Analysis Of Quality Item Tes National Standar School Exam Social Science Education (USBN-IPS) For Junior High School

by Sudarman Sudarman

Analysis Of Quality Item Tes National Standar School Exam Social Science Education (USBN-IPS) For Junior High School

Sudarman

1) Faculty of Teacher Training and Education, Mulawarman University, Samarinda, Indonesia
E-mail: sudarman@fkip.unmul.ac.id

Abstract. This research is an evaluative research that aims to know the quality of the problem in terms of material aspects, problem construction, language, distribution of cognitive domain level, validity, reliability, difficulty level, distinguishing power, and swizzle effectiveness. Objects studied are 40 items multiple choice questions with the number of respondents of this study as many as 102 students. Data collection methods used are documentation and focus group disscussion. This research uses qualitative and quantitative data analysis that interpret the calculation data. The results showed that the questions of USBN IPS SMP Kota Samarinda 2016/2017 academic year is a matter of good quality, because based on analysis and analysis of the contents of the criteria of a good question (77.5%). On the Aspect of the Level of question of items that are classified as difficult as 6 items or 15%, the items are classified as being 19 or 47.5%, and the items are easily classified amounting to 9 items or 22.5%, very easy and very difficult 3 each item or 7.5% respectively. There are 28 (70%) good quality questionnaires, 4 (10%) very good quality items, 6 (15%) items of poor quality, and 2 (5%) items quality is not good.

Keywords: Item, Feasibility, School Examination

I.   INTRODUCTION

Improving the quality of learning programs requires evaluation of the quality of the previous learning program. Thus, being able to update the educational program, including the learning program evaluation activities on the program that is being or had been conducted needs to be done well. To be able to develop better programs, the results of previous program evaluations are indispensable [1]. Evaluation of learning can be effective if using the right measuring tool. One of the measuring tools that can be used for instructional evaluation is the test. According to Mardapi [2] test is a mean to estimate the amount of a person's ability to stimulus or question. The test can also be interpreted as a number of questions that should be given answers in aiming to measure one's ability level.

Meanwhile, evaluation is a systematic and continuous process to determine the quality (value and meaning) of something, based on certain considerations and criteria in the framework of decision making[3]. It is supported by Arikunto [4] which stated that the activity of measuring, i.e., comparing something with a certain criterion or size. Measurement is quantitative, which means the measurement is manifested in symbols of numbers.

Similar with the above opinions, measurement is interpreted as a process to determine the quantity rather than something which are learners, learning strategy, school and so on. Conducting a measurement is required measuring tools. As in education, psychology, as well as other social variables, measurement activities usually use tests as a measuring tool.

Assessment is an important stage that provides evidence of the effectiveness of a teaching and learning process. Educators use assessment results for various reasons ranging from the grade level in which assessments are used to measure students' skills or to evaluate the learning process, at the national and international levels, the assessment is used to assess the curriculum or to compare the education system. In particular, summative judgments have substantial value for the students as long as it really forces them to learn ([5],[6]). The process of evaluating the final grade at the school level undergoes various policies and implementation developments, such as based on the Coordination Meeting of the Ministry of Education and Culture dated December 22, 2016 decided the policy on the implementation of the Ujian Nasional (National Examination) in 2017, that the implementation of the Ujian Nasional is still implemented and the School Exam is upgraded to become USBN (Standard School Examination National) for some subjects.

Following up the Ministry of Education and Culture Meeting, the Regulation of the Director General of Primary and Secondary Education Ministry of Education and Culture No. 08 / D / HK / 2017 on Standard Operating Procedures (POS) of USBN was passed. Based on the POS, the mechanism of preparing the USBN test questions as much as 75% -80% of them in the following package of subjects with their completeness (answer sheet formats, scoring guidelines for description test, answer key for multiple choices) is prepared by MGMP of cities and regencies.

USBN questions are compiled 88.8% consisting of multiple choice questions because multiple choice items are used to

measure the tests at school and college, and they get considerable composition on the assessment [7]. The advantages of using multiple choice tests are generating items that have a wide range of material of what has been taught [8]. Despite these advantages, the use of multiple choice questions is often criticized. Some researchers have pointed out that multiple choice items focus on what students can remember and do not assess the extent to which they can understand, apply and analyze learning-related information [9].However, it is clear that serious multiple-choice questions can serve to assess high-level cognitive processes, although making such items require more skills than writing memory-based items ([10],[11]). Analysis of the quality of the USBN test is very crucial to improve the quality of the test and to upgrade the quality of the next test. The tests are analyzed to know which one is good and bad question. A good test can be used as a measuring tool and a reference in the making of next test. The test considered bad test and revisable can be revised then be stored in the test bank to reuse. While the test considered bad and required a significant revision should be discarded.

According to Azwar [12], the analysis of items including the analysis of the degree of difficulty and the degree of difference testing is a rare classical analysis. However, by not analyzing the item, the quality of the tested item becomes unmeasurable and invalid. It is due to the development of the test quality is not based on good calculations. With the analysis of the item, the quality of the tests will be known, and it will help teachers knowing what matters related to the development, the preparation, and the use of good tests to be maintained.

The problems occur in the academic year 2016/2017 are it is the first time for the test of USBN for IPS SMP prepared by the MGMP team and quality analysis of the subject matter of USBN IPS Subject has never been conducted. Seeing the condition, the researcher is interested to conduct a study on the quality analysis of the USBN items. The research is used to know whether the items have good quality so that they can measure the achievement of learning objectives accurately.

II. METHODS

The research is an evaluation research, which the design and evaluation procedure in collecting and analyzing data is conducted systematically to determine the value of quality of USBN IPS subject. Data collection strategy of the research is through documentation study, test draft, USBN questions and Student Answer Sheet and conducting FGD with the teacher who made the test. The number of respondents is 102 students from 10 junior high schools in Samarinda.

Data analysis is used two methods, which are: (1) using the method of review based on the assessment rubric of the items, (2) The researcher uses computer program in ITEMAN (Item and Test Analysis) version 3.00. The program is used to analyze the validity, reliability, degree of difference, difficulty level and effectiveness of distractors.

III. RESULT AND DISCUSSION

The results of the analysis of material conformity aspects, The constructions of questions and language on multiple choice and description tests (questions about higher order thinking skill). Each aspect consists of a conformity indicator between the USBN IPS questions based on what has been determined. In the aspects of the review questions are focused on the question conformity toward the indicators, the questions not containing elements (Tribes, Religions, Races, Intergroup, Pornography, Politics, Propaganda, and Violence), using an interesting stimulus, using a contextual stimulus including the quality of questions able to measure the cognitive level of reasoning (analyze, evaluate, create), choice of homogeneous and logical answers and every question has only one correct answer.

In the construction aspect to examine whether the subject matter is formulated briefly, clearly, and firmly, the main formulation of the questions and the answers choices is a necessary statement only, the point of not giving a clue to the key answer. The subject matter is free of double negative statements, the use of images, graphs, tables, diagrams, or likely is clear and functional. The length of the answer is relatively same, the answer choices do not use the statement "all the above answers are wrong" or "all of the above are correct "and likely, the answer choices in the form of numbers / time is arranged according to the order of magnitude of the numbers or chronologies and the items are not dependent on the answers of other questions.

The following review results for recommendations as follows;

TABLE. RECOMMENDATION OF STUDY RESULTS

No.	Recommendation	Amount	%
1	Rejected	9	22,5
2	Revised	14	35,0
3	Used	17	42,5

Based on the table above, there are 9 problems or 22% of the multiple choice questions recommended for not being used because of the mismatch among the problems and indicators that have been set in the grid problem. The questions categorized rejected is questions on the numbers 3, 7, 10, 12, 14, 20, 22, 26, 27.

Although there are rejected questions but there are 17 questions or 42.5% questions valid to use which are the question no.1,2,4,8,16,21,23,24,29,30,31,34,37,38, 39, 40, while there are 14 questions or 35% that need to be revised. The main construction aspects that need to be clarified and confirmed, the answer choice in numerical form should be arranged in small order to large or vice versa and graphic images, tables on the questions must have optimal function in the explanation of the questions and has to use of communicative language, so they are easy to understand by the students..

The empirical composition of the cognitive level of the C1 questions (remembering) considering is 7% questions, C2 questions (understanding) is 29%, C3 questions (applying) is 19%, C4 questions (analyzing) is 29%, C5 questions (evaluating) 13% and C6 questions (creating) is 3%. Following chart to show the percentage,

The Cognitive Dimension of USBN Questions Study

Remember 7%
Create 3%
Evaluate 13%
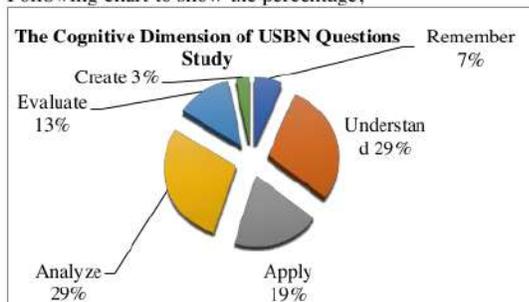Understand 29%
Analyze 29%
Apply 19%

Figure 5.1 Graph based on cognitive dimension tables

The analysis results on the aspects of validity, reliability, difficulty level, degree of difference, and deception effectiveness on multiple choices questions of USBN subjects IPS SMP, Empirical validity test is conducted using the formula point biserial correlation using of ITEMAN program. Result of calculation with significance level 5% and n = 102 hence obtained r table equal to 0,195. If r arithmetic> r table then the item is said to be valid, following table validity distribution.

TABLE. DISTRIBUTION OF QUESTIONS USBN IPS SMP ACADEMIC YEAR 2016/2017 BASED ON VALIDITY INDEX

No	Validity Indeks	Questions	Amount	Percentage
1.	≥ 0,195 (valid)	1,3,4,5,7,8,9,10,11,14,15,16,19,22,23,24,26,27,28,29,30,31,32,33,36,38,39	27	67,5%
2.	<0,195 (not valid)	2,6,12,13,17,18,20,21,25,34,35,37,40	13	32,5%

Source: Primary Data Processed

Referring to Anas Sudijono [13], one of the characteristics of a good achievement test is to have validity. A achievement test with high validity can be judged as reliable and accurate in measuring student learning outcomes.

Following up on the results of the item analysis, valid items can be reused and inserted in the question bank, but the validity analysis only has 67.5% as result. It means it needs revision for 32.5% invalid questions of USBN.

The results of reliability analysis obtained the reliability results 0.66 so it can be concluded that the item is reliable. The results of the analysis also illustrates the problem of USBN has a high coefficient and standard error of measurement (standard error of measurement) is low ". One of the characteristics of the question has a high reliability if the test consists of many items with valid categories. In addition, the high reliability of the reliability index is influenced by several factors: test length, score distribution, difficulty level, and objectivity [3].

The calculation of the degree of difference is done by dividing the subject into two parts 50% of the upper group and 50% of the lower group. The result of the calculation is interpreted in 5 criteria that is: D = negative means there is no degree of difference and preferably discarded, D ≤ 0,20 means weak degree of difference, D = 0,21- 0,40 means enough degree of difference, D = 0,41 - 0.70 means a good degree of difference, and D = 0.71 - 1.00 means excellent degree of difference. the result of calculation, ri item about which have bad difference factor amounted to 9 or equal to 22,5%, item of matter which have degree of difference good enough amount 15 item or equal to 37,5%, item having degree of difference either amounted 9 grain or equal to 22.5% and 5 item or 12,5% having negative degree of difference.

The degree of difference becomes one of the important elements in the preparation of the problem because a good question is a point that can distinguish the clever students and students who are less clever in this matter it was answered correctly by clever student. Following up from the analysis of the items after analyzing the degree of difference as follows: (1) items that have degree of difference are stored in the question bank. They can be reused during upcoming achievement test results. The items with a weak distinguishing feature have two possibilities that are not explored, they are traced to revise later and then reused in future upcoming achievement test results to determine whether their difference is improved or discarded, whereas items whose difference index numbers are negative, should be discarded because of the quality the point is very bad.

The result of the difficulty level of the problem is interpreted in 3 criteria namely: question with P 0,00 to 0,30 is difficult question; question with P 0.31 to 0.70 are moderate; and the question with P 0.71 to 1.00 is an easy matter.

Based on the results of the calculation of Level of Problems item with ITEMAN version 3.00, the items are classified as hard as 6 grains or 15%, the items are classified as being 19 or 47.5%, and the item is easily 9 or 22, 5%, very easy and very difficult cach 3 item or 7.5%.

The effectiveness of the deception choice is calculated by the Swizzle formula is calculate through Anates version 4 program. Swizzle said good if selected ≥ 5% of the number of learners. Based on the results of the analysis, there are 28 (70%) items of good quality, 4 (10%) very good items, 6 (15%) less good items, and 2 (5%) poor items.

Based on the results of quantitative analysis that includes the analysis of validity, the level of difficulty, difference, and effectiveness of deception, necessary following up on the questions. There are 3 possible follow-up actions that can be done next w are stored, revised, or discarded. A good item can be saved for future tests. The item that is less good can be improved and tested again in the next test. While a bad item can be discarded if it is not possible to be re-revised. Good items must meet the criteria, both in terms of validity, difficulty level, differentiation, and swizzle effectiveness. If

all four are good, then the item is eligible to be used as an evaluation tool. However, if there are aspects that do not meet the requirements then it should be revised again.

IV. CONCLUSIONS

1. Based on the results of the analysis of items in terms of Validity, Reliability, The degree of difference, Difficulty Level, and Swizzle Effectiveness , it can be concluded that the test of USBN IPS SMP Kota Samarinda 2016/2017 academic year can be considered as good quality test, because based on analysis and review content of the material, construction and linguistic aspects that meet the criteria of good questions only amounted to 31 out of 40 points (77.5%). Based on the analysis of items together there are 17 problems or 42.5% feasible problem to use that is the problem in which from various aspects have been eligible for use there are 14 problems or 35% that need to be improved but from the aspect of quantitative analysis already meet for use.

2. Based on the results of the calculation of Level of Problems item with the program Anates version 4, the grains are classified amounted to 6 grains or 15%, the item is classified as being 19 or 47.5%, and the item is relatively easy to number 9 grains or 22.5%, very easy and very difficult each 3 item or 7.5%.

3. Based on the results of analysis of the degree of difference and degree of swizzle, there are 28 (70%) items of good quality, 4 (10%) items of excellent quality good, 6 (15%) items less good quality, and 2 (5%) the item is not good quality.

REFERENCES

[1] Widoyoko, S. Eko Putro. 2013. *Teknik Penyusunan Instrumen Penelitian*. Yogyakarta: Pustaka Pelajar.

[2] Mardapi D. 2008. *Teknik Penyusunan Instrumen Tes dan Nontes*. Yogyakartata : Mitra Cendikia.

[3] Arifin.Zaenal.2013. Evaluasi Pembelajaran: Prinsip, Teknik, Prosedur.bandung:remaja Rosdakarya

[4] Arikunto, Suharsimi (2013). *Dasar-dasar Evaluasi Pendidikan (Jilid 2)*. Jakarta: Bumi Aksara.

[5] Race, P. (2005). *Making Learning Happen*. London: SAGE

[6] Struyven, K., Dochy, F., & Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education: a review. *Assessment and Evaluation in Higher Education, 30* (4), 325–341.

[7] Mavis, B. E., Cole, B. L., & Hoppe, R. B. (2001). A survey of student assessment in U.S. medical schools: The balance of breadth versus fidelity. *Teaching and Learning in Medicine,* 13,74-79. http://dx.doi.org/10.1207/S15328015TLM1302_1

[8] Bacon, D. R. (2003). Assessing learning outcomes: A comparison of multiple-choice and shortanswer questions in a marketing context. *Journal of Marketing Education, 25,* 31-36. http://dx.doi.org/10.1177/0273475302250570

[9] Walsh, C. M., & Seldomridge, L. A. (2006). Critical thinking: Back to square two. *Nursing Education, 45,* 212-219.

[10] Buckles, S., & Siegfried, J. J. (2006). Using in-depth multiple-choice questions to evaluate indepth learning of economics. *Journal of Economic Education, 37,* 48-57.http://dx.doi.org/10.3200/JECE.37.1.48-57

[11] Palmer, E. J., & Devitt, P. G. (2007). Assessment of higher order cognitive skills in undergraduate education: modified essay or multiple choice questions? *BMC Medical Education, 7,* 49. http://dx.doi.org/10.1186/1472-6920-7-49

[12] Azwar, Saifudin. 2000. *Tes Prestasi: Fungsi dan Pengembangan Pengukuran Prestasi Belajar*, edisi 2. Yogyakarta: Pustaka Pelajar.

[13] Anas Sudijono. (2011). *Pengantar Evaluasi Pendidikan*. Jakarta: Raja Granfindo Persada.

Analysis Of Quality Item Tes National Standar School Exam Social Science Education (USBN-IPS) For Junior High School