Analysis of Multiple Choice Questions for Japanese Language Final-Term Exam at Senior High School

Muhammad Muhammad[(✉)], Syamsul Sodiq, and Miftachul Amri

Universitas Negeri Surabaya, Surabaya, Indonesia
muhammadmuhammad.21006@unesa.ac.id

Abstract. The teacher must undertake item analysis to improve the quality of the resultant questions and identify which queries are acceptable, unacceptable, and incorrect. Research questions in detail indicate how they were constructed and how they can be improved in the future. The Japanese language final-term exam questions for class 12 language 1 SMAN 2 Malang have never been validated for validity and reliability. Class 12 language 1 SMAN 2 Malang students who take the Japanese final-term exam are the subjects of this study. This research focuses on the quality of the material, reliability, and questions. To ensure the validity and reliability of the results, the Japanese language final-term exam scores of 12th-grade language 1 students were compiled using a quantitative descriptive technique. Eleven questions were genuine, whereas nine items were determined to be ineffectual. Compared to Siregar's reliability standard of > 0.60, these items are regarded as having low reliability with a value of 0.5.

Keywords: Question points analysis · Final-term exam · Japanese · SPSS

1 Introduction

Evaluation is a process that is carried out to measure and assess. Evaluation can also mean collecting information in the form of appropriate data with planned goals, so evaluation plays a role as giver-related information with student learning outcomes. The results of measurements and assessments also become data sources of obtained information to show whether destination classroom teaching is achieved or not. The results obtained can be used as teacher feedback and also to determine the effectiveness of the learning. The feedback gained is used to improve or learn the perfect program [1].

Evaluation is an activity to get the facts with use steps taken critical and thorough. The evaluation is carried out since problems need precise answers, such as why the capital is jammed?, and why student achievement in class is low? [2]. Assessment is also a systematic activity to obtain information about student learning outcomes which will later be used as consideration for making decisions about grade promotion or graduation.

There are several benefits of evaluating learning processes and outcomes: (1) gaining an understanding of the implementation and learning outcomes that have been ongoing

or carried out by teachers, (2) making decisions about the implementation and learning outcomes, and (3) improve the quality of learning processes and outcomes for quality improvement.

The evaluation is used to know how much good students understand the given material. Overviews tell teachers and students how well they are doing so they can keep learning. In addition, assessment can help teachers know how well students are progressing and how well their learning objectives align with predetermined competency standards and essential competencies. Exams can help teachers know what students are good at and how to assist them in improving (enrichment). Likewise, student weaknesses can be used as a method to assist or provide advice (remedial). Then, the assessment can select students based on type and education.

Based on the type of evaluation, the test is divided into two, namely the written test and the oral test. The test is used as a tool for students to measure their progress. Students can use the test to measure success and improve their education outcomes. Second, the test can be used to measure the curriculum's success. So far, the planned curriculum has been followed. [3].

According to Zainul and Nasoetion [4], evaluation is a decision-making process based on information obtained from learning measurements using test and non-test tools. Sudjana [5] stated that the evaluation tool is valid for certain purposes but not automatic for other purposes. The validity of the evaluation depends on the situation and the evaluation purpose. The purpose of the study's validity is to test the validity of the evaluation instrument. From this, it can be concluded that validity is size-level item validity.

According to Arifin [1], reliability is the level or degree of device consistency. Besides, according to Amirono & Daryanto [6], equipment reliability is the condition of an instrument that provides reliable or consistent measurement results. A reliable tool produces the same results for different people at different times. Equipment reliability is related to equipment effectiveness. Most valid devices are guaranteed to be reliable, but devices that are not always reliable are declared valid. Reliability test equipment is used to determine the reliability of the tool.

All schools hold exams at the end of each semester. The purpose of conducting the final-term exam is to measure the achievement of students' abilities at the end of the semester. Foreign language students, especially Japanese language, face many challenges as their skills develop, including lack of language mastery, inability to write Kanji characters, and various other difficulties.

By knowing the results of the final-term exam scores, the teacher can add further discussion of the topic to the learning material. Test in the form of a multi-disciplinary observatory examination with five counter decisions. The questions tested include subjects and lesson objectives. Calculation of the validity, reliability, and difficulty level of the questions is done with a computer program called SPSS. After conducting the inspection, the feasibility evaluation carried out can be explained.

2 Methods

This research was conducted quantitatively on grade 12 language 1 students whose research parameters included the validity and reliability of the items. The data calculation

is carried out using the SPSS version 25 application. The steps for using the SPSS version 25 program are as follows:

a. Open the SPSS version 25 program by double-clicking it.
b. Entering data in the form of the number of questions, students, and the scores obtained during the test.
c. The results of data processing in the form of reliability and items (validity) will be seen.

2.1 Validity

According to Arikunto [7], the formula used in the validity test is as follows:

$$r_{xy} = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{\{N\Sigma X^2 - (\Sigma X)^2\}\{N\Sigma Y^2 - (\Sigma Y)^2\}}}$$

Description:
$r_{xy}r_{xy}$= Correlation coefficient between variables X and Y
X = Item score question
Y = Total score
N = Total student
According to Sugiyono [8], the validity of the normal weight is 0.5. Therefore, the results must be compared with the standard score of the correct question if the correlation score obtained is lower than the standard score (Significant).

2.2 Reliability

The reliability of an instrument can be measured using a reliability calculation. The SPSS application is used to test 20 multiple choice questions in the Japanese language final-term exam assessment. This study used the Cronbach Alpha formula to perform the dependency test. Questionnaires or descriptions in the form of questions can use the Alpha formula. According to Arikunto [7], to determine the reliability of the instrument is if the score is neither 1 nor 0. Below is Cronbach's Alpha formula.

$$r_{11} = \left[\frac{k}{k-1}\right]\left[1 - \frac{\Sigma\sigma_b^2}{\sigma_t^2}\right]$$

Description:
$r_{11}r_{11}$ = Reliability
k = Number of questions $\Sigma\sigma_b^2 \Sigma\sigma_b^2$= Number of item variants $\sigma_t^2 \sigma_t^2$= Total variance
The reliability is sufficient if the alpha value is more than 0.7. Alpha > 0.80 indicates that all items are reliable, and all tests consistently show strong reliability. Alternatively, the following interpretation has been offered: Alpha > 0.90 means the system is completely reliable. Between 0.70% and 0.90%, alpha is considered reliable. Moderate reliability occurs if the alpha is between 0.50 and 0.70. Low Alpha indicates a low level of reliability. One or more items may not be reliable if the Alpha is low.

3   Results and Discussion

The analysis of this study was carried out in accordance with the purpose of the study, namely examining the quality of the final-term exam questions. The analysis consists of three parts: validity, reliability, and difficulty level of the questions. There are 20 multiple-choice final exam questions, and the number of students in the class is 31.

3.1   Validity Question Points

The validity test conducted using the SPSS application aims to determine the number of valid and invalid questions. The following table depicts the use of the SPSS application on the Japanese language final-term exam questions.

Whether the question is valid or not is determined by the correlation coefficient of significance as the benchmark. Table 2 shows the results obtained from the validity of the final-term exam items conducted on 11 language 1 students. The significance limit in this validity test is > 0.5, and if it is more than 0.5, the item is declared invalid.

The results in the table show that only 11 questions are significant or valid, while the rest (9 questions) have low or invalid validity. Of the 20 questions, 3 questions have a minus correlation score or less than 0.5, so the question is automatically invalid. Valid questions show the relationship quality between measurements that are in accordance with the test questions.

3.2   Reliability Question Points

Count reliability is used to measure the consistency of the measuring instrument. Of the 20 multiple choice questions in the Japanese final-term exam, which were analyzed using the SPSS program, a reliability index of 0.05 was obtained. According to Siregar [9], the minimum coefficient for these questions to be reliable is 0.60. Based on the benchmark, if $r11 \geq 0.60$, then the questions tested can be interpreted as questions that have high reliability. Meanwhile, for $r11 \leq 0.60$, the questions tested have low reliability. So, the reliability in this study is included in the category of questions with low reliability since it has a reliability index of 0.05 (Table 1).

Table 1. The results of the validity question points

Question Number	Correlation Number
1	(0.434341)
2	(0.515732)
3	(0.683222)
4	(0.731387)

<div align="right">(<i>continued</i>)</div>

Table 1. (*continued*)

Question Number	Correlation Number
5	(0.649565)
6	(0.465688)
7	(0.260598)
8	(0.365592)
9	(0.465688)
10	(0.7633)
11	(-0.01303)
12	(0.650371)
13	(-0.12837)
14	(0.220468)
15	(-0.21322)
16	(0.291661)
17	(0.342349)
18	(0.034207)
19	0.038849
20	0.155705

Table 2. The reliability statistics result

Reliability Statistics	
Cronbach's	N of items
0.511	20

4   Conclusion

Analyzing question items is a process that must be taken and carried out by a teacher to determine the quality of the questions given. This analysis activity can assist the teacher in determining whether the question is worthy enough to be tested or not. Validity identifies which questions are appropriate for evaluation by referring to the competencies that have been prepared using the SPSS formula. Of the 20 question items, 11 valid questions and 9 invalid questions were found.

Reliability refers to how far the results of a measurement can be trusted. A measurement result can only be trusted if relatively the same results are obtained in several times after carrying out measurements to group the same subject. The reliability instrument of the coefficient of determination for the question item scores uses Cronbach's Alpha formula, where 20 question items have low reliability.

The reliability of the interpretation coefficient is a relative interpretation, which means there is no absolute limit stating how many minimum coefficients must be achieved before a measurement can be considered reliable. However, it does offer information about the correlation between the observed variance scores and the actual variance scores of any given group. There are several conditions that must be met to do so, such as the validity and reliability of the questions on the mid-term or the final-term tests.

Authors' Contributions.   Muhammad Muhammad was the mind behind the study's conception and planning. Syamsul Sodiq was the one who carried out the research. Miftachul Amri made a contribution in figuring out how to interpret the results. Every author contributed insightful criticism and played a part in conducting the research and drafting the the manuscript.

References

1. Z. Arifin, *Evaluasi Pembelajaran*. Bandung: Remaja Rosdakarya, 2013.
2. M. Mulyadi, "Penelitian kuantitatif dan kualitatif serta pemikiran dasar menggabungkannya," Jurnal Studi Komunikasi dan Media, vol. 15, no. 1, pp. 127-138, 2011.
3. S. Anas, *Pengantar Evaluasi Pendidikan*. Jakarta: Raja Grafindo, 2012
4. A. Zainul dan N. Nasoetion, *Penilaian Hasil Belajar*. Pusat Antar Universitas, Direktorat Jenderal Pendidikan Tinggi: Departemen Pendidikan Dan kebudayaan, 1997.
5. N. Sudjana, *Penilaian Hasil Proses Belajar Mengajar*. Bandung: Remaja Rosdakarya, 2014.
6. A. Amirono dan D. Daryanto, *Evaluasi dan Penilaian Pembelajaran Kurikulum 2013*. Yogyakarta: Gava Media, 2016.
7. S. Arikunto, *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara, 2013.
8. D. Sugiyono, *Metode Penelitian Kuantitatif, Kualitatif dan R&D*. Bandung: Alfabeta, 2018.
9. S. Siregar, *Metode Penelitian Kuantitatif: Dilengkapi dengan Perbandingan Perhitungan Manual dan SPSS*. Jakarta: Kencana, 2017.