998) asked Fuji immigrants to self-assess their proficiency and the results were valid. Second, Chang et al. (2013) researched high school students' self-assessment of their portfolios and the results were valid and reliable. The third is Mistar (2011) who proves that gender and age do not influence students' self-assessment of listening and speaking skills. Although all three produced similar findings, the studies differed in subjects, skills, levels, and goals. Determining the scope is very important because according to Ross's (2006) suggestion that research on self-assessment cannot be compared in different fields. Therefore, research on self-assessment is still very limited compared to the breadth of its scope and needs.

In this research the skill chosen was reading, especially at junior high school level. As mentioned above, the subjects of previous research were mostly adult learners. Research on young students needs to be explored. Moreover, research on reading at this level has never been studied by previous researchers, previous research has always been related to other aspects such as speaking and listening which

are considered important because they are verbal aspects that are directly used in practice with the person you are talking to using conversation with someone. It is rare for researchers to raise research aspects that are based on reading because they are considered boring aspects for students due to the increasing interest and interest in reading among teenagers in this era. Assessment of reading ability is very important. They are not aware that reading is an important aspect in increasing knowledge considering that reading is one of the important skills as an input in the acquisition of a second language. Reading can help students develop their understanding and vocabulary which is useful for use when having conversations with other people. In fact, it is also written in our daily slogan that we should leave behind, namely "books are windows to the world"

However, reading also has a broad scope, such as reading to find information using scanning techniques or reading to enjoy and understand a text. Therefore, researchers will narrow it down to reading comprehension. This decision is supported by the benefits of implementing self-assessment in reading classes which has been proven by several studies (Nourdad & Asghari, 2017; Permata et al., 2015), but no one has tested the validity and reliability of student self-assessment. reading comprehension.

Due to the previous gap. Researchers want to test the validity and reliability of self-assessment of reading comprehension in junior high schools. with the aim of finding out the validity of self-assessments across genders, whether there are differences and also finding out whether the results of these assessments will affect their grades. which is if their self-assessment value is in accordance with their real value, which is if their self-evaluation is high then their real value should also be high and vice versa if their self-evaluation is low then their real value should be low.

Researchers hope to be able to provide solutions and complete the research. namely 1. Prove whether the self-assessment of male students is different from that of female students? 2. Does the higher a student's score on a self-assessment, the higher their score on a reading comprehension test?

RESEARCH METHOD

This study is a quantitative research designed in mixed method; causal-comparative and correlative research. Causal comparative research as stated by Soesilo (2018) is causal comparative research, which is a type of research to test a hypothesis regarding the cause and effect relationship of several variables. In this case, variables What will be studied is tutoring (X) and understanding learning styles (Y). 1) Causal-comparative research is used to examine the validity of self-assessment by investigating whether gender serves as a construct-irrelevant factor influencing self-assessment. If there is no significant difference between the scores of male and female students, it indicates that the self-assessment is valid. 2) Correlational research is employed to assess reliability by comparing self-assessment results with a reading comprehension test. If the findings are significant and positive, it suggests that self-assessment is reliable.

The analysis used to carry out different test. This can be in the form of a t-test or Anova, which uses interval and ratio data, which can be used to carry out the difference test. Before carrying out statistical analysis of the difference test,

researchers must test the data to ensure whether the data distribution is homogeneous and normal.

The basic principles of difference test analysis are as follows: 1. Parametric statistical analysis (t-test, anova) is used if there are two or more groups that have different values (scores). In the t-test, only one group is used to differentiate the Pre-Test results from the Post-Test results. 2. If the data, as a dependent variable, differs on an interval or ratio data scale, then the difference will be in the form of the average data from each group. However, as an independent variable, the differentiator can be ordinal or nominal data. If the data, as a dependent variable, is on a nominal or ordinal scale. Non-parametric statistical analysis, such as Mann Whitney, is used if the data, as the dependent variable, is either ordinal or nominal. 3. In correlation tests, generally, if there is a relationship between the variables used (X->Y), then that will also produce differences.

The subjects of this research were 38 students grade IX at Sunan Bonang Middle School, Situbondo, East Java who were deemed to meet the criteria set by the researcher. They are students who can assess themselves and their abilities. They consisted of 18 men and 24 women.

This research was carried out in the following stages:

1. Students are assessed in a 45-minute reading comprehension test. Students are asked to read a passage that has been provided by the teacher within the given time limit.
2. After reading and taking the reading test, students are then asked by the teacher to fill out a self-assessment questionnaire for 10 minutes.
3. Reading comprehension test scores and self-assessments are measured
4. The scores are calculated in SPSS to determine the difference between the self-assessment scores of male and female students; and the level of significance.
5. Scores are calculated in SPSS to measure the correlation coefficient between self-assessment scores and reading comprehension tests; and the level of significance.

Data were collected using two instruments: (1) a self-assessment questionnaire (Appendix 1) and (2) a reading comprehension test (Appendix 2). In the questionnaire, students were asked to evaluate their reading skills, with each skill measured through 10 items. For each item, students had to select one option from 1 to 5, indicating their ability to answer the questions in the reading test. The Likert scale corresponds to 1 for 'very difficult', 2 for 'difficult', 3 for 'neutral', 4 for 'easy', and 5 for 'very easy'. The second instrument consisted of a 20-item multiple-choice test designed to assess students' reading comprehension.

According to Azwar (1999), basically research Causal-comparative is "ex post facto", meaning that the data needed by researchers for research is data that is presented after all the events of interest happen. Then the researcher selects one or more of the effects (dependent variables) and tests the data by tracing the time that has passed, looking for the causes, looking at the relationships, and understanding the meaning of it.

Two statistical analyses were employed to evaluate the collected data. The independent samples t-test is a statistical method used to determine whether there

is a statistically significant difference between the means of two independent groups. This test is effective for comparing the means of two groups to assess any differences between them. This section outlines the fundamental concepts and procedures associated with independent t-tests. First, an independent t-test was conducted to examine the impact of gender on self-assessment scores, comparing the mean self-assessment scores of male students to those of female students. Second, Pearson Product-Moment correlation analysis was utilized to investigate the relationship between self-assessment scores and test scores.

RESULT AND DISCUSSION

The findings of this study are presented according to the research questions. As mentioned earlier, this study explores two questions as follows.

Do male students self-assess differently from females?

Table 1. The Result of Independent Samples T-Test.

Gender	N	Mean	Std. Deviation	t-value	.Sig
Male	18	25.39	4.791	1.503	.142
Female	20	23.40	3.299		

According to the comparison results shown in Table 1, the mean self-assessment score for male students is 25.39, while the average score for female students is 23.40. At first glance, male students appear to self-assess higher than female students. However, the level of significance ($p<.143$) is below the .01 threshold, indicating that the difference (t-value) is not statistically significant. Therefore, it can be concluded that gender does not influence self-assessment. In other words, self-assessment is considered valid because gender is not a construct-irrelevant variance in self-assessment.

Do the higher the students' scores in self-assessment, the higher their scores in reading comprehension test?

The correlation results between students' self-assessment and reading test scores are presented in Table 2. In this table, the correlation coefficient is .489, with a significance level of .002. This indicates a positive and significant correlation (at the .01 level). Therefore, it suggests that higher self-assessment scores are associated with higher scores on the reading test. Additionally, this infers that students' self-assessments accurately reflect their actual reading comprehension skills, demonstrating the reliability of the self-assessment.

Table 2. The result of Pearson-Product Moment Correlation

		Self-Assessment	Reading Test
Self_Assessment	Pearson Correlation	1	.489[**]
	Sig. (2-tailed)		.002
	N	38	38
Reading_Test	Pearson Correlation	.489[**]	1
	Sig. (2-tailed)	.002	
	N	38	38

The first focus of this research is the validity of Self-Assessment. Validity is a parameter for evaluating if the variable being measured is the variable that the researcher intends to investigate (Newton & Shaw, 2014). In other words, valid means accurate or precise. Validity of assessment is a vast topic. Therefore, Brown (2004) divides validity into content, construct, criterion, consequential, and face validity. This study focuses on construct validity, where the evidence is the absence of construct-irrelevant factors that can affect the relevance of self-assessment. According to Ross (2006), one of the construct-irrelevant factors is gender.

This study has revealed that self-assessment has construct-related evidence because it was found that gender does not affect self-assessment. It refutes Langan et al. (2008), Langan's finding that male students overestimate their ability while females underestimate it. Confidence, courage, and indifference to the judgments of others, which are identical traits of men, are assumed to affect self-assessment. Those are also known as self-efficacy, which Yan et al. (2019) mentioned it as a factor influencing self-assessment. However, these two studies were not proven in the findings of this current research.

On the other hand, the findings that self-assessment is reliable is in line with several studies, such as (Chang et al., 2013; Mistar, 2011; Shameem, 1998), which also found that self-assessment is valid. However, the ways in examining the validity are different. Chang et al. (2013) examined the validity by comparing self-assessment with the final scores of a course and teacher assessment. Meanwhile, Shameem (1998) compared it to a performance test on speaking skill. Mistar (2011) tested self-assessment on listening and speaking. There have certainly been other earlier studies in this area. However, Ross (2006) stated that since the discussion of self-assessment is diverse, research findings across skills, education levels, validity and reliability types, and so forth, are not comparable. For instance, the study conducted by Mistar on the validity and reliability of self-assessment in speaking and listening skills at the university level cannot be generalized or compared to other skills at the elementary and secondary school levels. Therefore, researchers' involvement in this area will ultimately enhance the theory of self-assessment.

The second focus of this research is the reliability of self-assessment. Reliability in assessment is the quality of assessment to be trusted or believed that can be proven by consistency (Race, 2001). An assessment is deemed reliable if its results are consistent or stable over time. In this study, the reliability of self-assessment was evaluated by comparing it to a reading test. The findings indicated that self-assessment is reliable, as the correlation between the two assessments was both positive and significant. This result contradicts the finding of Matsuno (2009) that high-ability students tend to underestimate themselves, whereas low-level students overestimate. In this study, the situation was quite the opposite. As the reading test scores increased reflecting their actual abilities so did the self-assessment scores.

In Contrary, the finding that self-assessment is reliable is in line with Chang et al. (2013); Mistar (2011); and Shameem (1998), but the types of skills and education levels they studied are different from this current study. However, a study by Sung et al. (2005) found that students aged 14-15 could do a reliable self-assessment that was relevant to their original abilities and the teacher assessment. This supports the validity and reliability of self-assessment among junior high school

students. This finding is particularly beneficial for teachers, as it can help reduce their workload; by having students self-assess, teachers can gain insights into their true abilities.

Based on the researcher's experience in guiding self-assessment at SMP Sunan Bonang Situbondo, it was also found to be advantageous for students in terms of acquiring new knowledge. Previously, they only focused on identifying which answers were correct or incorrect in a multiple-choice reading test. However, with the 10 self-assessment items explained beforehand, students became more aware of the types of questions involved in reading comprehension. Those questions cover four of the five subskills of reading comprehension; lexical, literal, interpretive, applied, and affective comprehension (McAndrews, 2020). However, only applied comprehension is not covered. So, the results of this self-assessment provide information for students and teachers on which skills they excel in and on which skills they are weak. It is associated with the suggestion of Andrade (2019) that although self-assessment can be used for summative purposes because of its validity and reliability, applying it as a technique in teaching is also fruitful, especially in terms of increasing students' autonomous learning.

CONCLUSION

As mentioned earlier, the first research question examines whether there is a difference in self-assessments between male and female students. This question seeks to investigate the validity of self-assessment. The data analysis revealed no significant differences in self-assessment scores between male and female students, indicating that gender does not serve as an irrelevant factor affecting validity. Consequently, the researchers conclude that self-assessment of reading comprehension in junior high school is valid. The second question explores whether a higher self-assessment score correlates with a higher score in the reading comprehension test. Based on the research and analyses conducted, the researchers determined that this relationship is indeed confirmed, as there is a positive and significant correlation between the scores from the two assessments. This suggests that self-assessment is consistent and reliable.

REFERENCES

Boud, D. (1995). Enhancing Learning through Self Assessment. Taylor & Fancis.

Brown, H. D. (2003). Language Assesment Principles and Classroom Practice. In Pearson Longman.

Chang, C. C., Liang, C., & Chen, Y. H. (2013). Is learner self-assessment reliable and valid in a Web-based portfolio environment for high school students? Computers and Education, 60(1), 325–334. https://doi.org/10.1016/j.compedu.2012.05.012

Falchikov, N., & Boud, D. (1989). Student Self-Assessment in Higher Education: A Meta-Analysis. Review of Educational Research, 59(4). https://doi.org/10.3102/00346543059004395

Langan, A. M., Shuker, D. M., Cullen, W. R., Penney, D., Preziosi, R. F., & Wheater, C. P. (2008). Relationships between student characteristics and self-

, peer and tutor evaluations of oral presentations. Assessment and Evaluation in Higher Education, 33(2). https://doi.org/10.1080/02602930701292498

Lesmana, N., & Rokhyati, U. (2020). The Implementation of Doing Self-Assessment in Higher Education. Journal of English Language Studies, 5(1), 60–72. https://doi.org/http://dx.doi.org/10.30870/jels.v5i1.7210

Lianisa. (2014). The Use of Self-Assessment toward Students' Reading Behavior (A Case Study). Journal of English and Education, 2(2), 26–35. https://ejournal.upi.edu/index.php/L-E/article/view/4606

Matsuno, S. (2009). Self-, peer-, and teacher-assessments in Japanese university EFL writing classrooms. Language Testing, 26(1), 075–100. https://doi.org/10.1177/0265532208097337

McAndrews, S. L. (2020). Literacy Assessment and Metacognitive Strategies: A Resource to Inform Instruction, PreK-12. Guilford Press, 508. http://login.ezproxy1.lib.asu.edu/login?url=https://www.proquest.com/books/literacy-assessment-metacognitive-strategies/docview/2488225891/se-2?accountid=4485

Mistar, J. (2011). A Study of The Validity and Reliability of Self-Assessment. TEFLIN Journal, 22(1), 45–58. https://eric.ed.gov/?id=EJ1129982

Newton, P. E., & Shaw, S. D. (2014). Validity in Educational & Psychological Assessment. In Validity in Educational & Psychological Assessment. https://doi.org/10.4135/9781446288856

Nourdad, N., & Asghari, R. (2017). The Effect of Reflective Reading on Reading Comprehension of Iranian EFL Learners. International Journal of Applied Linguistics and English Literature, 6(6). https://doi.org/10.7575/aiac.ijalel.v.6n.6p.267

Permata, I., Selutan, S., Sutapa, G., & Rosnija, E. (2015). The implementation of self-assessment in teaching reading comprehension on narrative text. Jurnal Pendiidkan Dan Pembelajaran, 4(11), 1–11. http://jurnal.untan.ac.id/index.php/jpdpb/article/view/12487

Race, P. (2001). A Briefing Document on Self, Peer and group Assessment. 9.

Ross, J. A. (2006). The Reliability, Validity, and Utility of Self-Assessment. Practical Assessment, Research and Evaluation, 11(10). https://doi.org/https://doi.org/10.7275/9wph-vv65

Shameem, N. (1998). Validating self-reported language proficiency by testing performance in an immigrant community: The Wellington Indo-Fijians. Language Testing, 15(1), 86–108. https://doi.org/10.1177/026553229801500104

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

Vasileiadou, D., & Karadimitriou, K. (2021). Examining the impact of self-assessment with the use of rubrics on primary school students ' performance. International Journal of Educational Research Open, 2–2(January), 100031. https://doi.org/10.1016/j.ijedro.2021.100031

Yan, Z., Brown, G. T. L., Lee, J. C., Qiu, X., Yan, Z., Brown, G. T. L., Lee, J. C., & Qiu, X. (2019). Student self-assessment : why do they do it ? Educational Psychology, 0(0), 1–20. https://doi.org/10.1080/01443410.2019.1672038

Soesilo, T. D. (2018). Penelitian Inferensial dalam Bidang Pendidikan. Salatiga: Satya Wacana University Press.

Boud, D. (1995). *ENHANCING Self Assessment*. 239.

Brown, H. D. (2003). Language Assesment Principles and Classroom Practice. In *Pearson Longman*.

Falchikov, N., & Boud, D. (1989). Student Self-Assessment in Higher Education: A Meta-Analysis. *Review of Educational Research*, *59*(4). https://doi.org/10.3102/00346543059004395

Langan, A. M., Shuker, D. M., Cullen, W. R., Penney, D., Preziosi, R. F., & Wheater, C. P. (2008). Relationships between student characteristics and self-, peer and tutor evaluations of oral presentations. *Assessment and Evaluation in Higher Education*, *33*(2). https://doi.org/10.1080/02602930701292498

Lianisa. (2014). The Use of Self-Assessment toward Students' Reading Behavior (A Case Study). *Journal of English and Education*, *2*(2), 26–35.

Matsuno, S. (2009). Self-, peer-, and teacher-assessments in Japanese university EFL writing classrooms. *Language Testing*, *26*(1), 075–100. https://doi.org/10.1177/0265532208097337

Ross, J. A. (2006). The Reliability, Validity, and Utility of Self-Assessment. *Practical Assessment, Research and Evaluation*, *11*(10).

Yan, Z., Brown, G. T. L., Lee, J. C., Qiu, X., Yan, Z., Brown, G. T. L., Lee, J. C., & Qiu, X. (2019). Student self-assessment : why do they do it ? *Educational Psychology*, *0*(0), 1–20. https://doi.org/10.1080/01443410.2019.1672038