THE ANALYSIS OF ASSESSMENT ITEMS TEST OF MATH SUBJECT IN MIDTERM ON ODD SEMESTER OF XI IPS 2 CLASS STUDENTS AT SMA

Lisa Octaviani [a], Anggit Prabowo[b]
Program Studi Pendidikan Matematika Universitas Ahmad Dahlan
Jalan Ring Road Selatan, Tamanan, Banguntapan, Bantul Yogyakarta
[a] lisaoctt@gmail.com , [b]anggitprabowo@gmail.com

ABSTRACT

The test instrument for the odd midterm assessment questions in the XI grade of the social science class of mathematics subjects has not yet been analyzed for its characteristics. Therefore, this study aims to analyze the characteristics of midterm assessment items of the mathematic subject in State Senior High School (SMA Negeri) 1 Ngaglik. The research is a research with post facto with a quantitative descriptive approach with subjects, namely the instrument items for midterm assessment questions in the form of open questions as many as five questions. The research object is a twentieth grade in SMA Negeri 1 Ngaglik that consists of 32 students. An analyzing data used Microsoft excel on manually of validity, power difference difficulty index, and reliability. The result showed that: 1) midterm test instrument of mathematics subjects are categorized very high, which means the items are valid with a V-Aiken index value of 1 for each; 2) from 5 questions, the difficulty index obtained was one which was categorized as an easy question and four moderate questions; 3) from 5 questions, two questions have good power difference category, and three questions were categorized sufficient; 4) the reliability obtained was 0,501 which was categorized as medium.

Keywords: Midterm test questions analysis, validity, power difference, difficulty index, reliability

INTRODUCTION

Based on law Number of 14 the Year of 2005 article 1 paragraph of 10, competence is a set of knowledge, skills, and behaviors that must be owned, internalized, and mastered by the teacher or lecturer in carrying out professional duties. The teacher competency is divided into 4, among others, pedagogic competencies, which are the ability to manage the learning of students; personality competence, which means the personality ability of a teacher who is dependable and can be a role model for students; social competence which means the ability of teachers to communicate and interact efficiently with students, fellow teachers, parents/guardians of students and the surrounding community; and professional competencies obtained through professional education.

In the world of education, competencies that must be mastered by a teacher are pedagogies competencies because teachers are required not only to provide material, but teachers are required to carry out assessments and evaluations. Assessment and evaluation are needed to know the level of completeness of teaching and learning activities so that teachers can improve the learning program's quality. Evaluation is one of the important components that must be considered by the teacher in order to determine the effectiveness of learning evaluation is an action or a process to determine something value (Sudijono, 2011: 1). In the world of evaluation, education can be interpreted as a process carried out by someone (evaluator) to determine the extent to which a program's success has been achieved, which is carried out continuously. The purpose of learning evaluation is to determine the learning system's effectiveness and efficiency, concerning the goals, material, methods, media, learning resources, environment, and the assessment system itself (Arifin; 2012).

Meanwhile, based on the Minister of Education and Culture Number 66 of 2013 concerning Educational Assessment Standards, assessment of education is the process of gathering and processing information to determine the achievement of student learning outcomes, which is carried out systematically and continuously to become meaningful information. When teachers are required to evaluate and evaluate students, this means that teachers are also required to have the ability to develop instruments, especially test instruments. The test's function is a measuring tool (Arifin, 2012: 6),

meaning that the test instrument is used to measure students' ability to be used as material for evaluation by the teacher.

Giving tests in the world of education can usually be done when the material discussion of each basic competency has been completed, at midterm or the end of the semester. Giving tests is done repeatedly to be able to measure students' abilities and the results obtained remain. Test instruments have several types of forms, such as multiple-choice form tests, essay form tests, true-false form tests, matchmaking, etc. est instruments are used to measure students' abilities so that test instruments must genuinely be valid and reliable. To find out whether the test instruments given to students are indeed valid and reliable, the thing to do is to analyze the test instrument's characteristics.

The characteristics of the test instruments include validity, reliability, index of difficulty, and power difference. Validity comes from the word validity, which means the extent to which a measuring instrument's accuracy and accuracy perform its measuring function (Azwar, 2010: 5). According to Arikunto (2013: 80), in a broad outline, there are two kinds of validity, namely logical validity and empirical validity. Empirical validity is divided into 2, namely content validity and constructs validity. Simultaneously, empirical validity is divided into 2, namely, current validity and predictive validity. So, validity is used to measure the extent to which a measuring instrument's accuracy and accuracy. According to Sabri (2013:5), reliability is expressed as the constancy of particular instruments producing the same result in repeated measurements. According to the meaning, the word reliable means trustworthy (Asrul et al., 2014: 125). That is, a test is said to have a high level of trust if the test can provide the right results. Reliability means the level of stability of a test instrument, meaning that the measurement results can be trusted if the measurements are repeated and have relatively the same results. According to Mehta dan Mokhasi (2014:18) Difficulty index is merely the proportion of total students in the two groups who have answered the item correctly. The difficulty index aims to identify test instruments that are good, not right, and questions that are not good. Good questions are not too easy or not too tricky (Arikunto, 2013: 222). Questions that are too easy will make students not improve their effort in solving problems, and problems that are too difficult will cause students to become hopeless, which will make students not eager to try to work on the problem because the questions given are not by the abilities possessed by students. According to Arikunto (2013: 226), the differentiating power of questions is the ability of a question to distinguish between students who are smart (highly capable) with stupid students (low ability). Before calculating the power difference, all groups will be divided into 2, namely the upper and lower groups, which have been sorted by students who have the highest score. A good power difference is a different power that can distinguish between students who are smart with stupid students.

In Indonesia, the tendency is that teachers do not research to improve the quality of learning, find new learning method ideas, and even evaluate tools. The teacher waits for the university's research results and then tries to apply if it is considered easy to do and instead forgets if it is considered difficult for himself (Leonard, 2015: 16). So far, many teachers have only made test instruments without analyzing the characteristics of the test instruments. It is the same as in SMA Negeri 1 Ngaglik. Teachers at SMA Negeri 1 Ngaglik, especially mathematics teachers, have not yet analyzed the characteristics of the Semester Assessment questions on the questions that have been made so that based on observations made at SMA Negeri 1 Ngaglik, it was found that the semester XI grade subjects were questions in the form of questions the characteristics have not been analyzed.

Based on the discussion above, there is a need for research to identify the characteristics of the XI IPS 2 class test instrument SMA Negeri 1 Ngaglik. The objectives of this study are as follows: 1) It knows the validity of the odd semester Middle Semester Assessment question in the XI IPS 2 class at the SMA Negeri 1 Ngaglik. 2) Knowing the difficulty index of the question of odd semester Middle Semester (PTS) Assessment in the XI IPS 2 class at the SMA Negeri 1 Ngaglik. 3) It knows the power difference of the question of odd semester Middle Semester (PTS) Assessment in the XI IPS 2 class at the SMA Negeri 1 Ngaglik. 4) It knows the reliability of the question of odd semester Middle Semester (PTS) Assessment in the XI IPS 2 class at the SMA Negeri 1 Ngaglik.

METHODS

This study uses a quantitative descriptive approach. The research object used was the XI IPS 2 class students of SMA Negeri 1 Ngaglik, totaling 32 students, while the research subject was a midterm assessment question test instrument in the form of a description. The time used in this study is the odd semester of the 2018/2019 academic year. This research was conducted at SMA Negeri 1 Ngaglik, Sleman, Yogyakarta. Data collection techniques using documentation techniques and interviews. The documentation technique was carried out by getting midterm maths question sheets, grids, answer keys, and class XI maths question sheets at SMA Negeri 1 Ngaglik. The interview technique was conducted to find out the mid-semester assessment at SMA Negeri 1 Ngaglik. The data obtained were then analyzed quantitatively descriptive using Microsoft Excel to identify the characteristics of mid-semester assessment items in SMA Negeri 1 Ngaglik.

The first step is to analyze the questioned validity. The validity used is the content validity tested by two validators. After being tested, it is then inputted in Microsoft Excel, and the formula used is the V-Aiken formula. According to Hendryadi (2014: 3), the Aiken index formula, as follows.

$$V = \frac{\sum_{i=1}^{c-1} in_i}{N(c - Io)}$$

Information :

V = Validity
N = many experts
c = highest score (ex 2)
n_i = r – 1
r = ratings given by experts
l_o = lowest score

The next step is to analyze the index of difficult questions. Analyzing index difficulties between multiple-choice questions and different description questions. The initial stage of analyzing the index of difficulty in multiple-choice questions is to give a score of 1 to the correct student answer and give a score of 0 to the wrong student answer. Simultaneously, the problem description's initial stage is by adding up all students' scores given to each item. The next step is to calculate with the Aiken index formula.

The next step is to analyze the different power of students. All students will be divided into two groups: upper class and lower class. The grouping is based on students who get the highest score for students who get the lowest score. Similar to the index of difficulty, analyzing students' different power between multiple-choice questions and description questions is also different. Analyzing different power needs to sort the scores obtained by students. In multiple-choice questions, a score of 1 is given to the student's answer who answers correctly and gives 0 to the wrong student answer. At the same time, the description questions are by adding up the scores obtained by each student.

The last step is to analyze the reliability of the question. Analyzing the reliability of the problem is by using the Cronbach Alpha formula. According to Arikunto (2013: 122), The Cronbach Alpha formula is as follows.

$$r_{11} = \left(\frac{n}{n-1}\right)\left(1 - \frac{\sum \sigma^2}{\sigma^2}\right)$$

where

r_{11} = overall reliability
$\sum \sigma^2$ = number of variance scores for each item
σ^2 = total variant

RESULTS AND DISCUSSION

The validity used is content validity. Content validity is a validity test carried out on its contents to ascertain whether the learning outcomes test items measure precisely the conditions that want to be measured (Purwanto, 2009: 120). Content validity testing is done by examining the suitability of the items in the midterm assessment questions with the grid so that the content validity testing is carried out by people who have competence in the field. Two validators carry out this content validity test. The results of the validity of the midterm assessment items can be presented in table 1.

Tabel 1. The results of the validity of the difficulty of the questions

	Bullet									
	1		2		3		4		5	
Experts	**item**	**s**	**items**	**s**	**item**	**s**	**item**	**s**	**item**	**s**
1	2	2	2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	2	2	2	2
∑s		4		4		4		4		4
V		1		1		1		1		1

Table 1, the results of the calculation of validity with the Aiken Index obtained results, which is one categorized very high, meaning that the midterm assessment questions items in the XI IPS 2 class with a grid are appropriate. The greater the CVR of 0, the more important and the higher the validity of the contents. (Hendryadi, 2014: 4). So, from the results above, it is found that the CVR value of 1 has fulfilled good content validity.

The test difficulty index is the test's ability to capture the number of test subjects who can work correctly. The problem is said to have a useful index of difficulty if the question is not easy and not complicated. Arikunto (2013: 225) states that the categories used to interpret the difficulty index are $p > 0.70$ included in the easy category, $0.30 \leq p \leq 0.70$ included in the medium category, and $p < 0.30$ included in the difficult category. The results of the index analysis of the difficulty of the questions indicate that item number 1 results in 0.69 being categorized medium, items number 2 obtained the results of 0.81 categorized easily, items number 3 obtained the results of 0.65 which are categorized medium, items number 4 obtained the results of 0.43 are categorized medium and items number 5 are 0.64 which are categorized as medium. The results of the index analysis of difficulties in midterm assessment questions are presented in table 2.

Tabel 2. The results of the index analysis of the difficulty of the questions

Item number	**Difficulty index**	**Category**
1	0,69	Medium
2	0,81	Easy
3	0,65	Medium
4	0,43	Medium
5	0,64	Medium

In table 2, the results of the difficulty index analysis obtained below. There is 1 question with an index of difficulty with easy categories; four questions have an index of difficulties with the medium category. Problem number 2 has a difficulty index of 0.81, which is the most challenging index. This proves that question number 2 is a straightforward category. The number of students who can work on question number 2 indicates that students of class XI IPS 2 can understand mathematical induction well. A better understanding of the concepts that are owned will help students do mathematical induction correctly. Simultaneously, the four questions with difficulty indexing with the medium category are found in items 1, 3, 4, and 5. The difficulty index is good, which is not easy or difficult. This means the items number 1, 3, 4, and 5 have been stated to be good. Because the items are too easy or difficult for students, they cannot distinguish students' abilities (Purwanto: 2009).

The differentiating point is the ability of a question item to distinguish groups in aspects measured according to the group's differences. Arikunto (, 2013: 232) states that the categories used to interpret different powers are DP <0 including categories not used, 0.00 ≤ DP ≤ 0.20 including wrong categories, 0.21 ≤ DP ≤ 0.40 including good categories, 0.41 ≤ DP ≤ 0.70 including the excellent category, and 0.71 ≤ DP 00 1.00 including the excellent category. Different power from the category is not used because the difference in power obtained is negative, meaning that the upper class of students cannot work on the problem than the lower class of students. The results of the different power analysis show that item number 1 shows that 0.26 is categorized sufficiently, item number 2 is obtained as a result of 0.26, which is categorized sufficient, item number 3 is obtained as a result of 0.31, which is categorized sufficient, item number 4 is obtained the results of 0.49 which are categorized as good and item number 5 obtained the results of 0.41 categorized as good. The results of different power analysis midterm assessment questions are presented in table 3.

Tabel 3. The results of the different power

Item number	Different power	Category
1	0,26	Enough
2	0,26	Enough
3	0,31	Enough
4	0,49	Good
5	0,41	Good

In table 3, the results of the analysis show that three questions have different power with good categories, and two questions have different powers with good categories. This means that the items in the assessment of Middle Semester XI grade 2 at SMA Negeri 1 Ngaglik are quite capable of distinguishing students who have high abilities with students with less ability. Good questions are questions that can be answered correctly by smart students only (Arikunto: 2013).

A test is said to be reliable if it gives the results of the measurement of learning outcomes, which are relatively consistent consistently. In this case, the reliability analysis of the midterm assessment questions uses the Cronbach Alpha formula. The results of the reliability analysis of the midterm assessment questions obtained the reliability coefficient of 0.501, which means that the items have reliability in the medium category. Reliability with this medium means that the instrument will get results that are entirely consistent when used to make measurements again. This is because the higher the test's reliability coefficient, the higher the stability or accuracy (Santoso: 2013).

CONCLUSION

Based on the analysis of the characteristics of items in the midterm evaluation questions, the mathematics subjects of class XI IPS 2 in SMA Negeri 1 Ngaglik, Sleman Regency, 2018/2019 showed that of the five items mentioned above;

1. Valid categorized review of the validity of the content;
2. There is 1 question has a difficulty index with easy category, and four questions have a difficulty index with the medium category;
3. Three questions have different power with enough categories, and two questions have different powers with good categories;
4. The midterm assessment instrument's reliability coefficient is 0.501, which means that the items have reliability in the medium category.

REFERENCES

Arifin, Zaenal. 2012. Evaluasi Pembelajaran. Direktorat Jenderal Pendidikan islam Kementrian Agama RI

Arikunto, Suharsimi. 2013. Dasar-Dasar Evaluasi Pendidikan. Jakarta: Bumi Aksara.

Asrul, R. Ananda and Rosnita. 2014. Evaluasi Pembelajaran. Medan: Citapustaka Media

Azwar, Saifuddin. 2010. Reliabilitas dan Validitas. Yogyakarta: Pustaka Pelajar.

Hendryadi. 2014. *Content Validity*. Teori Online Personal Paper. 01:1-5.

Leonard. 2015. Kompetensi Tenaga Pendidik Indonesia: Analisis Dampak Rendahnya Kualitas SDM Guru dan Solusi Perbaikannya. Jurnal Formatif, 5(3):192-201

Mehta, Gyata, and Varsha Mokhasi. 2014 Item Analysis of Multiple Choice- an Assessment of the Assessment Tool. *International Journal of Health Sciences and Research*. 4: 197-202.

Pemerintah Indonesia. 2005. Undang-Undang Republik Idonesia Nomor 14 tahun 2005 Tentang Guru dan Dosen. Jakarta: Sekretariat Negara.

Pemerintah Indonesia. 2013. Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 66 tahun 2013 Tentang Standar Penilaian Pendidikan. Jakarta: Sekretariat Negara.

Prabowo,Anggit. 2016. Analisis Butir Soal Ujian Akhir Semester Mata Kuliah Analisis Kurikulum dan Materi Pembelajaran Matematika SMA: *The Progressive and Fun Education Seminar*.

Purwanto. 2009. Evaluasi Hasil Belajar. Yogyakarta: Pustaka Pelajar.

Sabri, Shafizan. 2013. Item Analysis Of Student Comprehensive Test For Research In Teaching Beginner String Ensemble Using Model-Based Teaching Among Music Students In Public University. *International Journal of Education and Research*. 01:1-14

Santoso, Yohan. 2013. Analisis Butir Soal Ujian Tengah Semester Ganjil Mata Diklat Teori Produktif Untuk Siswa Kelas X SMK Muhammadiyah 1 Bantul 2012/2013. Universitas Negeri Yogyakarta.

Sudijono, Anas. 2011. Pengantar Evaluasi Pendidikan. Jakarta: Rajawali Pers.