Proceeding of International Conference on Arts and Humanities:
International Conference on Cultural Studies and Applied Linguistics (ICCSAL)

Analysis of Objective Test Type of Multiple Choice in Final Semester Assessment Problems (PAS) Japanese Language Class VI

Cahyani Khayati[1,*] Miftachul Amri[1] Syamsul Sodiq[1]

[1] *Universitas Negeri Surabaya, Indonesia*
**Corresponding author: cahyani.20012@mhs.unesa.ac.id*

ABSTRACT
Researchers will describe the quality of the Final Assessment (PAS) items for the 2019 and 2020 Odd Semester Academic Year in the Japanese language subject class VI MI Khairu Ummah by looking at the terms of validity, reliability, and level of difficulty. The researcher took a sample of 35 students and 20 multiple-choice items related to PAS items to the 2019 and 2020 academic years. The primary data source was the teacher's objective multiple-choice items type test to know the Odd Semester Final Assessment. The quality of the items in Japanese class VI has as many as 12 items (60%) are valid, and 8 questions (40%) are invalid. After calculating the validity test, the reliability results for the 2019 academic year amounted to 0.7549 where r11 is greater than 0.60. Therefore, it can be concluded that it is a high category. In the reliability test of the Final Semester Assessment test for the 2020 academic year, it was found that the r11 reliability coefficient was 0.5035 where r11 was smaller than 0.60. It can be concluded that test items for the 2020 academic year in the form of multiple-choices that presented 20 items and were followed by 35 students had moderate reliability. The calculation results of the items difficulty level in the Odd Semester for the 2019 Academic Year, it is found that there were 5 items (20%) were classified as easy, 9 items (45%) were moderate, 6 items (30%) were classified as difficult. The 20 items presented in the Odd Semester 2020 Academic Year have 4 items (20%) were easy, 10 items (50%) were moderate, and 6 items (30%) were classified as complex.

Keywords: *Item analysis, Validity, Reliability, Difficulty level.*

1. INTRODUCTION

In improving human resources with knowledge, of course, education is needed. However, in addition to gaining knowledge, education can also gain expertise and values contained in life. According to Nata [4], education is a vital activity in an effort to increase human resources, which is carried out through the transfer of knowledge, expertise and values contained in life as a provision for students to reach maturity and personal maturity. A program involves a number of components that work together in a process to achieve the programmed goals. To find out whether the implementation of the program can achieve its objectives effectively and efficiently, it is necessary to conduct an evaluation.

Evaluation is a process that determines the conditions under which a goal has been achieved. Under Law - Law of the Republic of Indonesia Number 20 the Year 2003 on National Education System, states that the evaluation of education is control activities, and the determination of the quality of education to the various components of education at every track, level, and type of education as a form of education. According to Eka [1], this evaluation is carried out in the context of controlling the quality of education nationally as a form of accountability for the implementation of education to interested parties, including students, institutions, and educational programs on formal and non-formal channels for all levels of units and types of education.

Based on the previous presentation. So it is clear that learning evaluation is essential and must be implemented. Likewise, with the evaluation of learning outcomes, a teacher can determine the level of students' understanding of the material taught by evaluating student learning outcomes. In addition, teachers can also use the evaluation results as a reference for improving the learning process in the future. Silverius [7] argues

that the evaluation results from the student's perspective provide information about the extent to which he has mastered the subject matter presented by the teacher. That way, students can determine which attitude should be done after knowing the results of the learning they have gone through. As for the evaluation, there are two evaluation tools, namely tests and non-tests [8]. The test is an educational evaluation program, one of the assessment tools used to assess learning outcomes given by teachers to students. As an evaluation tool in education, the test has a significant role in measuring student achievement. A teacher can use two kinds of tests to evaluate teaching and learning results, namely the *standardized test* and the *teacher-made test*.

The tests used in schools are mostly made by the teachers themselves (*teacher-made tests*). The test made by the teacher certainly has its own advantages and disadvantages. Teacher-made tests are tailored based on the content and goals specific to the class or school in which the teacher teaches. According to Purwanto [6] most of the items used have not been analyzed for each test instrument to have low or moderate reliability. Therefore, it is necessary to know that a good test instrument must meet the requirements of a good test so that it can be said to be a high-quality test instrument. In classical test theory, many item characteristics to be tested are the level of difficulty, the power difference, and the effectiveness of detractors. However, it is the instrument used must also be valid, reliable, objective and practical. Every good item has a moderate difficulty level, the power difference is high, and posing to function effectively.

Based on the form of the questions, the tests are grouped into: a) Objective tests, namely written tests that require students to choose the answers that have been provided or provide limited short answers. Ben tuk-forms of objective tests are (1) True-False Test, according to Metler [2] test point because consisting of a statement accompanied by alternative answers are answers or statements are true and false. (2) Matching Test consists of two parallel columns, one column consisting of information or statements, while the other column consists of answers to statements in the other column. (3) Multiple Choice Test is a test that consists of a statement or notification about an incomplete understanding. Moreover, to complete it must choose one of several possible answers that have been provided. b) Subjective Tests (*Essays*) Subjective tests are items that contain questions or assignments whose answers or work on the questions must be done by expressing the thoughts of the test takers. The characteristics of the question are preceded by words such as; describe, explain, compare, why, how, conclude, etc.

This form of multiple choice objective test into a test that is often used, both at the time of the summative and formative tests. In the practice of teaching summative tests, according to Zaenal [10], it is known as the Final Semester Examination (UAS) or Semester Final Assessment (PAS) depending on the unit of time used to complete the material. The benefits of this summative test are to determine grades, determine the increase in the child's grade, and fill in student learning progress notes that are useful for parents, guidance, and counseling at school, and valuable for other parties if the student will move to another school.

Madrasah Ibtidaiyah Khairu Ummah is an Islamic Aqeedah-based educational institution to create a generation of leaders who build noble civilizations. In learning MI Khairu Ummah not only infuse for students to deepen the Islamic Aqeedah, but students are given pieces of language skills languages Arabic, the language of English and Japanese. This is what makes Japanese a foreign language and is first learned by students of *madrasah ibtidaiyah* or elementary schools. In addition, learning Japanese is included in the category of language with a famous culture for its discipline. Japanese language in MI Khairu Ummah is not like learning grammar which is usually only taught to grade VI and grade V students. Thus, to know the Japanese language learning students of class V and class VI MI Khairu Ummah each semester will perform tests summative or Final Assessment (PAS) by using teacher-made questions Japanese (*teacher-made test*) so that each school year has a different problem in testing students. A teacher should know whether the question has a good quality question or not. However, most of the artificial matter tested on the student-teacher is not all about the analysis done, which means it has not been known level of validity, reliability, and difficulty level. Of course, this will be ungood if a teacher does not carry out the analysis.

Previous research by Mutiah [5] aims to determine the physics learning outcomes of class XI IPA students assessed using objective and subjective tests. In addition, to find out the differences in physics learning outcomes of students in class XI IPA SMA Negeri 1 Sidenreng Rappang, which were assessed using objective tests and subjective tests. This research is an experimental study using a post-test-only control groups design. Furthermore, previous research by Yuniar et al. [3] focuses on testing objective questions in the subject of Social Sciences (IPS) class V SD Negeri 7 Ciamis. The reason for doing this research, because there are still many questions made by teachers who do not meet the criteria for making good questions. The research method used is a descriptive research method with a qualitative approach.

From the two previous research, there are differences and similarities to the research to be carried out. The research equation is the use of objective tests as a tool to measure learning outcomes. In contrast, the

difference in research consists of quantitative analysis that leads to quantitative descriptive analysis methods.

The researcher will describe the quality of the Final Assessment (PAS) items for the odd semester of the 2019 academic year and the odd semester PAS of the 2020 academic year

in the Japanese class VI subject of MI Khairu Ummah. By looking at the items' validity, reliability, and level of difficulty, the researchers took a sample of 35 students and 20 *multiple choice questions* from each PAS item for the 2019 academic year and 2020 academic year.

2. METHODS

The research uses a quantitative descriptive analysis method in which the data is calculated using statistics on a *laptop* using the SPSS and ANATES application version 4.09. The following statistical analysis will be used in this study: Analysis of the validity and reliability using a *laptop* program was carried out using the SPSS program. The steps in the sequence using the SPSS program are as follows:

a. Open Programs. To open the SPSS menu, program collection is done by pressing the start button in the bottom left corner, selecting All programs, and selecting SPSS for Windows.

b. Input the data. After the data has been entered, we save the data in the file we want, then press file, select save as, in the file box fill in our name.

c. Processing data. To analyze the data is done with the following steps:

   1) Open the file,

   2) Click the Analyze menu in the toolbar, select the Scale sub-menu, then Reliability Analysis.

   3) It will appear on the Reliability Analysis dialog display screen.

   4) In the Model section, leave the selection on Alpha. Unless the data is discrete, namely only 1 and 0, we select Split-half in the Model section. Ignore the List Item Models selection box.

   5) Click the Statistics button until the Reliability Analysis: Statistics dialog appears on the screen.

   6) Ignore the others and press the Continue button to return to the previous dialog box.

   7) Press OK to process data.

   8) Wait until Output appears.

After knowing the validity and reliability, the next step is to analyze the level of difficulty of the questions. According to Wahyuni et al. [9], a good test difficulty level is a question that is neither too easy nor too difficult. Problems that are too easy do not stimulate students to increase their efforts to solve them. On the other hand, too difficult questions will cause students to become discouraged and not have the enthusiasm to try because they feel unable or out of reach.

The amount of difficulty index between 0,00 up to 1.0. This difficulty index shows the level of difficulty of the questions. A question with a difficulty index of 0.0 indicates that the item is too difficult. On the other hand, an index of 1.0 indicates that the question is too easy. To determine the level of difficulty of the objective test can be taken with the following formula:

$$TK = \frac{\sum JB}{TS} x\ 100\%$$

Information:

TK : Difficulty of an item.

JB : The number of students who answered the question correctly.

TS : number of samples/ Total number of students taking the test [8] (Suharismi, 2020:210).

The results of these calculations will be compared with the classification of the difficulty index as follows:

- Questions with TK 1.00 – 0.30 are complicated questions.
- Questions with TK 1.00 – 0.30 are medium questions.
- Questions with TK 1.00 – 0.30 are easy questions.

After analyzing the data by analyzing the validity, reliability, and difficulty level, it will describe the results of "Analysis of Objective Test Type of Multiple Choice in Final Semester Assessment Problems (PAS) Japanese Language Class VI MI Khairu Ummah".

3. RESULTS AND DISCUSSION

A Japanese language subject teacher compiled the Final Semester Assessment Test (PAS) for Japanese language subjects for Class VI Khairu Ummah for the odd semester of the 2019 academic year and 2020 academic year. Each PAS question for the odd semester 2019 academic year and the odd semester 2020 academic year has 20 questions in the form of multiple choice. One number of alternative answers that question has an answer a, b, c, and d provided that if test able to

answer the item correctly then given a score of 1 and if one is given a score of 0. Subsequently, the following item analysis test and detailed answers students test:

3.1. Item Validity and Reliability

Test the validity and reliability using SPSS and ANATES applications. It is used to determine the number of valid and invalid item items (invalid) and the number of reliable or unreliable questions. Here is an analysis of the validity and reliability of the *multiple choice* Final Assessment Odd Semester 2019 and Final Assessment Odd Semester 2020.

Table 1. Validity test results using SPSS

Information	Interpretation of validity	Total of Questions	Question Item Number
Questions for the Odd Semester of the 2019 Academic Year	Valid	12	1, 2, 3, 4, 5, 11, 12, 13, 15, 16, 17, 18
	Invalid	8	6, 7, 8, 9, 10, 14, 19, 20
Questions for the Odd Semester of the 2020 Academic Year	Valid	12	1, 2, 3, 4, 5, 6, 8, 9, 11, 15, 18, 19
	Invalid	8	7, 10, 12, 13, 14, 16, 17, 20

The *multiple choice* test items for the odd semester assessment of Japanese subjects for Class VI MI Khairu Ummah in the odd semesters for the 2019 academic year and the 2020 academic year have 12 questions (60%) are valid and 8 questions (40%) these items are invalid. Set e was counting validity, and then the next step is to count the re test abilitas using ANATES . Based on the ANATES calculation , namely in the test reliability test where the reliability coefficient r11 is 0.7549 where r11 is greater than 0.60, it can be concluded that the Final Semester Assessment test for the 2019 academic year is in the form of *multiple choice* which presents 20 questions and is followed by 35 students. Therefore, Japanese class VI subjects of MI Khairu Ummah in odd semesters have good/high reliability. In the reliability test, test Rate Final 2020 academic year form of *multiple choice* that presents 20 items and was followed by 35 students, subjects of Japanese Class VI MI Khairu Ummah semester, in which the reliability coefficient r11 by 0 , 5035 where r11 is smaller than 0, 60, it can be concluded that the Odd Semester Final Assessment test items for the 2020 academic year are in the form of *multiple choice* which presents 20 questions and is followed by 35 students with moderate reliability.

The results of the validity and reliability of multiple-choice objective test items that aim to measure the success of the teaching and learning process of teachers and students in learning Japanese can be used as an effective form of test. There are 12 valid questions so that they become a teacher's reference to provide a special teaching model related to the material contained in items 1, 2, 3, 4, 5, 11, 12, 13, 15, 16, 17, 18 in the odd semester of the following teaching period. And on the material contained in items 1, 2, 3, 4, 5, 6, 8, 9, 11, 15, 18, 19 in the even semester of the following teaching period.

3.2. Item Difficulty Level

The analysis results of the item difficulty index number calculation can be seen in the ANATES application. Therefore, information was obtained about the level of difficulty of the items in the form of *multiple-choice* exam results for the odd semester Japanese language subjects for Class VI MI Khairu Ummah in the odd semester for the 2019 academic year and the 2020 academic year in the following table:

Table 2. Difficulty level results using ANATES

Information	Difficulty Level	Total of Questions	Percentage
Questions for the Odd Semester of the 2019 Academic Year	Easy	5	25%
	Medium	9	45%
	Difficult	6	30%
Questions for the Odd Semester of the 2020 Academic Year	Easy	4	20%
	Medium	10	50%
	Difficult	6	30%

It can be seen that the 20 questions presented in the Odd Semester items for the 2019 academic year there are 5 questions (20%) are classified as easy, 9 items (45%) are classified as moderate, 6 items (30%) is difficult. On the other hand, the 20 questions presented in the Odd Semester questions for the 2020 academic year have 4 questions (20%) are classified as easy, 10 questions (50%) are classified as moderate, and 6 questions (30%) are classified as difficult.

4. CONCLUSION

After analyzing the data that has been collected, then an analysis of the validity, reliability, and level of difficulty is carried out. This is used to determine the quality of the Assessment Final Odd form of *multiple choice* subjects Japanese language classes VI MI Khairu Ummah academic year 2019 and 2020 by the number of about 20 items. It can be concluded that the quality items in Assessment Final Odd subjects Japanese language class VI have as many as 12 items (60%) is valid and 8 items (40%) is invalid. After calculating the validity test, the reliability of the Odd Semester Final Assessment test for the 2019 academic year is 0.7549 where r11 is greater than 0.60, it can be concluded that it is included in the good/high category. In the reliability test of the Final Semester Assessment test, for the 2020 academic year where the reliability coefficient r11 is 0.5035 where r11 is smaller than 0.60, it can be concluded that the odd semester Final Assessment test items for the 2020 academic year are in the form of *multiple choice* which presents 20 questions and is followed by 35 students have moderate reliability. The calculation results of the difficulty level in Odd

Semester items for the 2019 academic year there are 5 questions (20%) are classified as easy, 9 items (45%) are classified as moderate, 6 items (30%) are classified as difficult. The 20 questions presented in Odd Semester questions for the 2020 academic year have 4 questions (20%) are classified as easy, 10 questions (50%) are classified as moderate, and 6 questions (30%) are classified as difficult.

REFERENCES

[1] E. Prihatin, *Educational Administration Theory*. Bandung: Alfabeta, 2011.

[2] C. A. Metler. *Classroom Assessment a Practical Guide for Educators*, 2003.

[3] M. Yuniar, C. Rakhmat, A. Saepulrohman, "Analysis of HOTS (High Order Thinking Skills) in objective tests in Social Science (IPS) class V SD Negeri 7 Ciamis," *PEDADIDAKTIKA: Jurnal Ilmiah Pendidikan Guru Sekolah Dasar*, vol 2, no 2, 2015.

[4] A. Nata, *History of Islamic Education*, Jakarta : Kencana, 2011.

[5] N. Mutiah,, "Influence of objective tests on physics learning outcomes for class XI IPA students of class XI IPA SMA Negeri 1 Sidenreng Rappang," 2020.

[6] M. N. Purwanto, *Principles and Techniques of Teaching Evaluation*, Bandung: PT. Remaja Rosdakarya, 2010.

[7] S. Suke, *Evaluation of Learning Outcomes and Feedback*, Jakarta: Gramedia, 1991.

[8] S. Arikunto, *Basics of Educational Evaluation,* Yogyakarta: Bumi Aksara, 2016.

[9] W. Sri, and A. S. Ibrahim, *Language Learning Assessment*, Bandung: PT. Refika Aditama, 2012.

[10] Z. Arifin, *Evaluation of Learning*, Bandung: PT. Remaja Rosdakarya, 2016.