Mathematical Theory and Modeling
ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online) DOI: 10.7176/MTM
Vol.9, No.5, 2019

DEVELOPMENT OF NATIONAL STANDARD-BASED ASSESSMENT ASSESSMENT INSTRUMENTS FOR PROCEDURE IN MATHEMATIC LESSONS ELEMENTARY SCHOOL IN KECAMATAN WAY TUBA

Hendro Utomo[1], Rochmiyati[2], Darsono[3]
[1]hendroutomo8@gmail.com, [2]rochmiyatiazwardi@yahoo.co.id
[3]darsono.1954@fkip.unila.ac.id
University of Lampung Teacher Training and Education Faculty
Jln. Soemantri Brojonegoro No.1 Meneng Building Bandar Lampung 35145

Abstract
This study attempts to describe the fulfillment of procedures for the preparation of valid and reliable USBN assessment instruments, as well as measuring the difference power, level of difficulty, and distractors of USBN assessment instruments in Mathematics in Elementary Schools. It employs research and development (R & D) which refers to the theory of Borg & Gall. Participants were all Grade VI students of Elementary School in Way Tuba District. A purposive random sampling technique of 30 students in SD Negeri 01 Suma Mukti and 60 students at Bandar Sari Public Elementary School 01 was undertaken related to sample. Data was collected through questionnaires and test questions. Questionnaires were administered to measure the feasibility of school exam assessment instruments on mathematical subjects. In addition, test questions were involved to measure validity, reliability, power difference, level of difficulty, and question distractors. The results of the analysis indicated that the procedure for developing the USBN assessment instrument developed had advantages and had been tested for its feasibility and validity compared to USBN assessment instruments in schools in general.
Keywords: assessment instruments, mathematics, school examinations
DOI: 10.7176/MTM/9-5-07
Publication date: May 31st 2019

INTRODUCTION

Assessment is an important part of the process of implementing education, to measure and assess the level of achievement of competencies in the curriculum. It is also used to determine the strengths and weaknesses in the learning process, so that it can be used as a basis for decision making, and improvement of the learning process that has been done. Lessani (2014: 10), Pogoy & Chiu (2015: 7), Nurul and Budi (2017: 350) state that educators who are familiar with the TIMSS content domain have a significant effect on students 'mathematical achievements. The results of students' mathematical performance are influenced by the domain cognitive and preparation of knowledge of the teacher's concept of mathematics and the mathematics achievement of students in the TIMSS (Treds in the International Mathematics and Science Study). It is still low with a score below the international average. This means that mathematics achievement in students is still a problem, especially regarding to the results of mathematics learning performance and mathematical achievements. They are the results of a class assessment or assessment process conducted by independent institutions such as TIMSS and PISA.

In addition, the assessment results are expected to motivate students to continue to improve the quality of learning. This is in accordance with what was stated by Mardapi (2008: 5) that one of the efforts to improve the quality of learning is to improve the quality of the assessment system. A good assessment system will encourage educators to determine good teaching strategies and motivate students to learn better. Therefore a good, planned and sustainable assessment system is needed to support the success of the learning process.

Based on Minister of Education and Culture Regulation No. 23 of 2016 (2016: 11) concerning the standard of assessment, there are three forms of assessment instruments used in formal education at the elementary school level, namely (1) assessment instruments used by educators in the form of assessments in the form of tests,

observations, assignments. (2) Assessment instruments used by educational units in the form of final assessments and / or examinations of schools / madrasas fulfill the requirements of substance, construction, and language, and have evidence of empirical validity, (3) Assessment instruments used by the government in the form of UN meet the requirements substance, construction, language, and have evidence of empirical validity.

Furthermore, Permendikbud Number 3 of 2017 (2017: 9) states that the assessment instruments used by educational units in the form of final assessments and / or school / madrasah examinations applied today are National Standardized School Exams (USBN). The preparation of USBN questions as much as 20% -25% is made by the center as Anchor, and 75% -80% is submitted to the related subject teachers through the Teacher Working Group (KKG) forum. The role of the teacher in making USBN questions can be a benchmark or benchmark in mapping the teacher's abilities by the Ministry of Education. Through this USBN activity, teachers are expected to be able to develop quality questions. The questions made must meet the criteria and standards in writing. Through this policy the making of questions is expected to be adapted to the conditions of the region and the condition of the school given the different geographical conditions of Indonesia, so that the tests actually measure the actual abilities of participants.

Ministry of Education and Culture (2018: 4) National Standardized School Examination (USBN) is an activity to measure student achievement of competencies carried out by Education Units by referring to Graduates 'Competency Standards to gain recognition of students' learning achievements. Ministry of Education and Culture (2016: 32) the target to be achieved in school examinations is National Standardized School Exams, but local implementation; Question Instruments have high validity; Organizing School Exams with Professional Management; and Commitment to improve quality. Based on this, a good test gauge must be valid and reliable.

Moreover, assessment is the right term for assessing the learning process of students. Assessment is often referred to as one form of assessment in the form of the results of the learning process of students that is used for the basis of decision making about students, both concerning the curriculum, learning programs, school climate, and school policies. Linn and Gronlund (in Kusaeri, 2012: 9) suggest that assessment is a general term that includes procedures used to obtain information about learners' learning (observation, average written test implementation) and the format of assessment of learning progress. Wiggins (1989: 98) states that assessment is a tool that chronologically helps teachers monitor students. Whereas Stinggins (1994: 9) defines assessment as a process, progress, and student learning outcomes (outcomes).

According to the USBN POS (2018: 15) subjects that are examined in USBN for elementary schools are Indonesian Language, Mathematics, and Science. Mathematics is a compulsory subject that is always tested in the School Examination. USBN on this mathematics subject which aims to find out the competencies of students and includes to know the cognitive of students in mathematics learning.

In case of Mathematics, it is a lesson that is arranged in a regular, logical, tiered manner from the easiest to the most complicated. Russefendi in Suwangsih (2006: 3) explains that mathematics emphasizes activities in the world of ratios (reasoning), not emphasizing the results of experiments or observations. Mathematics is formed because of human thoughts, which are related to ideas, processes, and reasoning.Given the importance of assessment instruments in assessments, a teacher as a teacher is required to be able to develop assessment instruments that can measure students' abilities. Based on these problems, one solution that can be done is the need for research to develop procedures for compiling USBN Assessment Instruments in Elementary Mathematics Subjects in Way Tuba District.

METHODS

Types of Research and Procedures

This research is research and development (R & D). It aims to produce certain products. The research and development used is the design model Borg & Gall (1983: 781) which consists of 10 steps. Steps to be followed to produce the product, namely: initial research and information gathering, planning, development of initial product formats, initial trials, product revisions, field trials, product revisions, operational field trials, final product revisions, implementations. By still referring to the research and development (R & D) model by Borg and Gall (1983: 781), in this development process researchers only conducted steps one through the seventh step, namely the research step and initial information gathering until the main field trials.

Population and Samples

The population in this study were all sixth grade students with a total of 225 students in 8 public elementary schools in Way Tuba sub-district. The sample in this study was determined by purposive random sampling technique by considering if the target sample studied had certain characteristics so that other samples that did not meet the expected characteristics were not possible. The samples which were undertaken were class VI SD N 01 Suma Mukti students (20 students) and SD Bandar 01 Elementary School students (60 students).

The research instruments used were questionnaire sheets and written tests. Test techniques and non-test techniques were administered as data collecting techniques. The test technique was used to measure the knowledge possessed by each student, while the non-test technique was used in the form of questionnaires and documentation. Questionnaire was used to get the value from the validator about the feasibility of USBN assessment to collect data from the preliminary study phase, product development and product testing.

Qualitative data analysis was performed to analyze the data obtained in the form of validation score assessment for material experts, linguists, and evaluation experts to determine the level of feasibility of products produced to be used as assessment tools, the results of teacher questionnaires to measure the usefulness and feasibility of products. On the other hand, quantitative data analysis was involved to measure the instrument test which includes validity, reliability, difficulty index, difference power, and distractor.

Validity test was done to measure the level of validity or validity of an instrument, the formula used was as follows.

Information :
= biserial correlation coefficient
= average value of the subject that answers correctly for the items that are sought for validity
= average total value
= standard deviation of total value
= the proportion of students who answered correctly
= the proportion of students who answered incorrectly
Source: Sudjiono (2008: 185).

Test the validity of researchers using the IBM SPSS Statistics 2.0 program. The instrument is declared valid, or vice versa if> with α 0.05, the coefficient is declared significant. In this validity test using a significance level of 0.05 with n = 30-2 = 28. A total of 40 questions tested obtained 32 valid questions and 8 invalid questions. .
Valid questionsl were used again, but the invalid questions were revised so that they can be reused.

Another important requirement is reliability. Reliability testing is an index that shows the extent to which a measurement tool can be trusted or relied upon. Instrument reliability is needed to obtain data in accordance with the measurement objectives. To achieve this, use the following formula.

Information:
= instrument reliability
n = many items
= average total value
Source: Purnomo (2015: 147)

Table 1 List of Interpretations of the Coefficient "r"
R coefficient of reliability
Between 0.800 - 1.00 Height
Between 0.600 - 0.800 Enough
Between 0,400 - 0,600 Somewhat low
Between 0.200 - 0.400 Low
Between 0,000 - 0,200 Very Low (uncorrelated)
Source: Purnomo (2015: 147)

After calculating the reliability of the Mathematical USBN assessment instrument, the reliability coefficient was 0.982. Based on these results it can be concluded that the tests used have high reliability criteria.
Calculation of distinguishing power was done to show the extent to which each item was able to distinguish students who master the material and students who do not master the material. The formula used to determine the distinguishing power (DP) in (Kusaeri, 2014: 108) is as follows.
Information :
DP = differentiation of questions
BA = many top group participants who answer questions correctly
BB = many lower group participants who answer questions correctly
N = number of students working on the test
Source: Kusaeri (2014: 108-109)

Table 2 Criteria for Problem Distinction
No. Decision Category Distinguishing Power Range
1. 0.40 - 1.00 Very Satisfying Accepted

2. 0.30 - 0.39 Satisfying Received

3. 0.20 - 0.29 Not satisfactory Rejected / revised

4. 0,00 - 0,19 Very unsatisfactory Total revised

Source: Kusaeri (2014: 108-109)

After calculating the differentiating power of the USBN Mathematics assessment instrument, as many as 40 questions tested the power level the difference was obtained by the results of 39 questions with very satisfying categories and 1 question with satisfying categories.

Difficulty level analysis is intended to find out whether the questions are made relatively easy or difficult, besides fulfilling validity, reliability, is the balance of the level of difficulty. The items of the learning outcomes test can be said to be good if the level of difficulty is moderate.

To find out the extent of the difficulty of the question (index of difficulty), use the following formula:

Information:

P = index of difficulty

B = the number of students who answered the question correctly

= number of all test takers

Source: Arikunto (2013: 56)

Table 3 Difficulty Index Criteria

No. Difficulty Index Difficulty Index

1 0,00 - 0,30 Difficult questions

2 0.31 - 0.70 Medium Questions

3 0.71 - 1.00 Easy Questions

Source: Arikunto (2013: 56)

The results of the calculation of the difficulty index in the field test at SDN Bandar Sari class VI-B obtained 4 easy problem category questions (numbers 3, 5, 35, and 37) and 36 moderate problem questions (numbers 1, 2, 4, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 36, 38, 39, and 40).

In the question of multiple choice forms there are alternative answers (options) which are deceptive. Good questions, the trickster will be chosen evenly by students who answer incorrectly. On the other hand, items that are not good, the trick will be chosen unevenly. Impostors are considered good if the number of students who choose deception

it's the same or near the ideal. The deception index is calculated using the following formula:

Information:

1Pc = index of deception / distractor

nPc = number of students who choose the trickster

N = number of all subjects taking the test

nB = number of subjects who answered correctly on the item

Alt = many alternative answers / options (3, 4, or 5)

In addition to qualitative analysis, this study also used descriptive analysis to describe the results of the study. This study used a list of values on the results of assessment use tests to determine the ability of students. Data retrieval was done after obtaining an analysis of the validity of the items in the field trial process. The results of the analysis of the items in which the items were feasible were used and then tested for 60 students of class VI SDN 01 Bandar Sari and SDN 01 Suma Mukti as a usage test. The data obtained were then analyzed descriptively by calculating the value of learning outcomes. The formula used to calculate the value of students according to Zainal (2011: 229), is as follows.

Information:

B = Number of correct answers

N = Number of questions

RESEARCH RESULTS AND DISCUSSION

Research Results Development Procedure for Preparation of National Standardized School **Exam Assessment Instruments**

The results of this development research were in the form of a USBN assessment instrument that was feasible to be used. The material developed was on grade 6 elementary school mathematics subjects. The procedure used in this study refers to the Borg & Gall model which contains the main steps of the study as follows:

Initial Data Information Collection

The researcher started by having observation, then analyzed the needs of students and teachers. The study began with observations to see the problem of the USBN assessment instrument in the Elementary School District of Way Tuba District. The results of this initial research were material for information and reference in the preparation of test instruments prepared by researchers. The preliminary data obtained were: Test instruments at the elementary level in Way Tuba Subdistrict were in the category of basic mathematical abilities. The teacher had not yet analyzed the USBN problem grid, the questions made for USBN training were not based on Basic Competencies (KD) that must be achieved, the teacher had not compiled a question card, the analysis of students and items, and the test instruments developed had not accommodated the objectives of mathematics subjects, namely the problems of reasoning models, contextual problem solving with the lives of students.

Planning

The planning stage for developing the USBN assessment instrument includes reviewing indicators of assessment instruments, compiling grid assessment instruments, determining the behavior patterns to be measured, drafting a test grid, determining the number of items, determining the type of item, and arranging the grid in the final form in accordance with the predetermined proportion.

Development of Initial Product Draft

The next step of the framework that had been compiled before was the development of a prototype of the School Examination assessment instrument on Mathematics subjects.

Test the product

This trial was conducted to determine the validity of the USBN assessment product according to theoretical criteria. This validity includes expert assessment tests and small group trials. Appraisal of experts is used as a basis for revising and improving prototypes. Expert judgment is carried out by submitting the assessment instrument grid. Based on the validation carried out by the material experts, the value obtained was 85%, the validation value obtained from linguists was 90%, and the validation value obtained from the expert evaluation was 86.2%. While the results of small group trials through readability tests by the teacher showed results of 95.8%, and the readability test conducted by students with high, medium and low abilities showed an average yield of 93.3%.

Suggestions and input given by material experts include: the suitability between indicators and achievement of competencies need to be considered, the number of items per indicator, and each question must represent an easy, medium, and difficult instrument. Suggestions and input given by linguists include: the choice of answer in the form of numbers must be arranged in the order of the size of the value of the number, each item must use good and correct language and pay attention to spelling, and images, tables, and the like must be clear and function. While the advice given by expert evaluations is: options / options A B C D may not use lowercase letters, and the order of choice of answers from the smallest to the largest value or may be from the largest to the smallest.

Product Revision

Furthermore, the reviewers improve the USBN assessment instrument as suggested by the validator. The researcher conducted several revisions such as those suggested by material experts, namely adjusting indicators with achievement of competence, number of items for each indicator, and each question must represent an easy, medium and difficult instrument. The revision suggested by linguists is: the choice of answer in the form of numbers must be arranged in the order of the size of the number, each item must use good and correct language and pay attention to spelling, and images, tables, and the like must be clear and functioning. While what is suggested by expert evaluation are: options / options A B C D may not use lowercase letters, and the order of choice of answers from the smallest to the largest value or may be from the largest to the smallest.

Main field trial

In this main field trial, researchers tested the validity, reliability, power test differences, and test the level of difficulty on the USBN questions developed. The researcher conducted a validity test using the IBM SPSS Statistics 2.0 program. A total of 40 questions tested obtained 32 valid questions and 8 invalid questions. Valid questions were used again, but the invalid questions were revised so that they can be reused. After revisions to invalid questions, the final results showed there were 40 valid questions.

The researcher tested the reliability using the Cronbach's Alpha through calculation of SPPS. After calculation, it can be concluded that this USBN instrument had a high level of reliability with a result of 0.982. After conducting the reliability test, the researcher then conducted a different power test question. Different power tests obtained using the help of the IBM SPSS 2.0 program. The results of the calculation of the different items in the field test at SDN Bandar Sari class VI-B obtained 40 questions about the very satisfying question

categories, meaning that the questions had good differential power with the decision of the question instruments received.

After conducting a different power test, the researcher then tests the difficulty level on the question. This level of difficulty test was also calculated by using IBM SPSS 2.0 program. The calculation results of the difficulty index in the field test at SDN Bandar Sari class VI-B obtained 4 easy problem category questions (numbers 3, 5, 35, and 37) and 36 questions about the medium problem category. (number 1, 2, 4, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 36, 38, 39, and 40).

Final Product Revision

Based on the results of field trials, researchers still need to revise the final product, after the product design was validated through discussions with experts and other experts, then the weaknesses were known. These weaknesses were then tried to be reduced by means of design improvements. The material development in the instrument was validated by each part by the material validator and evaluation. It was found that the validated instrument still needed revision so that the final product expected to be tested in the trial was expanded to produce the USBN assessment instrument on mathematics subjects.

DISCUSSION

Development of procedures for preparing USBN Assessment Instruments

Development of procedures for preparing assessment instruments for national standard school examinations is based on the theory of national standardized school examination assessment instruments so that the product procedures for preparing assessment instruments for national standardized school tests are more meaningful for students and teachers in understanding procedures for preparing national standardized school examination instruments. in assessment instruments National standardized school examinations consist of question grids, question cards, question instruments, and answer keys that can be used as guidance for teachers in terms of understanding national standardized school examination assessments.

This is in accordance with the steps of the USBN assessment instrument according to Subali (2010: 41) suggesting that in order to obtain an assessment tool or a good measuring instrument it is necessary to develop a correct procedure or steps which include assessment planning that contains the purpose and objectives of the assessment, namely: 1) compilation of the grid; 2) preparation of instruments / measuring instruments; 3) review to assess the quality of the measuring instrument / instrument qualitatively, ie before use; 4) testing of measuring instruments, to investigate empirical validity and reliability; 5) implementation of measurements; 6) assessment which is the interpretation of measurement results; 7) utilization of assessment results.

The development of the USBN Assessment instrument in this study is valid and reliable. In accordance with the research of Rahayu, Purwoko, and Zulkardi (2008: 19-35) states that the instruments developed have met the criteria of valid, practical and effective. The process of developing the USBN assessment instrument through the stages of validation by experts before the instrument is tested. This aims to determine whether the quality of the instrument is appropriate or not. Small group trials were conducted after expert validation, this stage was carried out to determine the feasibility of USBN assessment instruments on mathematics subjects. The results of a small group trial for readability by a teacher obtained in the category of "very feasible" and the results of the readability test of students, the results obtained in the category "very feasible". This is in line with the research of

Tondowala & Sulvia (2012: 1-43). The results of the study show that the test instruments developed have good validity and empirical tests.

Field Trial

After conducting a small group trial, the researchers immediately carried out a field trial, namely at Bandar Sari Elementary School to determine the validity, reliability, level of difficulty, and the different power of the USBN instrument. The results of the validity and reliability tests of instruments that have 40 items with 30 respondents in SDN Bandar Sari Class VI-A use person correlation with the help of IBM SPSS 2.0 program, is r count > rtable, it can be concluded that 8 items are declared invalid and 32 items declared valid. After the question was revised, the question was tested again in a different class, namely at SDN Bandar Sari class VI-B with the number of students 30 students, it can be concluded that 40 questions were declared valid.

The reliability test results were obtained by 30 students in the field test at SDN Bandar Sari class VI-A and 30 students in class VI-B, the results of the calculation of Cronbach's Alpha reality using IBM SPSS 2.0 assistance, stated to have high reliability. This is in accordance with the opinion of the Ministry of Education and Culture (2016: 32) the target to be achieved in school examinations is the National Standardized School Examination, but local implementation; Question Instruments have high validity; Organizing School Exams with Professional Management; and Commitment to improve quality. Based on this, a good test gauge must be valid and reliable.

Based on the results of the calculation of the study of 30 students in the field test at SDN Bandar Sari class VI-A, the results of the calculation of the difficulty index in the field test at SDN Bandar Sari class VI-A obtained 4 questions in the easy question category, 35 questions in the medium problem category, and 1 about difficult categories. While the results of the calculation of the difficulty index in the field test at SDN Bandar Sari class VI-B obtained 4 questions in the easy problem category and 36 in the medium problem category. This is in line with the research of Tondowala & Sulvia (2012: 1-43). The results of the study show that the test instruments developed have good validity and empirical tests.

CONCLUSION

This research produced a valid and reliable USBN assessment instrument product on mathematics subjects in grade VI students; The product of the USBN assessment instrument on the mathematics subjects of grade 6 elementary school students produced good power that had moderate difficulty, a valid instrument, and high instrument reliability.

REFERENCES

Arikunto, Suharsimi, 2013. Research Procedures A practical approach. Rineka Cipta. Jakarta.

Borg, W. R. & Gall, M.D. (1983). Educational researcher: An introduction, (4th ed.). New York: Longman.

Ministry of Education and Culture. 2013. Teacher Training Materials Implementation of 2013 Curriculum. Education and Culture Human Resources Development Agency and Education Quality Assurance Ministry of Education and Culture.

Kusaeri and Suprananto, 2012. Educational Measurement and Assessment. Graha Ilmu. Yogyakarta

Mardapi, D; Kartowagiran, B, Nurcholis, and Fauzan. 1999. "Ebtanas Implementation Procedure". Research Report. Pusbangsisjian Cooperation between Balitbang Dikbud and the IKIP Yogyakarta Research Institute.

Minister of Education and Culture. 2016c. Permendikbud No.23 of 2016 concerning Assessment Standards. Ministry of Education and Culture. Jakarta

Purnomo, Edy. 2015. Basics of Educational Evaluation. Earth Literacy. Jakarta.

Stiggins, R. J. 1994. Student-Centered Classroom Assessment. Maccmillan College Publishing Company. New York.

Subali, B. 2010. Assessment, Evaluation, and Remedial of Biological Learning. Yogyakarta State University. Yogyakarta.

Sudjiono. 2008. Qualitative Research Procedure. PT. Rineka Cipta. Jakarta.

Tondowala, Sulvia Fery Hanry. 2012. Development of Anderson and Krathwohl's Taxonomy-Based Multiple Choice Objective Tests for Reading Ability of Grade VIII English in Middle School in Poso District, Central Sulawesi Province. Thesis, Research and Educational Evaluation, Postgraduate Program, University of Education Ganesha.

Wiggins, G. 1989. "A True Test: Toward More Authentic and Equitable Assessment". Phi Delta Kappa International. 70 (9): 703 - 713.

Zainal, A. 2011. Alternative Assessment. Director General of Higher Education. Jakarta.