The Differences of Students' Learning Outcomes Using Oral and Written Test on Biology Subject at Senior High School

Marwah,* Mursito Bialangi & Mohammad Jamhari

Pendidikan Sains Program Magister/Pascasarjana – Universitas Tadulako, Palu – Indonesia 94118

Email corresponding author: marwamappakoe@gmail.com

Abstract

The purpose of this study was to determine the learning outcomes of students who used oral tests in biology subjects. Learning outcomes using oral tests and written tests. The research population is students of class X MIA and XI MIA. The results of the calculation of the statistical analysis of the oral and written tests showed that in the oral test, there were 48 students in the poor category, 7 students in the sufficient category, while in the very good category, there were 0 students. In the written test, there were 33 students who received the less category, 14 students in the sufficient category, 5 students in the good category, and 3 students in the very good category. The results of the Wilcoxon test calculations show that the output of the Test Statistics Asymp.Sig (2-tailed) is 0.000. Because the value of 0.000 is smaller than 0.05, it can be concluded "Ho is rejected". This means that there is a difference in student learning outcomes between those given oral tests and those given written tests in biology subjects, with the average written test being higher than the average oral test, both in class X MIA and class XI MIA.

Introduction

Education is basically an interaction between educators and students to achieve educational goals that take place in a certain environment. Such interactions are called educational interactions, namely between educators and students (Nana, 2003).

Educators in this case can be a teacher or a lecturer, according to the level of education being taught. The teacher as one of the parties with an interest in education in schools has several roles, including as a guide, educator, and trainer for students. Therefore, a teacher is expected to have adequate potential in carrying out these various roles (Muttaqqin & Kusaeri, 2017).

The teaching and learning process that is carried out must always be improved so that the results to be achieved are better. One of the efforts to improve the quality of the results process and academic achievement as part of improving the quality of education is through an assessment system. One of the assessment techniques often used by educational institutions is a test (Suharman, 2018).

Assessment of learning outcomes can be done using test and non-test techniques. However, the form of test assessment is most often used compared to non-test measuring instruments (Ruhimat, 2018).

The test is a form of evaluation tool to measure how far the teaching objectives have been achieved. A good test must meet several requirements, namely, it must be efficient, must be standard, have norms, objective, valid, and reliable (Kadir, 2015).

Tests as an assessment tool can be defined as questions given to students to get answers from students in oral form (oral test), in written form (written test), or in the form of action (action test). In general, tests are used to measure and assess student learning outcomes, especially cognitive learning outcomes relating to mastery of learning material (Yussa'diah, 2013).

Learning outcomes are often used as a measure to determine how far a person has mastered the material that has been taught. Learning outcomes consist of two words, namely: results and learning. The definition of results (product) is something that is obtained from doing an activity. Whereas learning is a process of someone trying to obtain a relatively permanent form of behavior change (Purwanta, 2010).

Learning outcomes are students' self-assessments, observable, demonstrable and measurable changes in the abilities or achievements experienced by students as a result of learning experience (Young et al., 2003). So, what is meant by learning outcomes is the ability possessed in the form of knowledge (cognitive), attitude (affective), and skills (psychomotor), all of which are obtained through the teaching and learning process (Yusuf, 2009).

The written test is a test that requires students to write the required answers (Saidah, 2016). While the oral test is in the form of questions given by the teacher orally and students respond to these questions verbally as well, thus fostering a bold attitude to argue. Answers can be in the form of words, phrases, sentences, or paragraphs (Umi, 2018).

Based on the results of interviews with biology teachers at SMA Labshool Palu, information was obtained that in the use of evaluation tests, written tests were more frequent, especially essay tests. Occasionally the teacher uses multiple choice form tests. These written tests are often used because they are easy to prepare and arrange and do not take a long time to complete. However, the tests used in the evaluation are not maximum, because they have shortcomings, namely written tests that easily lead to cheating and falsification of answers because teachers find it difficult to control whether the test is done on their own or someone else's effort.

To overcome this, the teacher needs to try another evaluation test, namely the oral test. Therefore, there is no harm in the form of assessment of learning outcomes using oral tests and written tests (Samad & Arshad, 2013).

Both tests can be used simultaneously because they have several advantages, namely, the written test is easier and faster to do and does not require a lot of time when carrying out the test. In the oral test, students can directly find out the ability of students to express opinions (Purwanto, 2004).

In connection with the learning outcomes measured using oral tests and written tests, the researchers conducted a study entitled the differences of students' learning outcomes using oral and written tests on biology subjects at SMA Labschool Palu.

Materials and Method

This research is quantitative and descriptive. This study aims to explain the existing phenomena by using numbers to describe individual or group characteristics. This research was conducted at SMA Labschool Palu which is located at Jalan Setia Budi, Palu City. The selection of this research location is based on the consideration that the availability of data is needed and in accordance with the problems posed in the study. This research was conducted in the odd semester, from July to August 2020, Academic Year 2019/2020

The populations in this study were students of class X Mia and class X Mia at SMA Labschool Palu, totaling 153 students. The samples in this study were students who could represent the total population of SMA Labschool Palu as much as 110 students consisting of 55 students of class X MIA and 55 students of class XI MIA.

The sampling technique in this study is random sampling, which is how to take random samples by actually providing the same opportunities (Sugiyono, 2008).

The type of data in this study is quantitative data, namely data obtained based on the results of giving oral and written tests on the subject of the ecosystem for class X Mia and excretion for class XI Mia.

The data sources in this study are primary data sources and secondary data sources. Primary data was obtained from oral tests and written tests given to students and the results of interviews with the teacher as support. Meanwhile, secondary data were obtained from literature, articles, journals, books, and internet sites relating to the research conducted.

The data collection technique in this research consists of three stages, namely (1) the preparation stage: formulating the problem to be studied, conducting a literature review, determining the research location, determining the research population and sample, compiling a proposal, which is then presented at the proposal seminar, compiling the instruments that will be used in research, (2) implementation stage: giving oral tests, giving written tests and (3) final stage: activities carried out in the final stage are data tabulation, processing and analyzing sample data in reporting research results. The instrument in this study was in the form of six oral test question and 10 written test questions.

Matondang (2009) stating that validity comes from the word validity which means the extent to which the accuracy and accuracy of a measuring

instrument (test) in performing its measuring function.

Validity concerns the appropriateness of the appraisal tool against the concept being assessed so that it actually assesses what should be assessed. Reliability is one of the main characteristics or characteristics of a good measuring instrument. A test is said to be reliable if it always gives the same results when tested on the same group at different times or occasions (Matondang, 2009). The difficulty level of an item is defined as the proportion or percentage of subjects who answered

certain test items correctly while the difficulty index is a number that indicates the difficulty or ease of an item (Rasyid & Mansur, 2009).

The data analysis steps in this study are as follows:
1) Students are given an oral and written test
2) Determines the answer score based on test category
   a. Oral test answer scoring
      Scoring of answers on the oral test is accelerated in three aspects of the assessment which can be seen in Table 1.

Table 1. Rubric for assessment of oral test answers

No.	Scored aspect	Score
1.	Courage to Answer	
a.	Fluently answers questions	2
b.	Haltingly answered the question	1
c.	Not answering questions	0
2.	Clear Pronunciation of Answers	
a.	The answer is very clear	3
b.	The answer is less clear	2
c.	The answer is not clear	1
3.	Truth of Answer	
a.	In accordance with theoretical studies	3
b.	Approaching theoretical studies	2
c.	Not in accordance with theoretical studies	1
Maximum Score		8 / Problem
Maximum Total		48

b. Scoring of written test answers

The written test in this study used a multiple choice test or commonly called a multiple choice test where in one question there is only one question choice of the correct answer and the other one is wrong. Multiple choice test scores can be seen in Table 2.

Table 2. Rubric for written test answer assessment

Option	Score
Correct	1
Incorrect	0

1. Calculating the total score of students by adding up the scores of each question for each student.
2. Calculate the frequency distribution of the oral and written test scores.
3. Determine the mean difference of the two paired samples using statistical formulas. The formula used is the Wilcoxon test with the help of SPSS.
4. Determine the statistical hypothesis

Results and Discussion

The test results were analyzed to determine the validity, reliability and difficulty level of the questions used. Based on the test results, a total of 6

questions for the oral test and 10 questions for the written test, both class X MIA and class XI MIA are used as standard tests or in other words all questions are used. The results of the oral and written test validity tests can be seen in Tables 3-4.

Table 3. Results of the validity of class X MIA oral test questions on the subject of ecosystems

No. Question	*Pearson Correlation*	Validity
1.	.733 *	Valid
2.	.671 *	Valid
3.	.745 *	Valid
4.	.572 *	Valid
5.	.690 *	Valid
6.	.582 *	Valid

Table 4. Results of the validity of class XI MIA oral test questions on the subject of excretion

No. Question	*Pearson Correlation*	Validity
1.	.743 *	Valid
2.	.733 *	Valid
3.	.781 *	Valid
4.	.777 *	Valid
5.	.642 *	Valid
6.	.717 *	Valid

The test results were analyzed to determine the validity, reliability and difficulty level of the

questions used. Based on the test results, a total of 6 questions for the oral test and 10 questions for the written test, both class X MIA and class XI MIA are used as standard tests or in other words all questions are used. The results of the oral and written test validity tests can be seen in Tables 5-6.

Table 5. Results of the validity of class X MIA oral test questions on the subject of ecosystems

No. Question	*Pearson Correlation*	Validity
1.	.733 *	Valid
2.	.671 *	Valid
3.	.745 *	Valid
4.	.572 *	Valid
5.	.690 *	Valid
6.	.582 *	Valid

Table 6. Results of the validity of class XI MIA oral test questions on the subject of excretion

No. Question	*Pearson Correlation*	Validity
1.	.743 *	Valid
2.	.733 *	Valid
3.	.781 *	Valid
4.	.777 *	Valid
5.	.642 *	Valid
6.	.717 *	Valid

Table 7. Results of the validity of class X MIA written test trial on the subject of ecosystems

No. Question	*Pearson Correlation*	Validity
1.	.624 *	Valid
2.	.557 *	Valid
3.	.560 *	Valid
4.	.625 *	Valid
5.	581 *	Valid
6.	.600 *	Valid
7.	.452 *	Valid
8.	.415 *	Valid
9.	.517 *	Valid
10.	.464 *	Valid

Table 8. Results of the validity of the written test test questions for class XI MIA on the subject of excretion

No. Question	*Pearson Correlation*	Validity
1.	.682 *	Valid
2.	.576 *	Valid
3.	.737 *	Valid
4.	.576 *	Valid
5.	.712 *	Valid
6.	.698 *	Valid
7.	.615 *	Valid
8.	.309 *	Valid
9.	.456 *	Valid
10.	.574 *	Valid

The data in Tables 7-8 shows that each of 6 items for the oral test and 10 items for the written test are declared valid. It is known that if the calculated value is greater than r-table it is said to be valid and if the calculated value is smaller than r-table it is said to be invalid. The calculated value in the table above is the value of Pearson Correlation. While the r-table value with a significance level of 5% and the number of students 55 is 0.266. In the reliability test, a certain limit is used to determine the reliability of an instrument. The limitation of the reliability value according to Priyatno (2010) namely the reliability of less than 0.6 is not good, while 0.7 is acceptable, and above 0.8 is good. The reliability test results of the oral test and written test of class X MIA and XI MIA can be seen in Table 9.

Table 9. The results of reliability test questions on oral and written tests for class X MIA and class XI MIA

Test	Class	*Cronbach's Alpha*
Oral	X	0.725
	XI	0.798
Written	X	0.729
	XI	0.793

The results of the data in Table 9 show that the Cronbach's alpha value in class X MIA and XI MIA in the oral test refers to Priyatno's (2010) the values of 0.725 and 0.798 are above 0.6, so that the question instrument has proven to be reliable and acceptable. The same thing is shown in the written test for class X MIA and XI MIA, the data obtained are 0.729 and 0.793, which means that it is above 0.6 so that the question instrument can be accepted. The results of the difficulty level test on the oral and written tests of class X MIA and XI MIA can be seen in Tables 10 - 11.

Table 10. Results of the difficulty level test questions for class X MIA and XI MIA oral test

Question Number					
1.	2.	3.	4.	5.	6.
Moderate	Moderate	Moderate	Moderate	Moderate	Moderate

Table 11. Results of the difficulty level of the written test questions for Class X MIA and XI MIA

Question Number					
1.	2.	3.	4.	5.	6.
Moderate	Moderate	Moderate	Moderate	Moderate	Moderate

The results of the difficulty level test are used to see the level of difficulty where the data obtained

all meet the criteria as a good test because they are in the medium category with the difficulty index interval from 0.31 to 0.70. Furthermore, the results of statistical analysis tests of oral tests and written tests to see the distribution of frequencies can be seen in Tables 12-13.

Table 12. The result of an evaluation frequency distribution of class X MIA oral test

Score	Category	Frequency	Percentage
80 - 100	Very well	-	-
70 - 79	Good	-	-
60 - 69	Enough	7	12.7%
<60	Less	48	87.2%

Table 13. The result of an evaluation frequency distribution of MIA class XI oral test

Score	Category	Frequency	Percentage
80 - 100	Very well	3	5.4%
70 - 79	Good	5	9.0%
60 - 69	Enough	14	25.4%
<60	Less	33	60%

The data in Tables 14-15 show that the value range 80-100 is included in the very good category with a student frequency of 0 in class X MIA with a percentage of 0 and 3 in class XI MIA with a percentage of 5.4%. The range of values 70 -79 is included in the good category with a student frequency of 0 in class X MIA with a percentage of 0 and 5 in class XI MIA with a percentage of 9.0%. The range of values 60-69 were included in the sufficient category with a frequency of 7 in class X MIA with a percentage of 12.7% and 14 in class XI MIA with a percentage of 25.4%. The value range <60 is included in the low category with the frequency of students being 48 in class X MIA with a percentage of 87.2% and 33 in class XI MIA with a percentage of 60%.

Table 14. The result of the distribution of frequency of assessment of class X MIA written test

Score	Category	Frequency	Percentage
80 - 100	Very well	5	9.0%
70 - 79	Good	18	32.7%
60 - 69	Enough	18	32.7%
<60	Less	14	25.4%

Table 15. Frequency distribution of the written test result assessment for class XI MIA

Score	Category	Frequency	Percentage
80 - 100	Very well	14	25.4%
70 - 79	Good	12	21.8%
60 - 69	Enough	20	36.3%
<60	Less	9	16.3%

The data in Table 14 and Table 15 show that the value range 80-100 is included in the very good category with a frequency of 6 students in class X MIA with a percentage of 10.9% and 14 in class XI MIA with a percentage of 25.4%. The range of values 70 -79 is included in the good category with a student frequency of 18 in class X MIA with a percentage of 32.7% and 12 in class XI MIA with a percentage of 21.8%. The range of values 60 -69 were included in the sufficient category with a frequency of 18 in class X MIA with a percentage of 32.7% and 20 in class XI MIA with a percentage of 36.3%. The value range <60 is included in the low category with a frequency of 13 students in class X MIA with a percentage of 23.6% and 9 in class XI MIA with a percentage of 16.3%. After analyzing the oral test and the written test to see the results of the frequency distribution, Furthermore, a statistical analysis of the differences between oral and written tests was carried out. The results of the known normality test can be seen in Table 16.

Table 16. Class X MIA and XI MIA normality test results

Class	Oral test (Sig)	Written test (Sig)
X MIA1	.143	.022
X MIA2	.849	.016
X MIA3	.364	.006
XI MIA1	.418	.007
XI MIA2	.169	.004
XI MIA3	.093	.046

The normality test data in Table 14 shows that class X MIA and class XI MIA have an abnormal distribution, this is because the significance of the written test is less than 0.05 (<0.05). The results of hypothesis testing using the Wilcoxon test can be seen in Tables 17-18.

Wilcoxon test (sign test) is included in non-parametric statistics. This test uses two interconnected (paired) samples, which aim to measure the significance of the difference between the two paired data groups with an interval scale but with an abnormal distribution. The results of the Wilcoxon test oral test and written test in class X MIA and XI MIA can be seen in Table 17.

Table 17. Wilcoxon test results oral and written test in class X MIA and XI MIA

Class	Oral - Written Test Asymp.Sig	Negative Ranks	Positive Ranks
X MIA1	.000	18a	0b
X MIA2	.000	18a	0b
X MIA3	.000	19a	0b
XI MIA1	.000	18a	0b
XI MIA2	.000	18a	0b
XI MIA3	.000	19a	0b

The data in Table 17 shows that the negative ranks or the difference (negative) between the biology learning outcomes for the oral test and the written test are X MIA1 18, X MIA2 18, X MIA3 19, and XI MIA1 18, XI MIA2 18, XI MIA3 19, this shows there is a decrease (reduction) from written test scores to oral tests. Positive ranks or the difference (positive) between the biology learning outcomes for the Lisa test and the written test is 0. This value of 0 indicates that there is no improvement in biology learning outcomes from the written test to the oral test. The following is the average difference between the oral and written tests can be seen in Table 18.

Table 18. Results of difference in the average oral test and written test in class X MIA and XI MIA

Class	Oral test	Written test
X MIA1	51.00	65.00
X MIA2	43.83	60.56
X MIA3	45.84	61.58
XI MIA1	55.50	66.67
XI MIA2	51.89	63.89
XI MIA3	61.05	67.89

The data in Table 18 shows that the average written test is higher than the average oral test, both class X MIA and class XI MIA. Based on the Test Statistic output, it is known that Asymp. Sig (2-tailed) has a value of 0.000. Because the value of 0.000 is less than 0.05, it can be concluded that "Ho is rejected". This means that there are differences in student learning outcomes between those given oral tests and those given written tests on biology subjects in class X MIA and XI MIA SMA Labschool Palu.

This research was conducted at SMA Labschool Palu with the research subjects being students of class X MIA and students of XI MIA. This research aims to know (1) student learning outcomes using oral tests on biology subjects, (2) student learning outcomes using written tests on biology subjects and (3) knowing whether there is any difference between the oral test results and the written test results. The data analyzed were the learning outcomes of the Labschool Palu SMA class X MIA and XI MIA. Learning outcomes are obtained from two techniques, namely oral tests and written tests. The aspects analyzed in this study include statistical analysis of oral test results, statistical analysis of written test results, and statistical analysis of differences between oral test results and written test results.

The statistical analysis of the results of the oral test was carried out by giving questions orally to each student in turn. The oral test requires mastery of the material in order to answer the questions given. The questions given are 6 questions with the category of assessment using three aspects, namely the courage to answer a score of 0-2, clarity of pronunciation to answer scores of 1-3, and correctness of answers to scores of 1-3. The results of the analysis obtained are seen from the results of the frequency distribution of the assessment on the oral test, where the number of students who get the oral test score in the category <60 is 48 students. Meanwhile, students who get scores in the category> 60 are still lacking, this is due to the lack of mastery of the material and also the personality of the students who are closed. This is in line with Elfiza's (2017) opinion that personality is an internal factor that affects student achievement in learning, especially speaking skills. Extrovert and introverted personality influence how students handle feelings generated during the learning process.

Other factors that influence the lack of oral test scores obtained are the feeling of reluctance and the lack of prior communication between students and researchers who will give oral tests. This is in accordance with the opinion of Arifin (1995), which states that communication is a process of delivering messages or interactions from senders to recipients. Therefore, communication must have feedback between the communicator and the communicant.

Likewise, education requires good communication, so that what is conveyed in terms of lessons, by communicators to communicants can be digested optimally, so that the educational goals to be achieved can be realized.

Statistical analysis of written test results, just like oral tests, written tests also require mastery of the material to be able to answer the questions given. The difference between the two tests is that the written test has a longer time to think so that the opportunity to answer the question correctly is much greater, while the oral test gives less time to think but gives students the opportunity to explain more argumentative answers so that it becomes a consideration for researchers in grading. The questions given are 10 questions with the assessment using true-false options with a score of 0-1.

The analysis results obtained were seen from the results of the assessment frequency distribution on the written test, there were still many students who scored in the <60 category. This is due to conceptual errors where students already understand the concept of the material being tested but cannot express the concept in understanding

multiple choice answers so errors occur in the choice of answers. On the other hand, the majority of students scored a written test in the excellent category. This is inseparable from the understanding of the material that has been given and it is normal for students to choose answers with origin but right on the correct answer choices.

The results of difference between the oral test and the written test can be seen from the data that has been obtained, namely the average score of the oral test results is lower than the written test. The average number of oral tests was X MIA1 51.00, X MIA2 43.83, X MIA3 45.84 and XI MIA1 55.50, XI MIA2 51.89, XI MIA3 61.05. While the average scores of written test results are X MIA1 65.00, X MIA2 60.56, X MIA3 61.58 and XI MIA1 66.67, XI MIA2 63.89, and XI MIA3 67.89.

The results of this study are in line with research conducted by Rokhmawan & Adhy (2012), namely that after undergoing the learning process with the same method, the experimental group I was given a written test, and the experimental group II was given an oral test, the experimental group I had a higher average learning result. Or better than the average value of the learning outcomes of experimental group II. The amount of difference in learning outcomes is 8.50%. Based on the criteria for differences in learning outcomes, it shows that the difference in learning outcomes between the written test and the oral test is very low.

Other research on biology subjects also stated that the average learning outcomes of students who used written tests were higher than oral tests, with an average written test of 74.43 and oral tests with an average of 64.02 (Fitriani, 2013).

The results of this study are not in line with the research of Oktaviyanti & Rosyidah (2019), which states that the results of oral and written tests are in a good category with an average of 79.6 and 78.65. While the results of this study found that the oral test was in the poor category and the written test was in the sufficient category.

The factors that affect the differences in the results of the oral test and the written test (multiple choice), seen from the research that has been conducted, namely in terms of different test forms, meaning that multiple-choice has the potential to get better results than the oral form. This is in line with research by Nita (2014), which suggests that overall, the characteristics possessed by multiple-choice tests make it easier for students to answer the test because they are given alternative answers that can help students collect information relevant to the problem being asked. This brings an advantage because students can get a correct score even if only based on half of the knowledge. In another case with the oral test, the scoring is based on the level of courage to answer, clarity of pronunciation of answers, and correctness of answers. If students only use half of their knowledge to answer the oral test, of course, the score they will get is not optimal.

Another factor that supports the differences in the results of oral and written tests proposed by Widyoko (2012), namely for multiple choice, there is a drawback that there is an effect of test takers' habits on multiple choice form tests on student learning outcomes. Thus, the more familiar a person is with this form of a multiple-choice test, the more likely he or she will score higher.

Another factor that distinguishes the results of oral tests and written tests is communication skills, this factor is in line with the research of Nanang & Rani (2016), which states that written communication, is better than oral communication. This is because students have limitations in speaking or expressing opinions so they are mostly silent or even just smile.

Another factor that distinguishes the results of the oral and written tests is the timing of the questions. It is known that the test administration is carried out remotely via the network (online). So that the time for administering the test is uncertain both morning, afternoon and evening. So it is likely that the written test (Multiple Choice) has a greater chance of answering correctly than the oral test because the test is administered online.

In general, the written test is better than the oral test. According to Purwanto (2004), the oral test has several weaknesses, including 1) if the relationship between the testers and those who are tested is not good, it can interfere with the objectivity of the test results. 2) the emotional state of students is strongly influenced by the personal presence of the educator in their presence. 3) the nervous nature of those who are tested can interfere with the smoothness of the answers given. 4) It takes a long time so it is not economical. 5) The freedom of students in answering questions is less.

Conclusions

Based on the results of research and data analysis, it can be concluded that: (1) there is an

effect of concept understanding between students who take virtual laboratory learning (2) there is a difference in conceptual understanding between students who have high learning interest and students who have low learning interest, but in the hypothesis third (3) there is no interaction between virtual laboratory learning and learning interest on students' conceptual understanding. This can be seen in the statistical test used where for the final test questions the Sig value was obtained <0.05 which means that Ho is rejected. So, there is a significant difference between students who are given virtual laboratory learning and students who are given real laboratory learning.

Acknowledgments

References

Arifin, A. (1995). *Ilmu komunikasi*. Jakarta: PT. Raja Grafindo Persada.

Elfiza, R. (2017). Pengaruh penggunaan game petualangan dan gambar untuk meningkatkan keterampilan berbicara siswa ekstrover. *Jurnal Kiprah, 5*(1), 43-44.

Fitriani. (2013). *Perbandingan hasil belajar biologi siswa SMA Al Islam 3 Surakarta melalui tes tulis dan tes lisan ditinjau dari konsep diri*. Teks Publikasi: FKIP Universitas Muhammadiyah Surakarta.

Kadir, A. (2015). Menyusun dan menganalisis tes hasil belajar. *Jurnal Al-Ta'dib, 8*(2), 70-77.

Matondang, Z. (2009). Keandalan instrumen penelitian. *Jurnal Tabularsa PPS UNIMED, 6*(1), 87-97.

Muttaqqin, Z.M., & Kusaeri. (2017). Pengembangan instrumen penilaian tes tertulis berupa uraian kajian ilmu keislaman berbasis permasalahan dalam sejarah peradaban Islam. *Jurnal Tatsqif, 15*(1), 1-23.

Nana, S.S. (2013). *Dasar psikologi proses pendidikan*. Bandung: Remaja Rosdakarya.

Nanang, S., & Rani, D. (2016). Analisis kemampuan komunikasi matematis siswa dalam pembelajaran memecahkan masalah bangun datar. *Jurnal Pendidikan Matematika, 7*(1), 1-9.

Nita, J. (2014). *Pengaruh jenis tes dan kemampuan kognitif terhadap hasil belajar materi sistem peredaran darah*. Artikel penelitian: Universitas Tanjung Pura Pontianak.

Oktaviyanti, I., & Rosyidah, A.N.K. (2019). Korelasi antara hasil tes lisan dan hasil tes tertulis pada mahasiswa PGSD UNRAM. *Jurnal Ilmu Pendidikan, 2*(1), 9-19.

Priyatno, D. (2010). *Teknik mudah dan cepat untuk melakukan analisis data penelitian dengan SPSS*. Yogyakarta: Gaya Media.

Purwanta. (2010). Penggunaan class based assessment untuk meningkatkan kualitas proses dan hasil belajar siswa kelas IV SD Negeri Samirono Yogyakarta pada mata pelajaran IPS. *Jurnal Penelitian, 13*(2), 147-168.

Purwanto, N. (2004). *Prinsip dan teknik evaluasi pengajaran*. Bandung: PT. Rosdakarya.

Rasyid, H., & Mansur. (2009*). Penilaian hasil belajar*. Bandung: Prima Wacana.

Rokhmawan., & Adhy, D. (2012). Perbedaan hasil belajar tes tulis dan tes lisan pada mata pelajaran konstruksi pondasi dangkal pada siswa kelas X program keahlian teknik gambar bangunan SMK Negeri 5 Semarang tahun ajaran 2008/2009. *Jurnal Scaffoldin, 1*(1), 1-6.

Ruhimat, M. (2018). Kompetensi pembuatan instrumen pengukuran hasil belajar oleh guru IPS di SMP Kota Bandung. *Jurnal Lentera Pendidikan, 21*(2), 176-187.

Saidah, K. (2016). Analisis Kesesuaian Instrumen Penilaian IPS buku guru kelas V tema sejarah peradaban Indonesia dengan kompetensi dan prinsip penilaian IPS Kurikulum 2013 SD. *Jurnal Pendidikan Dasar Nusantara, 2*(1), 21-31.

Samad., & Arshad, A. (2013). Penggunaan portofolio sebagai alat penilaian di kelas L2 Malaysia. *Jurnal Internasional Pendidikan Bahasa Inggris, 1*(1), 94-108.

Sugiyono. (2008). *Metode penelitian pendidikan, pendekatan kuantitatif, kualitatif*. Bandung: Alfabet.

Suharman. (2018). Tes sebagai ukuran prestasi akademik. *Jurnal Pendidikan Agama Islam, 10*(1), 93-94.

Umi, S. (2018). Penjaminan mutu penilaian pendidikan. *Jurnal Evaluasi, 2*(1), 283-284.

Widyoko, E.P. (2012). *Teknik penyusunan instrumen penelitian*. Yogyakarta: Perpustakaan Pelajar.

Young, M. E., Klemz, B. R., & Murphy, J. W. (2003). Enhancing learning outcomes the effects of instructional technology, learning style, instructional methods, and student behavior. *Journal of Marketing Education, 2*(5), 130-13.

Yussa'diah, T. (2013). *Perbandingan hasil belajar siswa antara yang menggunakan tes lisan dan tes uraian pada pokok bahasan persegi panjang dan persegi di kelas VII SMPN 8 Kota Cirebon*. Tesis: IAIN Syekh Nurjati Cirebon.

Yusuf, M.M. (2009). Pengaruh metode pembelajaran dan motivasi terhadap hasil belajar programmable logic controller (PLC) siswa kelas III jurusan ketenagalistrikan SMK Negeri 5 Makassar. *Jurnal Medtek, 1*(2), 2-3.