KnE Social Sciences

Knowledge E
Engaging minds

Conference Paper

Validity Test and Reliability of Indonesian Language Multiple Choice in Final Term Examination

R. Setiawaty, T.B. Sulistyorini, Margono, and L.E Rahmawati

Departement of Indonesian Education, Faculty of Teacher Training and Education, Muhammadiyah University of Surakarta, Surakarta, Central Java, Indonesia

Corresponding Author:
R. Setiawaty
A310140096@student.ums.ac.id

OPEN ACCESS

Abstract

The purpose of this research is to examine the validity and reliability of the multiple-choice items of the Even Semester Final Examination of Bahasa Indonesia subject of class 7A Junior High School 2 of Surakarta (SMPN 2 Surakarta) in the academic year 2015/2016. A qualitative research method is used by involving the primary data from the school's documentation, which in this case is SMP N 2 Surakarta. The respondents of this research comprised 26 students. The discussion of this research shows that based on the five percents of calculation table 0.388, the validity of multiple-choice items of-of significance of 5% with the correlation table of 0.388, the validity of the multiple-choice items of Even Semester Final Examination of Bahasa Indonesia subject that consists of 45 numbers, there are seven valid items, and 38 invalid items. The result of the reliability index calculation of the Even Semester Final Examination of Bahasa Indonesia subject of class 7A, SMP N 2 Surakarta is $0.3657 \leq 0.6$. This means that this instrument cannot be implemented to carry out measurements, so there must be an evaluation to improve its quality. Therefore, r_{table} obtained 0.388 with $r_{11} < r_{table}$, and it can be said that the result of measuring the Learning Result Test (THB) does not have a significant correlation and cannot be said to be reliable.

Keywords: evaluation, multiple choice, reliability, test, validity

1. Introduction

Learning is an activity that is implemented by students, teachers, and educators with a view to achieving successful learning. Learning activities apply the test and nontest techniques. Tests measure the outcomes of learners in the expectation of providing validated truth. The results obtained through the test technique will be used to make final decisions for teachers in providing both student assessment and evaluation and assessment of the learning process. In order to obtain the validated test information,

How to cite this article: R. Setiawaty, T.B. Sulistyorini, Margono, and L.E Rahmawati, (2018), "Validity Test and Reliability of Indonesian Language Multiple Choice in Final Term Examination" in *The 1st International Seminar on Language, Literature and Education*, KnE Social Sciences, pages 43–50. DOI 10.18502/kss.v3i9.2609

Page 43

the test itself should be validated too. The test equipment must have good-quality validity and reliability, in order to measure the learners' outcomes.

As a measurement of learning achievement, an evaluation tool (test) must have a good level of validated (validity) information about the learners, which is obtained from the implementation of a validated measurement test [1]. Validity is the evidence that has been collected from a learner's evaluation that can be used as a reference by the teacher to predict the test result. Validity is the interpretation of test scores, and not the test itself. Validity is concerned with the prediction of the test result scores. A high validity shows that the measurement result reflects the measured facts correctly [2].

Based on the type of data and analytical work, that validity can be divided into two categories: rational analysis and empirical data analysis [2]. According to rational analysis or logical considerations, validity can be divided into the validity of the content and the validity of the concept. Based on the empirical data, which is called empirical validity, the validity is divided into two kinds of validity in line and the validity of the forecast. The validity of the test can be influenced by several factors, including the factor of the test instrument itself, the implementation of measurement and scoring, and the factor of the answers of test participants. These things can affect the teacher's judgment by using the instrument of validity.

Validity and reliability are closely related, as stated by Nurgiyantoro; if validity focuses on an appropriate prediction of the score of the test result, reliability focuses on the constant score of the test result [2]. These two things are often discussed in correlation with the test. The reliability of a test shows how consistent the test result of a current measurement is compared to other measurements. A high or low level of reliability will influence the validity. A high reliability will enable the accomplishment of validity.

Data analysis is used by uses test validity and reliability of the instrument [3]. A valid research instrument is an instrument that is capable of measuring what should be measured. The technique used in the validity test is based on the product moment correlation formula of Karl Pearson. The research result can be said to be valid if the correlation coefficient is higher than r_{table} with a significance level of 5%.

The main problem that will be discussed in this research regards the instrument validity of the Even Semester Final Examination of Bahasa Indonesia subject 7th grade the SMP N 2 Surakarta academic year 2015/2016. In addition, it also finds out how the reliability of the test items of And to find out the reliability of the Even Semester Final

Examination of Bahasa Indonesia subject 7th grade SMP N 2 Surakarta academic year 2015/2016.

The purpose of the research on the test validity instrument is to reveal the validity of the Even Semester Final Examination of Bahasa Indonesia subject of 7th grade SMP N 2 Surakarta academic year 2015/2016. Besides that, it also finds out how the reliability of the test items of And to find out how the reliability of Even Semester Final Examination of Bahasa Indonesia subject of 7th grade SMP N 2 Surakarta academic year 2015/2016.

2. Methods

Based on the problem proposed in this research, the test validity and reliability of the test items of the Even Semester Final Examination of Bahasa Indonesia subject of class VII the SMP N 2 Surakarta academic year 2015/2016, the researcher will use both qualitative research. This research is applied by analyzing the students' answers regarding the test items of the final examination of Bahasa Indonesia subject. The data used in this research are primary data in the form of students' answer sheets, which are obtained directly. The data collection applies the documentation technique. The subject of this research is the 26 student respondents of the 7A class of SMP N 2 Surakarta, and the object of this research is the students' answer sheets of the 7A class of the Even Semester Final Examination of Bahasa Indonesia subject.

The instrument evaluation of this research comprises: (1) the test instrument of the Even Semester Final Examination of Bahasa Indonesia subject of class 7A Junior High School 2 of Surakarta (SMP N 2 Surakarta); (2) the key answers of multiple-choice test items and the answer sheets of class 7A students who attended the Final Examination of Bahasa Indonesia subject. The process of analyzing the test items is carried out by using the computer calculation program of Microsoft Excel. The calculation formula for test item analysis is based on the validity of test items of the Learning Result Test by using rpbis (biserial point correlation).

The interpretation of the instrument validity is done by means of valid and invalid criteria limits as shown in the correlation table. If rpbi> r_{table}, it is said to be "valid," and if rpbi <r_{table}, it is said to be "invalid" [4]. Furthermore, to measure reliability testing using the formula K-R 20.

$$r_{11} = \left(\frac{n}{n-1} \right) \left(\frac{S2 - \sum pq}{S2} \right)$$

r_{11}: Reliability of the overall assessment instrument

p: The proportion of the subjects who answered the item correctly

q: The proportion of the subjects who answered the item wrongly (q = 1-p)

Σpq: The number of multiples p and q

n: Number of test takers

S^2: Standard deviation

The criteria for determining the reliably are: (a) if r arithmetic > r table then the two scores of the results of the TEST Learning Result Test (THB) are significantly correlated. The correlation significance indicates the existence of consistency, so the THB can be said to be reliable; (b) if the reliability index $\geq$ 0.6 then the instrument has a good value of usefulness. This means that the instrument can be used to perform measurements [5].

3. Results

The Even Semester Final Examination of Bahasa Indonesia subject of class 7th grade SMP N 2 Surakarta academic year 2015/2016 is one of the test forms conducted by the school that was created by the MGMP Surakarta. The quality of the test instrument can be determined by the degree of difficulty, differentiation, the problem of deception, and the validity and reliability of the questions.

The research conducted in SMP 2 Surakarta shows that the test instrument of the School Final Examination of Bahasa Indonesia subject is in the form of a multiple-choice test item and essay. There are 45 multiple-choice items.

3.1. Validity

The research of analysis validity and reliability that has been done by taking the sample of students' answer sheets and the test instrument of the Final Term Examination of Bahasa Indonesia subject 7th grade the SMP N 2 Surakarta academic year 2015/2016 has shown that the test instrument of the School Final Examination in the form of multiple-choice test items occurred the invalid points, it shown on Figure 1. The test instrument consists of 45 multiple-choice items. Based on the computer calculation of Microsoft Excel, the validity of the Learning Result Test obtained showed that seven numbers are valid and 48 numbers are invalid (Table 1)

TABLE 1: Calculation Validity of Indonesian Language Multiple-Choice Items of Even Semester Class 7A SMP N 2 Surakarta Academic Year 2015/2016.

Category	Item
Valid	8, 10, 16, 21, 22, 23, 30
Invalid	1, 2, 3, 4, 5, 6, 7, 9, 11, 12, 13, 14, 15, 17, 18, 19, 20, 24, 25, 26, 27, 28, 29, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45

A valid Learning Result Test is one that can be measured accurately in relation to the object that will be measured, whereas an invalid Learning Result Test is inappropriate for measuring the condition that will be measured. So, determining the harmony between the Learning Result Test (THB) and the learning outcomes requires the validity of the test result to be tested. The validity of the multiple-choice test items is calculated by the rpbis (biseral point correlation) formula as follows [4]:

$$rpbis = \frac{Mp - Mt}{SDt}\sqrt{\frac{p}{q}}$$

where:

rpbis: biserial point correlation coefficient

Mp: the average score of the count for the item being answered correctly

Mt: the average score of the total score

SDt: standard deviation total score

p: the proportion of students who answered correctly the items tested for their validity

q: the proportion of students who answered incorrectly the items tested for their validity

Figure 1 shows the acquisition of a variable value for: (1) the average counted correct answer score (Mp); (2) the average score of the total score (Mt); (3) the acquisition of the variable value of the proportion of students who answered correctly the items tested for the validity (p); (4) the proportion of students who answered incorrectly valid test items (q); (5) the standard deviation of the total score (SDt); (6) the biserial point correlation coefficient (rpbi); (7) r_{table} (rt); and (8) interpretation of THB validity.

3.2. Reliability

Reliability is the constancy of a test to measure or observe something that becomes a measurement object [6]. "Reliability is expressed as the constancy of particular
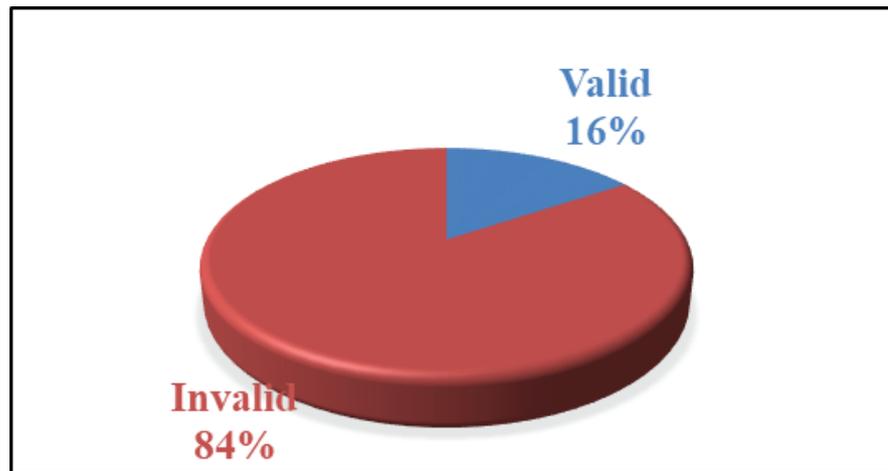
Figure 1: Calculation Validity of Class 7A SMP N 2 Surakarta.

instruments in producing the same in repeated measurements. An instrument is considered reliable if the instrument produces the same result every time when used to evaluate identical measurements" [7]. This means that reliability is seen as an instrument constancy to create a measurement. The instrument can be used if an instrument produces the same result to evaluate a measurement.

A reliable Learning Result Test provides a constant result learning measurement consistently. In contrast, a THB that is not reliable is not constant in giving the result learning measurement consistently. The calculation of reliability by using the KR20 method is as follows:

$$r_{11} = \left(\frac{n}{n-1} \right) \left(\frac{S2 - \sum pq}{S2} \right)$$

$$r_{11} = \left(\frac{26}{25} \right) \left(\frac{8.6923 - 5.6361}{8.6923} \right)$$

$$r_{11} = (1.04)(0.3516)$$

$$r_{11} = 0.3657$$

3.3. Discussion

According to the validity calculation of 5% significance of 26 students with correlation table 0.388 above indicated that 45 number of test items, 16% of test items are valid, and 84% of test items are invalid. The reason why the seven numbers of test items

are valid because of it suitable with the standard competence and the basic competence, and in addition the students' understanding of the test instrument is good, so the instrument can be said to be suitable or valid. In contrast, the reason why the test items are invalid is because there are some test items that are not suitable to the standard competence and basic competence, the making of item multiple choice that is not in accordance with the order of basic competence, the distractors are not working properly, and the lack of exercise held by the teacher approach to School Final Examination. In addition, some problems lie with the students themselves. The lack of understanding of the students of the material explained by the teacher is one of the reasons why the students cannot answer the test items, so this will make the test items invalid.

The multiple-choice items that have not been in accordance with the basic competence (KD) are first item number 34 with KD 5.1. This should be created in the first semester. There are basic competencies that are not written into the multiple-choice items of the School Final Examination, namely KD.9.2, KD.16.1, and KD.16.2. In addition, item number 41 is not in accordance with KD.11.3, is the finding information quickly from the table/diagram. However, the item provides a statement about finding important information from a paragraph.

The results of the reliability index of the test instrument of the Even Semester Final Examination of Bahasa Indonesia subject of class VII the SMP N 2 Surakarta are: (a) $0.3657 \leq 0.6$, so it can be predicted that the instrument has a bad usage value. This means that the instrument cannot be used to do the measurement, so it needs to set the remedial of that instrument; (b) r_{table} is 0.388 because $r_{11} < r_{table}$, so it can be said that the measurement result of the Learning Result Test is not correlated significantly and it cannot be said that it is reliable.

4. Conclusion

The results of the discussion show that based on the significant correlation 5% by using correlation table 0.388, the validity of multiple-choice test items of Even Semester Final Examination of Bahasa Indonesia subject which consist of 45 numbers, there are seven valid numbers, and 38 numbers are invalid. The calculation result of index reliability of the test instrument of the Even Semester Final Examination Bahasa Indonesia subject 7th grade SMP N 2 Surakarta is $0.3657 \leq 0.6$, so it can be interpreted that the instrument has a bad usage value. This means that the instrument cannot be used to do the measurement, so it needs to set the remedial of that instrument, (b) r_{table} is 0.388

because $r_{11} < r_{table}$, so it can be said that the measurement result of the Learning Result Test is not correlated significantly and it cannot be said that it is reliable.

References

[1] Biggs DA, Robert CJ: The Moral Competence Test: An examination of validity for samples in the United States. J Moral Educ. 2015; 44(4): 497-515.

[2] Nurgiyantoro B: Penilaian Pembelajaran Bahasa Berbasis Kompetensi. Yogyakarta: Fakultas Ekonomi dan Bisnis, UGM; 2014.

[3] Fitriani, Laila N, Yuniwati: Pengaruh SK Menpan Nomor 132/Kep/M.Pan/12/2002 Tentang Jabatan Fungsional Pustakawan dan Angka Kreditnya Terhadap Produktivitas Karya Ilmiah Pustakawan Universitas Diponegoro. Jurnal Ilmu Perpustakaan. 2013; 2(2): 1-9.

[4] Purwanto: Evaluasi Hasil Belajar. Yogyakarta: Pustaka Pelajar; 2011.

[5] Sudaryono: Dasar-dasar Evaluasi Pembelajaran. Yogyakarta: Graha Ilmu; 2012.

[6] Awaludin R: Tes Hasil Belajar Soal Ujian Akhir Semester Kelas XI Madrasah Aliyah Negeri Mata Pelajaran Fiqih Di Kota Palembang Tahun 2014. Tadrib Journal; 2 (2): 1-31.

[7] Sabri S: Item Analysis of Student Comprehensive Test for Research Teaching Beginner String Ensemble Using Model-Based Teaching Among Music Students In Public Universities. IJERN. 2013; 1(12): 1-14.