OPEN ACCESS

Development of Mathematics Achievement Test for Third Grade Students at Elementary School in Indonesia

Viktor Pandra, Sugiman, Djemari Mardapi

STKIP PGRI Lubuklinggau, State of Yogyakarta University, INDONESİA

ABSTRACT

The aims of this research are: (1) to measure the difficulty level of grain test of mathematics achievement of third grade student at elementary school, (2) to know the differences of mathematics achievement test for third grade student of elementary school, (3) instrument reliability test of mathematics learning achievement of third grade students of elementary school. The research method used is research development, the data analysis to know the information about grains of mathematics test instrument using ITEMAN program. The results show: (1) the test which developed has difficulty level in the range $0,30 \leq p \leq 0,70$ with accepted category, (2), the test which developed has differences level in range $0,30 \leq B \leq 0,39$ and $0,40 \leq B \leq 1,00$ with accepted category and satisfy, and (3) the test which developed has reliability coefficient of $0,783$, show that mathematics test instrument give measuring result that stabile and consistent.

Introduction

Assessment of learning outcomes must fulfill the function, namely: motivate and condition students to keep providing feedback information meaningful for students and teachers. In addition, in law number 20 of 2003 on the national education system it is stated that assessment of learning outcomes is done continuously to monitor: (1) process, (2) progress, and (3) improvement of student learning outcomes (MONE 2003: 49). Assessment of learning outcomes has a very important position in the learning process. Assessment is not only used as a tool for monitoring the learning process, but also as a means to obtain

CORRESPONDENCE V. Pandra ✉ viktorpandra@ymail.com

information about student learning progress, learning process, and improving student learning outcomes.

Some experts reveal even finding a number of facts, that the quality of education is directly related to the scoring system used. According to Indrajati Sidi (March 26, 2004), one of the educational issues related to the quality of education, is the low level of student competence, as a result of inadequate assessment system.

Kumaidi (2004) reveals that improving the quality of education requires improvements in the learning process in schools by applying systematic work methods that one of them can start from revamping the assessment system. This opinion implies that in order to improve the learning system in the school, it is necessary to obtain some information from the result of the systematic and professional assessment conducted by teachers, schools, and educational institutions.

The test as a measuring tool for obtaining some information on the development of learners should be of good quality and developed from the curriculum used, taking into account the competencies (core competencies and basic competencies) available to be used as a basis for improvement of the learning system. Similarly, the use of tests for the exam at the school level and classroom assessment by teachers is also still inadequate (Djemari Mardapi et al., 1999).

Based on the explanation above, the purpose of this research are to know: (1) to measure the level of difficulty to learn basic class mathematics, (2) to know the differences of mathematics achievement test for third grade student of elementary school, (3) instrument reliability test of mathematics learning achievement of third grade students of elementary school.

Theoretical Review

Mathematics Achievements Test

In the curriculum of elementary school mathematics education (Curriculum, 2004) mentioned that efforts to improve the quality of education needs to be done thoroughly covering aspects of knowledge, skills, attitudes and others.

The developments of these aspects are done to improve and developing life skills through a set of competencies, to make the students survive, adapt, and succeed in the future. These skills require systematic, logical, critical thinking skills that can be developed through problem solving in mathematics learning. Therefore, the construction of mathematical test encourages students to maximize their thinking ability and allow students the flexibility to develop problem-solving skills based on their experiences in everyday life.

The students' thinking ability in solving problem in mathematical will be reflected in their solving. Therefore, the steps to solve the mathematical test tend to be unlimited and the differences in sequences, depending on the students' ability in mastering mathematical. Mathematics' test tends to be expressed through questions or statements that are combined with various forms such as stories, tables, graphs and diagrams. Aspects of ability that

measured in the test of mathematics include aspects of memory, understanding, application, analysis, synthesis, and evaluation suitable with Bloom's taxonomy.

Test Construction

Based on Djemari Mardapi (2012:110) there are nine steps that must be reached in constructing test namely : (1) Arranging the specification of test, (2) writing test, (3) understanding test, (4) Test experiment, (5) Analyzing test, (6) test revise, (7) assemble the test, (8) do the test, (9) interpreting the test results. Setting the test specification is to describe the overall characteristics that a test should have. Procedures for the preparation of test specifications include: determination of test objectives, arranging the test grille, determining the form of the test, and determining the length of the test.

Hambleton & Swaminathan (1985: 226) states that the development process of the test with the response model item, includes: (1) preparation of test specifications, (2) pool preparation, (3) implementation of field testing, (4) 5) compilation of norms reference (for norm-referenced test), (6) pass score criterion specification (for criterion-referenced test), (7) study of reliability, (8) validity study, and (9) final test production.

Schmeiser & Welch (Brennan, 2006: 308) stated that,

The most important stage in development of an educational achievement test is design stage. In design stage all of the imfortant overall decisions about the test are made. These decisions include: affirming test philosopy; determining; identifying logistical and administrative constrains on test design (including test lenght and test timing); identifying relevant legal considerations that affect test design; establisihing a validation foundation for the test; designing test spesifications; and reviewing, refining, and reaffiming validity evidence for test design.

Therefore, suitable with these steps in developing test above, the development of the mathematics learning test in this research uses the following steps: (1) determination of the test objectives, (2) the preparation of the test grille (test specification), (3) writing test, (4) review and revision of questions, (5) field trials, (6) test selection, (7) test assembly, (8) test presentation, (9) scoring, and (10) reporting results test.

Degree of difficulty

The degree of difficulty of an item expressed by proportion or symbolized of P as one of parameters in analyzing the item. The degree of difficulty of an item can be calculated in various ways, one of the correct proportions of answers. Mathematically the right proportion of a particular item (p) is calculated from many testers who correctly answer (ΣB) divided by many testers (N). Mathematically written as follow :

$$p = \frac{\sum B}{N}$$

If the value of p approaches 0, then the item is too difficult, and if the value is close to 1, then the item is too easy. Items that are too easy or too difficult do not give much information about the test because they are unable to distinguish the participants' abilities so they need to be discarded.

Allen and Yen (1979: 121) stated that the level of difficulty item (p) the better between 0.3 to 0.7 as an information about the participants' maximum abilities. But the numbers 0.3 to 0.7 need to be tailored to the purpose of item development. Each item development requires different p values depending on the development objectives. Fulfillment of required p value, will obtain maximum development goals.

Different Power

The item different power is one of test parameter which give information about how much item that able to differences for the students who have high and low ability. It counts by several tolls namely coefficient correlation point biserial (r_{pbis}) and coefficient correlation biserial(r_{bis}).

The biserial point correlation coefficient is obtained based on the proportion of the testee that correctly answers the item (p), mean score on the test of the testee that has the correct answer on the item ($\overline{Y}_1$)and the mean total score $\overline{Y}$and standard deviation (S_Y). The changes are expressed in the following equation.

$$r_{pbis} = \frac{\overline{Y}_1 - \overline{Y}}{S_Y} \sqrt{\frac{p_X}{1 - p_X}}$$

If standard deviation Y is a biserial correlation (r_{bis}) is mathematically expressed in the following equation.

$$r_{bis} = \left(\frac{\overline{Y}_1 - \overline{Y}}{S_Y}\right)\left(\frac{p_X}{f(z)}\right)$$

f(z) is the normal curve ordinate with z obtained from p_X.

The coefficient of biserial point correlation function is always lower than the biserial correlation coefficient. The relationship between the two is expressed in the following equation.

$$r_{pbis} = r_{bis} \frac{y}{\sqrt{pq}}$$

Explanation :

r_{pbis} : biserial point correlation coefficient

r_{bis} : biserial correlation coefficient

y   : $Y - \overline{Y}$

p   : correct proportion of items

q   : $1 - p$

The good criterion of the item is when the discrimination index > 0.40 item is very good, (0.30 - 0.39) is good but needs improvement, (0.20 - 0.29) items with some records, usually required repairs, <0.19 items ugly, discarded or repaired via revision.

Reliability

Reliability is the consistency of measurement results or test results performed at different times with the same subject. Allen and Yen (1979: 72) state that tests are called reliable if the observational score has a high correlation with the actual score. They also stated that reliability is the correlation coefficient between two observed scores obtained from the measurements using parallel tests.

The reliability coefficients of a test can be determined in various ways, inter alia, the split method, Cronbach alpha, Guttman, and parallel methods (Ebel & Frisbie, 1986: 231) state that although there is no general provision it is widely accepted that used to make individual decisions must have at least a reliability coefficient of 0.63.

Methods

This research is a kinds of research development (research and development), namely the development of mathematics learning achievement test in Elementary School. The product that will be produced from this research is the third grade mathematics learning achievement test instrument.

The instrument development model is test using modification of Wilson Model and Oriondo and Antonio Model, with the following steps: 1) test assembly, 2) trial test, and 3) large scale test. This article was written using data up to test of test instrument.

Data Analysis Technique

Data analysis in the development of mathematics learning achievement test for third grade students aimed to obtain information related to: 1) difficulty level, 2) differentiation, and 4) instrument reliability. Data analysis to know the information of the item using ITEMAN program

Result and Discussion

Result

Trial of instrument test of mathematics achievement of third grade students of Elementary School was conducted in May s.d. June 2017 in three Elementary Schools at Lubuklinggau, such as State Elementary School 3 Lubuklinggau, State Elementary School 44 Lubuklinggau, and State Elementary School 58 Lubuklinggau. Trials aims to get information related to the level of difficulty, differentiation, and reliability of the instrument.

Degree of difficulty

The results on the difficulty table are point 4 with the difficulty index of 0.729 included in the revised category because the grain is too easy for the student in answering, the part of the drawing that corresponds to the item indicates the choice of answer.

Tabel 1. Result Analysis ITEMAN Instrument Test Achievement Mathematics Based Learning Difficulty Level

Level of Difficulty	Test Item Number	Total	Notice

P < 0.10 atau p > 0.90	-	-	Rejected
0.10 ≤ p ≤ 0.29 atau 0.70 < p ≤ 0.9	4	1	Revised
0.30 ≤ p ≤ 0.70	1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29	28	Accepted

Different power

The results on the different power table are number 2 with a different power index of 0.292 included in the revision category, therefore it is necessary to improve on the item in the form of the question sentence and answer form.

Tabel 2. Results of ITEMAN Analysis Instrument Test of Mathematics Learning Achievement Based on Different Power Level

The Level of Different Power	Test Item Number	Total	Notice
B ≤ 0.19	-	-	Does not Work
0.20 ≤ B ≤ 0.29	2	1	Revise
0.30 ≤ B ≤ 0.39	1, 5, 10, 17, 18, 22, 24, 26	8	Accepted
0.40 ≤ B ≤ 1.00	3, 4, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 19, 20, 21, 23, 25, 27, 28, 29	20	Satisfy

Instrument Reliability

Overall the test items mathematics learning achievement developed has a reliability of 0.784. According to Nunnally (1978), for the test of good reliability index description is 0.60 - 0.70 and for the objective test reliability index is 0.75 - 0.90.

Discussion

The third grade mathematics learning achievement test instrument that was developed aimed to find out information related to the difficulty level, differentiation, and reliability of the instrument. Based on the results of data analysis with ITEMAN program based on difficulty level, mathematics test instrument as much as 29 points there is 1 item that point 4 is a grain category that needs to be repaired. The images associated with point 4 indicate an overview of the choice of answers, so students can easily answer the item. Whereas 28 grains of third grade mathematics test Achievement instrument have medium difficulty level, so that the item can be accepted.

Based on data analysis with ITEMAN program based on different power level, mathematical test instrument there is 1 point that point 2 is a category need improvement. Item 2 has the same question form sentence as the answer form, thus making the students become confused in understanding the meaning

contained in the item. A total of 8 grains of mathematics learning achievement test of third grade elementary school have different level of power with medium category, while 20 other items with satisfactory category.

Overall grain test instrument of grade 3 achievement mathematics learning elementary school has a reliability index of 0.784. According to Nunnally (1978), for an objective test the good reliability index is 0.75 to 0.90, whereas Ebel & Frisbie (1986: 231) states that although there is no general provision, it is widely accepted that the tests used to make individual decisions must at least have a reliability coefficient of 0.63. Based on the description, it can be stated that the learning achievement test instrument of grade 3 primary school mathematics developed has a reliability coefficient both in other words gives a stable and consistent measurement result (Mehrens& Lehmann, 1973).

Conclusion

Based on the analysis and discussion, it can be concluded:

1.   Grain test instrument of mathematics learning achievement of third grade students of developed elementary school has difficulty level in the range $0.30 \leq p \leq 0.70$ with accepted category.

2.   The grain of the achievement test instrument of grade 3 mathematics of elementary school developed has different power level in the range $0.30 \leq B \leq 0.39$ and $0.40 \leq B \leq 1.00$ with acceptable and satisfactory category.

3.   The third grade mathematics learning achievement test instrument developed has a reliability coefficient of 0.783, indicating that the mathematics test instrument provides stable and consistent measurement results.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Viktor Pandra - STKIP PGRI Lubuklinggau, Indonesia.

Sugiman - State of Yogyakarta University, Indonesia.

Djemari Mardapi - State of Yogyakarta University, Indonesia.

References

Allen, M. J. & Yen, W. M. (1979). *Introduction to measurement theory.* Belmont, CA: Wadsworth, MC.

Brennan*, R.L.* (2006), *Educational measurement.* Iowa City: United State of America: American Council on Education and Praeger Publisher.

Depdiknas. (2003). *Undang-undang RI Nomor 20 tahun 2003, tentang Sistem Pendidikan Nasional.*

Djemari Mardapi. (1999). *Estimasi kesalahan pengukuran dalam bidang pendidikan dan implikasinya pada ujian nasional.* Pidato Pengukuhan Guru Besar. Yogyakarta: Universitas Negeri Yogyakarta.

Djemari Mardapi. (2012). *Pengukuran, penilaian, dan evaluasi pendidikan.* Yogyakarta: Nuha Litera.

Ebel, R.L. & Frisbie, D.A. (1986). *Essentials of educational measurement.* Englewood Cliffs, NJ: Prentice-Hall, Inc.

Hambleton, R.K. & Swaminathan, H. (1985). *Item response theory.* Boston, MA: Kluwer Inc.

Kumaidi. (2004). *Sistem asesmen untuk menunjang kualitas pembelajaran.* Jurnal pembelajaran, 27, 93-106.

Mehrens, W.A. & Lehmann, I.J. (1973). *Measurement and evaluation in education and psychology.* New York: Hold, Rinehart and Wiston, Inc.

Nunnally, J.C. (1978). Educational measurement and evaluation. New york: MacGraw Hill Book Company.