

THE ANALYSIS of TEST ITEMS of ECONOMICS-ACCOUNTING FINAL EXAMINATION for VOCATIONAL HIGH SCHOOL

ANALISIS BUTIR SOAL ULANGAN KENAIKAN KELAS SEMESTER GENAP EKONOMI AKUNTANSI KELAS XI IPS untuk SMK

Oleh: **Novita Isnaini**

Praktisi Pendidikan Alumni Prodi Pendidikan Akuntansi FE UNY

Email: novitaisnaini92@gmail.com

Prof. Sukirno, S.Pd. M.Si., Ph.D

Staf Pengajar Jurusan P. Akuntansi Universitas Negeri Yogyakarta

Abstract

This study was a descriptive quantitative. The technique applied for collecting data was documentation method. The data obtained were analyzed by using Anates Program Version 4.09. The results: (1) According to the validity, there were 22 valid items (55%) and 18 invalid items (45%); (2) According to the reliability, was 0.73; (3) According to the discriminating index, that 9 items had very good discriminating index (22,5%), 6 items very bad (15%), 10 items adequate (25%), 10 items good (25%), and 5 items had bad discriminating index (12,5%); (4) According to difficulty level, 1 item very difficult category (2,5%), 5 items difficult (12,5%), 24 items moderate (60%), 8 items easy (20%), and 2 items very easy category (5%); (5) According to distractor efficiency, 16 items were very good quality (40%), 10 items good quality (25%), 12 items deficient (30%), 2 items bad (5%), and no items were very bad quality; (6) According to the analysis question 13 items or 32,5% were good, fulfilling four criteria, 14 items or 35% were adequate, fulfilling three criteria, 13 items or 32,5% were bad, fulfilling less than equal two criteria.

Keywords: The Test Item Analysis, Accounting Theory

Abstrak

Penelitian ini adalah penelitian deskriptif kuantitatif. Teknik pengumpulan data dilakukan dengan metode dokumentasi. Data yang diperoleh dianalisis menggunakan program Anates Versi 4.09. Hasil dari penelitian ini adalah : (1) Berdasarkan validitas, soal valid berjumlah 22 butir (55%) dan butir soal yang tidak valid berjumlah 18 butir (45%); (2) Berdasarkan reliabilitas, 2016 memiliki reliabilitas keseluruhan 0,73; (3) Berdasarkan daya pembeda, buruk berjumlah 5 (12,5%), cukup baik berjumlah 10 (25%), baik berjumlah 10 (25%), dan yang sangat baik berjumlah 9 (22,5%); (4) Berdasarkan tingkat kesukaran, kualitas sangat mudah adalah 2 (5%), Butir soal yang memiliki kualitas mudah berjumlah 8 (20%), butir soal yang memiliki kualitas sedang berjumlah 24 (60%), butir soal yang memiliki kualitas sukar berjumlah 5 (12,5%), dan kualitas sangat sukar berjumlah 1 (2,5%); (5) Berdasarkan indeks pengecoh, kualitas sangat baik adalah 16 (40%), Butir soal yang memiliki kualitas baik berjumlah 10 (25%), butir soal yang memiliki kualitas cukup baik berjumlah 12 (30%), butir soal yang memiliki kualitas buruk berjumlah 2 (5%), dan tidak ada butir soal yang memiliki kualitas sangat buruk.

Kata kunci: Analisis Butir Soal, Ekonomi Akuntansi

INTRODUCTION

The development of science and technology makes a lot of changes to humans' life and bring people to the global competition. The challenge and development of education in Indonesia at this time and in the future will be more complex. This is caused by changing demand of the people for the quality and quantity of the education itself.

Regarding learning system (learning as a system), evaluation is one of important components and stage teachers must go through in order to know the effectiveness of learning. The results obtained from evaluation can serve as feedback for them to reform and improve learning program and activity. In school, it is common that teacher frequently gives daily test, mid test, block exams, written test, spoken test, performance test, etc.

Tests conducted in MAN 2 Yogyakarta are similar to those conducted in the other schools. In the middle of semester, it is conducted mid-term test; at the end of semester, it is conducted final test (*UAS*). Learning process of Economics and Accounting was divided into two lesson hours for Economics, and three hours for Accounting. Meanwhile, tests will be made of the combination of the two. This is done in order to collect

learners' scores according to their understanding of Economics and Accounting subject.

Evaluation is needed to achieve education level. It is a process of providing information that could serve as consideration for determining the worth and merit from the purposes achieved, design, implementation and effect to help in making decision, helping accountability, and improving understanding of phenomena.

In education field, viewed from the targets, evaluation is divided into macro and micro evaluation. Macro evaluation's target is educational program, namely program planned to improve education field. Micro evaluation is commonly conducted in class level. Thus, micro evaluation's target is learning program done in the class with teacher as one being in charge of school and lecturer as one being in charge of college (Djemari Merdapi, 2008: 2). Teacher takes responsibility to plan and implement learning program in the class, while headmaster is responsible for evaluating learning program teacher already planned and implemented.

Test is one of measuring tools, namely a tool used to collect information about characteristics of an object. Test is the

smallest unit of assessment. According to Djemari Merdapi (2008: 67) test is one way to estimate the extent of someone's ability to the stimulus or question. Moreover, test is defined as a number of questions should be answered with the aim of measuring the level of one's ability or expressing particular aspect from testee. The response testees show to a number of questions or statement described the ability in a specific area. Test is employed to measure learning results in the form of hard skills.

The analysis of the quality of the test items is a process which is needed to measure the quality of the test items so that those which have a good quality can be identified and can be used again. Meanwhile, the items which have low quality should be revised and those which have bad quality can be removed. If the quality of the items which are given in a test is doubted, the results of the test might have bad quality. The results from the test with the items which have bad quality cannot reflect the exact students' achievement because they will give wrong information about the students' performance.

In fact, the Accounting teachers in MAN 2 Yogyakarta have not completely analyzed the test items like what has been mentioned above. The teachers only

analyze them from the aspect of item facility and discrimination index and then decide whether those items are acceptable or not to be tested. The quality of the test items is not definitely known because those two criteria cannot truly explain it exactly. The teachers should carry out the analysis of the test items based on the five criteria to find out the quality of the test items exactly.

RESEARCH METHOD

Types Of Research

This study belongs to descriptive quantitative study. This is because it reveals real facts and then describe those facts quantitatively and intended to test research questions. This study was conducted in Madrasah Aaliyah Negeri (MAN) 2 Yogyakarta which is located in K.H. Ahmad Dahlan street, number 130, Ngampilan, Yogyakarta 55261.

Subject and Object

The subject of this study is all students class XI IPS at MAN 2 Yogyakarta which numbers 64 students and the object of this study is Test Items and Answer Key for The Analysis of Test Items of Economics-Accounting Final Examination of Even Semester for Grade XI IPS Students of MAN 2 Yogyakarta in Academic Year 2015/2016.

Collecting Data

Documentation method was employed to collect data regarding test items for the Even Semester Final Examination accompanied by answer key and answer sheet of testees involved in summative evaluation of the learning results for accounting subject obtained from teacher of Economic Accounting grade XI IPS at MAN 2 Yogyakarta. The Analysis of Test Items of Economics-Accounting Final Examination of Even Semester for Grade XI IPS Students of MAN 2 Yogyakarta are in the form of multiple choice or objective are analyzed first and scores are given to each student's answer. The scoring scale is 0-1, score 1 for the incorrect answers while score 1 for the correct answers.

Technique of Analysis

Analysis of test items means analysing each question tested in a test so that produces high-quality test items that can be used to measure students' learning achievement. There are some aspects need to be considered in doing analysis of test items.

FINDING AND DISCUSSION

This research was carried out in order to find the quality The Analysis of Test Items of Economics-Accounting Final Examination of Even Semester for Grade XI IPS Students of MAN 2 Yogyakarta in Academic Year 2015/2016 seen from

validity, reliability, discrimination index, difficulty level, and distractor aspects. The data were the test items of Economics Accounting Final Examination of Even Semester containing 40 multiple choice questions. The data were processed using *Anates Version 4.09* to figure out the validity, reliability, discrimination index, difficulty level, and distractors.

The results of The Analysis of Test Items of Economics-Accounting Final Examination of Even Semester for Grade XI IPS Students of MAN 2 Yogyakarta in Academic Year 2015/2016 are:

Validity

Validity can be examined rationally (rational validity) and empirically (empiric validity). To determine rational validity, a tracking can be done using content validity. Content validity can be identified by using question blueprints whether the test items are already appropriate with the indicators.

Correlation coefficient formula is functioned to examine the empirical validity. The total number of the students of XI IPS class are 64 students that consist of 22 students of XI IPS 1 class, 21 students of XI IPS 2 class, and 21 students of XI IPS 3 class. In reference to the number of the research subjects, 64 students, the score gained was 0.250.

Thus, the research result toward the analysis of the test items' validity according to ≥ 0.250 which means that the test items are valid. Otherwise, if the score is < 0.250 , the test items are not valid. 0.250 was resulted from Anates in which r table below is at 60 and 0.250.

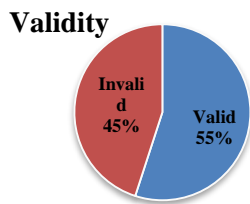


Figure 3. The distribution according to Empirical Validity

Reliability

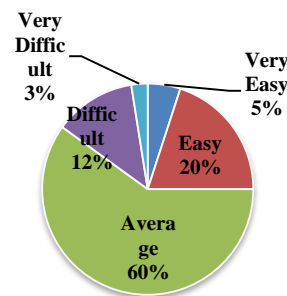
The research result of the analysis of test items' reliability according to $r_{11} \geq 0.70$, the examined test items have high reliability (reliable), otherwise, if $r_{11} < 0.70$, the examined test items have low reliability (unreliable). Based on the analysis of the test items of Economics-Accounting Final Examination of Even Semester for grade XI IPS students of MAN 2 Yogyakarta in the Academic Year of 2015/2016, the test items were 0.73 which means that $r_{11} > 0.70$, hence the test items were reliable.

Difficulty Level

The criteria used to interpret calculation result of difficulty level are 0%-15% categorized as very

difficult, 16%-30% categorized as difficult, 31%-70% categorized as average, 71%-85% categorized as easy, and 86%-100% categorized as very easy.

According The Analysis of Test Items of Economics-Accounting Final Examination of Even Semester for Grade XI IPS Students of MAN 2 Yogyakarta in Academic Year 2015/2016, there were 2 (5%) very easy, 8 (20%) easy, 24 (60%) average, 5 (12.5%) difficult, and 1 (2.5%) very difficult test items. The distribution of 40 test items according to the difficulty level is:



Discrimination index

Classification used to interpret the test result of discrimination index is $< 9\%$, so the test items are classified as bad. 10%-19% is categorized as bad, 20%-29% is enough, 30%-49% is good and $\geq 50\%$ is very good.

From the analysis, it was found that there were 6 (15%) very bad, 5 (12.5%) bad, 10 (25%) enough, and 9 (22.5%) very good test items.

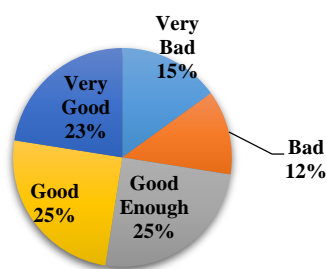


Figure 4. The distribution according to Discrimination index

Distractor efficiency

The pattern of answer distribution can be identified by calculating the number of test participants who choose a, b, c, d, and e or who does not choose any options. Distractor efficiency can be calculated using distractor efficiency by Anates version 4.09. A distractor is good, if it is $\geq 5\%$ than the number of participants.

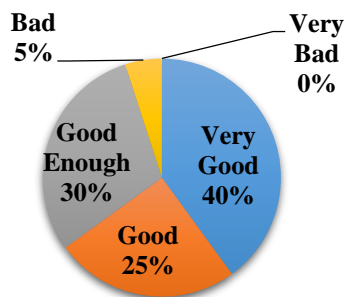


Figure 6. The distribution according to Distractor efficiency

Analysis of Test items According to Validity, Difficulty Level, Discrimination index, and Distractor efficiency

After doing the testing of each category, the test items were then analyzed based on the validity, reliability, discrimination index, and

distractor efficiency to figure out the quality of the whole The Analysis of Test Items of Economics-Accounting Final Examination of Even Semester for Grade XI IPS Students of MAN 2 Yogyakarta in Academic Year 2015/2016. Quality of test items is divided into five: very good, good, enough, bad, very bad based on some consideration below:

- a. A good test item is good if it fulfills four criteria. In validity, it is valid. In discrimination index, it is very good, good and enough. In difficulty level, it is average/medium. In distractor efficiency, it is very good, good, and enough.
- b. A test item is good enough, if it fulfills only three criteria, while the other one does not. The other aspect does not fulfill validity (invalid). In discrimination index, it is bad or very bad. In difficulty level, it is very easy, easy, and very easy. In distractor efficiency, it is bad or very bad.
- c. A test item is not good, if it does not fulfill two or more determined criteria.

The analysis of the whole The Analysis of Test Items of Economics-Accounting Final Examination of

Even Semester for Grade XI IPS Students of MAN 2 Yogyakarta in Academic Year 2015/2016 according to validity, reliability, discrimination index, difficulty level, and distractor efficiency is as follows:

Table 12. The Analysis of the whole The Analysis of Test Items of Economics-Accounting Final Examination of Even Semester for Grade XI IPS Students of MAN 2 Yogyakarta in Academic Year 2015/2016 according to Validity, Reliability, Difficulty Level, Discrimination Index, and Distractor Efficiency

No	Quality	Test items	Total	Percentage
1	Bad	1, 2, 3, 8, 10, 11, 14, 16, 21, 30, 33, 36, 39	13	32,5%
2	Enough	4, 7, 9, 15, 17, 18, 22, 23, 24, 27, 28, 34, 37, 40	14	35%
3	Good	5, 6, 12, 13, 19, 20, 25, 26, 29, 31, 32, 35, 38	13	32,5%

Source: Primary Data

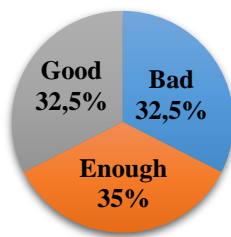


Figure 7. The Analysis of the whole The Analysis of Test Items of Economics-Accounting Final Examination of Even Semester for Grade XI IPS Students of MAN 2 Yogyakarta in Academic Year 2015/2016 according to validity, reliability, difficulty level, discrimination index and distractor efficiency

Source: Primary Data

In reference to the data, it is concluded that the quality of the test

items of Economics-Accounting Final Examination of Even Semester for grade XI IPS students of MAN 2 Yogyakarta in the Academic Year of 2015/2016 had 13 (32.5%) questions items fulfilling four criteria, 14 (35%) test items fulfilling three criteria, and 13 (32.5%) fulfilling only 2 or less than 4 criteria.

The causes of the failure are as the following:

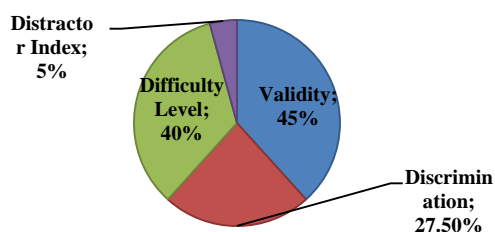
Table 13. Causes of Failure

No	Causes of Failure	Test items	Total	Percentage
1	Validity (Invalid)	1, 2, 3, 7, 8, 10, 11, 14, 16, 21, 22, 24, 28, 30, 33, 34, 36, 39	18	45%
2	Discrimination index (Very bad and bad)	1, 2, 8, 10, 11, 14, 30, 33, 36, 37, 39	11	27.5%
3	Difficulty Level (very difficult, easy, very easy)	1, 2, 3, 8, 9, 10, 15, 6, 17, 18, 21, 23, 27, 30, 33, 40	16	40%
4	Distractor efficiency (Less good and not good)	4, 8	2	5%

According to the table above, the biggest cause of failure was at validity showing 45%. Validity means that the test items were not appropriate yet. Diffiulty level got percentage of 40%. It showed that the questions were too easy or difficult, so there were some questions that were easy to answer by most of even all students.

The failure was also caused by discrimination index that showed 27.5%. It proved that discrimination index of the test items were not good and unable to distinguish whether the students understood the materials or not. Lastly, distractor efficiency of the test items showed 5% which means that the distractors did not function well.

Figure 8. Percentage of the Failure of the Test items



The good test items were put into a question bank and used, while the less good test items were revised and the causes of the problems were found and the bad ones were eliminated.

CONCLUSIONS AND SUGGESTIONS

Conclusions

In reference to the results of the analysis of the test items consisting of validity, reliability, discrimination index, difficulty level, and distractor index of The Analysis of Test Items of Economics-Accounting Final Examination of Even Semester for Grade XI IPS Students of MAN 2 Yogyakarta in Academic Year 2015/2016, the test items were in good

quality. According to the whole analysis, the data showed that there were 13 or 32.5% good items which fulfilled four criteria, 14 or 35% items which fulfilled three criteria, and 13 or 32.5% bad items which fulfilled only less than and equal to two criteria. The conclusions were obtained from the following analysis:

Suggestions

Based on the results of the analysis of the test items consisting of validity, reliability, discrimination index, difficulty level, and distractor index of The Analysis of Test Items of Economics-Accounting Final Examination of Even Semester for Grade XI IPS Students of MAN 2 Yogyakarta in Academic Year 2015/2016, the suggestions go as follows:

1. Good questions should be stored in a question bank and reused later by giving some modification on the next final examination. Less good questions should be revised and the causes of the problems should be found. Meanwhile, the bad ones should be eliminated.
2. The headmaster should give training to the teachers about how the analysis of test items for tests should be conducted.
3. Question makers should enhance their comprehension in writing test items. It is because test instruments consisting of 13 or 32.5% good items which

fulfilled four criteria, 14 or 35% items which fulfilled three criteria, and 13 or 32.5% bad items which fulfilled only less than and equal to two criteria.

REFERENCES

Djemari Merdapi. (2008). *Teknik Penyusunan Instrumen Tes dan Nontes*. Yogyakarta: Mitra Cendikia.

Marzuki. (2014). Analisis Butir Soal Ulangan Tengah Semester Genap Mata Pelajaran Produktif Akuntansi Kelas X Program Keahlian Akuntansi SMK Negeri 1 Yogyakarta TA 2014/2015. *Skripsi*. Yogyakarta: FE UNY