THE TEST ITEM ANALYSIS OF FINAL EXAMINATION ON ECONOMICS SUBJECT

ANALISIS BUTIR SOAL UJIAN AKHIR MATA PELAJARAN EKONOMI AKUNTANSI

Oleh: **Febrianti Dian Sari**
Prodi Pendidikan Akuntansi Universitas Negeri Yogyakarta
febriantidian1992@gmail.com

Abstract

This study aimed to examine the quality of the test item analysis of 1st semester final examination on economics subject in grade XII IPS SMA N 1 Kalasan academic year 2015/2016 viewed from the validity, reliability, discrimination index, the level of the difficulty, and distractor efficiency. 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 of this study were as follows: (1) According to the validity, there were 26 valid items (52%) and 24 invalid items (48%); (2) According to the realibility, the results of the study showed that the overall realibility of the test items for the first semester final examination on economic accounting subject grade XII Social at SMA N Kalasan in the academic year 2015/2016 was 0.61; (3) According to the discrimination index, the result showed that 10 items had very good discrimination index (20%), 12 items had very poor discrimination index (24%), 9 items had adequate discrimination index (18%), 12 items had good discrimination index (24%), and 7 items had poor discrimination index (14%); (4) According to level of the difficulty, 1 item belonged to very difficult category (2%), 5 items belonged to difficult category (10%), 26 items belonged to moderate category (52%), 8 items belonged to easy category (16%), and 10 items belonged to very easy category (20%); (5) According to distractor efficiency, 9 items were very good quality (18%), 12 items were good quality (24%), 14 items were deficient in quality (28%), 9 items were poor quality (18%), and 6 items were very poor quality (12%); (6) According to the analysis of question items done simultaneously, 14 items or 28% were good, fulfilling four criteria, 7 items or 14% were adequate, fulfilling three criteria, 29 items or 58% were poor, fulfilling less than equal two criteria.

Keywords**:** The Test Item Analysis, Accounting Theory, SMA N 1 Kalasan.

Abstrak

Penelitian ini bertujuan untuk mengetahui kualitas butir soal Ujian Akhir Semester Mata Pelajaran Ekonomi Akuntansi Kelas XII IPS di SMA N 1 Kalasan Tahun 2015/2016 yang ditinjau dari validitas, reliabilitas, tingkat kesukaran, daya pembeda dan indeks pengecoh. Penelitian ini merupakan 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, butir soal yang valid berjumlah 26 butir (52%) dan butir soal yang tidak valid berjumlah 24 butir (48%); (2) Berdasarkan reliabilitas, Hasil penelitian menunjukkan bahwa reliabilitas Soal Ujian Akhir Semester Gasal Mata Pelajaran Ekonomi Akuntansi Kelas XII IPS SMA N 1 Kalasan tahun ajaran 2015/2016 keseluruhan adalah 0,61; (3) Berdasarkan Daya Pembeda, hasil analisis butir soal dengan daya pembeda sangat baik berjumlah 10 (20%), soal dengan daya pemebda buruk 12 (24%), soal dengan daya pembeda cukup baik berjumlah 9 (18%), soal dengan daya pembeda baik berjumlah 12 (24%), dan soal dengan daya pembeda sangat baik berjumlah 7 (14%); (4) Berdasarkan Tingkat Kesukaran, soal yang termasuk kategori soal sangat sukar ada 1 soal (2%), soal yang masuk kategoi sukar 5 soal (10%), soal yang termasuk kategori sedang ada 26 soal (52%), tergolong mudah terdapat 8 soal (16%), dan soal yang masuk kategori sangat mudah terdapat 10 soal (20%); (5) Berdasarkan Indeks Pengecoh, butir soal yang memiliki kualitas sangat baik 9 (18%), butir

soal yang memiliki kualitas baik berjumlah 12 (24%), butir soal yang memiliki kualitas cukup 14 (28%), butir soal yang memiliki kualitas kurang baik berjumlah 9 (18%), butir soal yang memiliki kualitas tidak baik berjumlah 6 (12%); (6) Berdasarkan analisis butir soal secara bersama-sama sebanyak 14 butir atau 28% merupakan soal yang baik yang memenuhi empat kriteria, 7 butir atau 14% merupakan soal yang cukup baik yang memenuhi 3 kriteria, 29 butir atau 58% merupakan soal yang tidak baik yang memenuhi kurang dari sama dengan dua kriteria.

Kata kunci: Analisis Butir Soal, Ekonomi Akuntansi, SMA N 1 Kalasan

INTRODUCTION

Education is the gradual stage for humans to change their attitude, behaviour through the learning process. The meaning of education itself has already been written into *UU No. 20 Tahun 2003* about *Sistem Pendidikan Nasional*, namely:

"*Pendidikan adalah usaha sadar dan terencana untuk mewujudkan suasana belajar dan proses pembelajaran agar peserta didik secara aktif mengembangkan potensi dirinya untuk memiliki kekuatan spiritual keagamaan, pengendalian diri, kepribadian, kecerdasan, akhlak mulia, serta keterampilan yang diperlukan dirinya, masyarakat, bangsa, dan negara.*"

UU No. 20 Tahun 2003 describes the details of the nature and aspects of education term from which both can be analysed, compared, abstracted, and combined for an essential formulation. Moreover, the essential aspects included in the meaning of education are guiding (character building), developing, improving (from unknowing into knowing), the existence of interaction between teacher and students, long-life learning process, and complete self-actualization in terms of developing human's full potential.

One of major factors determining the quality of education is teacher. Teacher is the main actor taking charge of the development of the quality of human resource. Teacher directly deal with students in the classroom through learning process. The success in learning process heavily depends on the teacher's performance in teaching and this can only be performed by those who have part in particular study which is designed for preparing prospective teachers. This statement shows that teaching is a profession and being a teacher is a professional duty.

For teaching and learning process to take place, it is very necessary for teacher to be able to plan, teach, and also evaluate the learning process and outcomes by means of conducting the evaluation. Evaluation is an activity of collecting data in order to measure to what extent does something have met the objectives (Suharsimi Arikunto, 2013: 25). Measurement and assessment are included in the evaluation. The former means comparing something by using a measure while the last

means making decision about something by using several criteria.

There are two instruments used to evaluate learning outcomes at the school, namely test and non-test (Arikunto, 2013: 25). Teacher frequently uses the former to measure the students' learning outcomes. Teacher can detect to what extent do the outcomes have meet the learning objectives. Consequently, for the test to be able to measure the outcomes correctly, the right making of the test is very necessary. Test will be meaningful if it includes items testing the major objectives and representing the whole materials tested.

It is crucial for test to be a good instrument of evaluation. Arikunto (2013: 57-58) suggested that a good test is one which meets some requirements, such as possessing validity, reliability, objectivity, practicability, and economical aspect. A test is valid when it can measure something correctly. A test is realiable when the results obtained shows accuracy. A test is objective when there is no subjective factors influencing its application. A test is highly practical when it includes several criteria, such as easy to perform, easy to correct, easy to follow. A test is economical when it doesn't involve a lot of expense and is efficient. Arikunto (2013: 57-58) argued that the process of test analysis is necessary to examine the quality of test items.

The analysis of the quality of test items is a series of process should be conducted to find out the degree of the quality of the test, either entirely or partially. The analysis of test items is the investigation of the statements in the test to formulate the high-quality questions (Sudjana, 2013: 135). The aims of the analysis of test items are to identify the items into several categories, namely good, poor, and totally lacking by means of calculating several aspects, such as reliability, the level of difficulty, discrimination index, and distractor efficiency.

Viewed from the use of test in terms of measuring the student's performance, there are three types of test, namely: diagnostic test, formative test, and summative test. Summative test is a test conducted after the program ended (Arikunto, 2013: 38-39). It is moslty conducted in the middle or at the end of the program. The goal of summative test is to assess, to determine the school report, and to fill in student's learning progress (Arikunto, 2013: 39-40). Summative test itself consists of objective and subjective test. Objective test used is in the form of multiple choice which is a test consisting of the incomplete explanation or description of something by means of filling in the incomplete part through selecting one of several provided answers. Subjective test is mostly in the form of essay question, which

is the learning progress test needed the explanatory answer.

Semester final examination is one example of summative test as a means of elearning evaluation in the school. The test item of summative test used to assess the learning outcomes is the Test Items for the First Semester Final Examination on Economics Accounting Subject Grade XII Social at SMA N 1 Kalasan which is teacher-made. In accordance with the interview with the teacher of economic accounting subject on February 3rd 2016, it was found that the quality of the teacher-made test is still unknown. Further, it was also found that teacher, in terms of writing a set of tests, tends to follow the preceding tests. This manner will cause the teacher to overlook the quality of the test from these criteria, such as validity, reliability, the level of difficulty, discrimination index, and distractor efficiency.

The quality of test can be investigated by analysing the test items in order to find out which the test items that should be revised or omitted. The quality of the test will help teacher to get the result of evaluation which is reliable. Further, there are several reasons underlying teachers don't investigate the analysis of test items, namely there is no demand to analyse the test items, they aren't familiar with the steps which are understandabe in analysing the test items, and they don't have enough time. To sum up,

not all teachers are willing to conduct this kind of activity due to the fact it takes long time. In addition to this, there are also many agendas they should do, both in school and out of school.

According to the background of the problems, the writer is motivated to conduct the study which is related to the analysis of test used to evaluate students entitled "The Test Item Analysis of 1st Semester Final Examination on Economics Subject In Grade XII IPS SMA N 1 Kalasan Academic Year 2015/2016."

RESEARCH METHOD
Research Design

This study was descriptive analysis of the document, it used quantitative approach. According to Best in Sukardi (2011: 157) descriptive study is the research method which try to describe and interpret objects according to what it is. The approach used in this study is quantitative approach because all of the data or information obtained are in the form of numerical data and the data are in the form of figures and be analyzed statistical analysis using the program Anates Version 4:09.

Location and Time of The Research

The study was conducted in Grade XII Social SMA N 1 Kalasan started in February 2016, it was from the Final Examination

results which had been done in December 2015.

Research Subjects and Object

The subjects of this study were students of Grade XII Social with the number of 68 students. The object of this study is test items, students' answers and answer keys of the test item analysis of 1st semester final examination on economics subject in grade XII IPS SMA N 1 Kalasan academic year 2015/2016.

Procedure

A test aims to determine the level of learners' mastery towards the materials has been given, besides it is also to determine the level of learners' progress and the suitability between the learning outcomes and the standard of competencies. The test item analysis of 1st semester final examination is used to determine students' understanding in lesson. There are some points to be considered in analyzing the items are validity, reliabilty, discrimination index, the level of difficulty and distractor efficiency.

Items analysis can be used by the students to determine the level of students' ability during the lessons and be able to know which students have understood and which have not. The results obtained from analyzing items are knowing the items which are qualified and which are not. If there are items that are less well may be revised and re-used,

items that are not good can be discarded, while the good items disimpen in the question bank and can be reused. The following chart is a scheme of the frameworks.

Instrument Research

In this study, the techniques of data collection used is the method of documentation. Documentation technique is used to obtain the data of test items, answer key, and answer sheet of 1st semester final examination on economics subject in grade XII IPS SMA N 1 Kalasan academic year 2015/2016.

Data Analysis Teachniques

Analysing the data of the test item analysis of 1st semester final examination on economics subject in grade XII IPS SMA N 1 Kalasan academic year 2015/2016 were by finding the validity, reliability, discrimination index, the level of the difficulty, and distractor efficiency. These criteria are calculated by using computer-assisted through Anates Version 4.09.

a. Validity

According to Sukiman (2012: 178) correlation techniques that can be used for analyzing the validity of these items are point biserial correlation or product moment correlation. Point biserial

correlation index called γ_{pbi}. This correlation formula is as follows:

$$\gamma_{pbi} = \frac{M_p - M_t}{S_t}\sqrt{\frac{p}{q}}$$

Information:

Y_{pbi} : the biserial correlation coefficient

M_p : the average score of the subject answering corretly towards the items to which the validity investigated

M_t : the average of the total score

S_t : deviation standar of the total score

p : the proportion of students answering correctly

$$\left(p = \frac{total\ of\ students\ answering\ correctly}{total\ of\ the\ whole\ students}\right)$$

q : the proporstion of students answering incorrectly

(Suharsimi Arikunto, 2013: 79)

The index of point biserial correlation (γ_{pbi}) obtained from the calculation result consulted with r table of the level of significance 5% according to the number of the students' answer sheets being examined.

b. Reliability

The use of formula K-R 20 is as follows:

$$r_{II} = \left(\frac{n}{n-1}\right)\left(\frac{S^2 - \sum pq}{S^2}\right)$$

Information:

r_{11} = the whole test reliability

P = the proportion of students answering correctly

q = the proportion of students answering incorrectly (q= 1-p)

$\sum pq$ = the total of multiplicaltion between p and q

n = the number of items

S = deviation standard of the test (deviation standard is the deviance root).

(Suharsimi Arikunto, 2013: 100)

Furthermore, in interpreting the test reliability coefficient (r11), it normally uses standard as follows:

1) If r_{11} is equal or bigger than 0.70, it means that the test has high reliability (=reliability).

2) If r_{11} is less than 0.70, it means that the test has low reliability (unreiable).

(Anas Sudijono, 2012: 209)

c. Discrimination Index

According Suharsimi Arikunto (2009: 211) discrimination index is the ability of the test to distinguish between the competent students (high competency students) and the less competent one (low competency students). The formula that can be used to calculate the dicrimination index is as follows.

$$DP = \frac{B_A + B_B}{N_A}\ x\ 100\%$$

DP = the number of testees

B_A = the number of correct answers of the top group

B_B = the number of correct answers of the bottom group

N_A = the number of students from the part of A or B group

(Karno To, 2003: 14)

Further, discrimination index classified as follows:

Table 1. Classification of Discrimination Index

Discrimination Index	Interpretation
Negative - 9%	Very poor, unusable
10% - 19%	Poor, less usable
20% - 29%	Adequate, necessary to revise
30% - 49%	Good
50% and over	Very good

(Karno To, 2003: 14)

d. The Level of the Difficulty

According to Nana Sudjana (2013: 135) the level of difficulty of items is measured from the capability or students ability to answer, it is not measured from the teacher's standpoint as a test maker. The following formula is to determine the level of difficulty:

$$TK = \frac{B_A + B_B}{N_A + N_B} \times 100\%$$

Information:

TK = the level of the difficulty certain test item

BA = the number of students in group A answering correctly

BB = the number of students in group B answering correctly

NA = the number of students in group A (top/competent)

NB = the number of students in group B (bottom/less competent)

Interpretation:

0% - 15% : very difficult

16% - 30% : difficult

31% - 70% : moderate

71% - 85% : easy

86% - 100%   : very easy

(Karno To, 2003: 15)

e. Distractor Efficiency

According to Suharsimi Arikunto (2013: 219) the dispersion pattern of the test answer is the distribution of testees in terms of choosing the answers of multiple choice questions. The pattern of responses about testee obtained by counting the number of choosing choice answers a, b, c, d or who do not choose any (blank). Distractor efficiency can be calculated with the following formula:

$$IP = \frac{P}{(N - B)/(n - 1)} \times 100\%$$

Information:

IP = distractor efficiency

P   = the number of students choosing the distractors

N   = the number of testees

B   = the number of students answering each item correctly

n   = the number of alternative answers (option)

1   = constant

(Zainal Arifin, 201: 279)

These following criteria can be used to interpret the calculation result of each distractor on a test item.

Very good   IP   = 76% - 125%

Good   IP   =51% - 75% or
                        126% - 150%

Deficient IP= 26% - 50% or 151% - 175%

Poor IP   = 0% - 25% or 176% - 200%

Very poor IP= more than 200%

(Zainal Arifin, 2013: 280)

Table 2. Criteria of Distractor Effectiveness Assessment

Distractor functioning	Criteria
0	Very good
1	Good
2	Adequate
3	Poor
4	Very poor

f.   Items Quality

Items that have been analyzed according to each criteria, then analyzed as a whole based on the criteria of the validity, reliability, discrimination index, the level of the difficulty, and distractor efficiency to determine the quality of the items. Determining the quality of the items whether include good, adequate or poor are based on the following considerations:

1) Items were said to have good quality, if the items fulfill four criteria: validity, discrimination index, the level of the difficulty, and distractor efficiency

2) Those items are said to have adequate quality when the items only meet three criteria while one criteria included the category that is not according to predetermined aspect.

3) Those items are said to have poor, if only meet two criteria or less than 2 criteria

RESULTS AND DISCUSSIONS

The findings was obtained from of the test item analysis of 1st semester final examination on economics subject in grade XII IPS SMA N 1 Kalasan academic year 2015/2016 as follows:

Validity

Testing the validity of the test can be done in two ways, by rational (rational validity) and empirical (empirical validity). The rational validity can be determined through the content (content validity). Content validity can be determined by looking at the blueprint whether the test items are in accordance with the indicators to be achieved or not of the test item analysis of 1st semester final examination on economics subject in grade XII IPS SMA N 1 Kalasan academic year 2015/2016 consist of five basic competencies that are concept and international trade policy, analyzing international economic cooperation, describing accounting as an information

system, describing the concept of fundamental accounting equation, accounting cycle services company.

The empirical validity was tested using the formula of biserial point correlation coefficient. The total of students of grade XII Social is 68 students, consisting of 22 students of grade XII Social 1, 22 students of grade XII Social 2, and 24 students of grade XII Social 3. Based on the number of research subjects which are 68 students, it shows the result that is 0.233. The study of the analysis of items validity based on the standard if ≥ 0.233, the item is valid, but if <0.233 the item is not valid. Figures obtained from Anates 0.233 namely at r table at the bottom of which is at number 70 and declared value of 0.233. Based on the analysis of the test item analysis of 1st semester final examination on economics subject in grade XII IPS SMA N 1 Kalasan academic year 2015/2016 , it was known that 26 items was valid (52%) and 24 items was invalid (48%).

Item that is used in an evaluation or test should be a valid question of the validity of the content and the item. Valid items can be stored in the question bank and can be reused in the next achievement test. Invalid items can be repaired with customized indicators and improve the mastery of the technique of preparing items. It can be valid if may be invalid because its construction and include materials that represent the target ukuranya.

Thus the validity is an important thing in determining the characteristics of a good test.

Reliability

The findings of the research toward the analysis of the reliability of the items had the standard if r11 ≥ 0.70 then the item being tested has a high reliability (= reliable), but if r11 <0.70 then the item being tested has low reliability or unreliable (un-reliable). Based on the result of test item analysis of the First Semester Final Examination on Economics Accounting Subject Grade XII Social at SMA N 1 Kalasan in the Academic Year 2015/2016 , it was known that the item has r11 value 0.61 that is greater than 0.70 so that the item had low reliability. An instrument was said to have a high level of reliability if the item has been made having the same results to in a measure to be measured.

Discrimination Index

Classification that is used to interpret the results of the calculation are if   <9% are categorized as poor item, 10% -19% are poor item, 20% -29% are adequate item, 30% - 49% are good item and ≥ 50% are very goog items.

From the results of the analysis, it was found that 10 items (20%) were having very poor discrimination index, 12 items (24%) were having poor discrimination index, 9 items (18%) were having adequate discrimination index), 12 items (24%) were

having good discrimination index, and 7 items (14%) were having very good discrimination index. The distribution of 50 items based on the discrimination index are as follows.

Very good and good discrimination should be preserved and incorporated into the question bank. Fairly good items should be fixed in order to become good or very good items, while items that had poor or very poor criteria should be fixed by paying attention to the reason what caused the failure in terms of discrimination.

1. The Level of the Difficulty

    The criteria used to interpret the results of the level of the difficulty are 0% - 15% classified as very difficult item, 16% - 30% belong to difficult item, 31% - 70% classified as moderate item, 71% - 85% belong to easy item, and 86% - 100% classified as very easy item. From the results of the analysis, it was known that there was 1 item which belongs to very difficult item (2%), 5 items to difficult items (10%), 26 items to moderate items (52%), 8 items to easy items (16%), and 10 items categorized as very easy items (20%).

    Items are included in the category of medium should be maintained, and then inserted into a question bank. Items are classified as difficult to be held improvements by searching for the cause of the item was difficult to be answered

by the students. The cause may come from intent or briefing items that are less clearly so poorly understood by students. Questions must also be held relatively simple improvements such as replacing the phrase with a more complex alternative answers or replace with equivalent posing distinguished with the answer key that will require students to think more

2. Distractor Efficiency

    The pattern of the answer distribution can be determined by calculating from the number of test takers who chose the answer a, b, c, d and e, or who did not choose answer. From the distribution of the answers, it can be known distractor efficiency functioning. Distractor efficiency can be determined through Anates program version 4. Distractors was good if it is chosen by $\geq 5\%$ of the number of learners. In interpreting the distractor efficiency of each items, it used some criteria in the Likert scale as follows:

a. Distractor efficiency was said to be very good if the four distractors functioning.

b. Distractor efficiency was said to be good if there were three distrators functioning.

c. Distractor efficiency was adequated if there were two distractors functioning.

d. Distractor efficiency said poor if there was one distractor functioning.

e. Distractor efficiency was said to be very poor if all the distractors were not functioning.

Based on the analysis of the test item analysis of 1st semester final examination on economics subject in grade XII IPS SMA N 1 Kalasan academic year 2015/2016, it was known that there were 9 items that had a very good quality (18%), 12 items that had a good quality 12 (24%), 14 items that had enough quality (28%), 9 items that had a poor quality (18%), 6 items that had bad qualities (12%).

The items that have very good and good distractors can be reused by putting it to the question bank. Items that were having deficent distractors need to be fixed so it can be good items. Items that were having poor and very poor distractors should be fixed and look for the cause of the failure.

3. Analysis of Test Items viewed Validity, Reliability, Discrimination Index, The Level of the Difficulty, and Distractor Efficiency.

After having analyzed according to each criterion, those items are then analyzed as a whole based on the validity, reliability, discrimination index, the level of the difficulty, and distractor efficiency. The items were analyzed together to determine the overall quality of the test items analysis of the First Semester Final Examination on Economics Accounting Subject Grade XII Social at SMA N 1 Kalasan in the Academic Year 2015/2016. Items quality are divided into five criteria: good, fairly good, and bad based on the following considerations:

a. Items were said to have good quality, if it meets four criteria. If the criterion of validity obtain valid results. When the criteria for distinguishing aspects can meet pretty good, good, and excellent. When the level of difficulty fulfills the criteria of being. When the index criteria humbug or distractor fulfills the very good, good, and pretty.

b. Those items were said to have a fairly good quality, if the issue of the only meet three criteria alone while the assessment of the criteria included in the category that is not in accordance with a predetermined aspect. Aspects that do not meet the criteria in terms of validity, namely invalid. When the distinguishing aspect that is bad and very bad. When the aspect of the level of difficulty is very difficult, difficult, easy, and very easy. When the index

aspect humbug or distractor is not good and not good.

  c. Those items were not qualified to say, a matter of two or more criteria do not meet the specified aspect.

Table 3. The results of the analysis of the test item analysis of 1st semester final examination on economics subject in grade XII IPS SMA N 1 Kalasan academic year 2015/2016 based on the validity, reliability, discrimination index, the level of the difficulty, and distractor efficiency is as follows:

No	Item quality	Items	Total	Persentage
1	Poor	1,3,4,5,6,7,8,9,10,11,13,14,16,17,19,20,21,22,27,28,29,31,33,34,37,42,47,48,50	29	58%
2	Adequate	2,23,24,30,35,44,49	7	14%
3	Good	12,15,18,25,26,32,36,38,39,40,41,43,45,46	14	28%

Sources: Primary data are processed

Table 4. Item Failure Causes Problem

No.	Test Item Failure Cause	Items	Total of Test Items Failure	Percentage
1	Validity (Unvalid)	1,2,3,4,5,6,7,8,9,10,11,13,14,16,17,20,21,22,27,28,30,42,48,50	24	48%
2	Discrimination Index (very poor and poor)	1,3,4,5,6,7,8,9,10,11,13,14,16,17,20,21,22,27,34,42,48,50	22	44%
3	the level of the difficulty (very difficult, difficult, easy, dan very easy)	1,3,5,6,8,10,13,14,19,21,22,23,24,27,29,31,33,34,35,37,42,44,47,49	24	48%
4	Distractor Efficiency (adequate, bad)	1,6,9,14,19,21,27,28,29,31,33,34,37,42,47	15	30%

Sources: Primary data are processed

According to the table, the biggest causes of the failure were on the validity and the level of difficulty in which both have the same percentage that were 48%. Validity means that any items not already have great support or do not have compatibility, the alignment direction of the total score, and the level of difficulty, which means questions used are still too easy so it can not measure the ability of the test taker. Very difficult items can only be answered by a few students while too easy items can be answered by most or all students.

The next factor is discrimination index that were very poor and poor. Poor and very poor discrimination index could not distinguish the students who understood the material and who has not. The cause of the last failure is distractor efficiency which means that the items had distractors index who do not function properly.

Items that had not good quality should not be reused and should be discarded, items that have a poor quality should be improved and look for causes of the failure, then items that have good quality should be put into the question bank that can be reused and modified for next year.

CONCLUSION AND SUGGESTIONS

Conclusion

a. Based on the validity, the results the test item analysis of 1st semester final examination on economics subject in grade XII IPS SMA N 1 Kalasan academic year 2015/2016 valid items amounted to 26 grains (52%) and invalid items amounted to 24 grains (48%).

b. Based on the reliability, results show that the reliability of the test item analysis of 1st semester final examination on economics subject in grade XII IPS SMA N 1 Kalasan academic year 2015/2016 is 0.61.

c. Based on discrimination index, the results of the test item analysis of 1st semester final examination on economics subject in grade XII IPS SMA N 1 Kalasan academic year 2015/2016, it is found that 10 items (20%) were having very good discriminating power, 12 items (24%) were having poor discriminating power, 9 items (18%) were having adequate discriminating power, 12 items (24%) were good discriminating power, 7 items (14%) were having very good discriminating power.

d. Based on the level of difficulty, the test item analysis of 1st semester final examination on economics subject in grade XII IPS SMA N 1 Kalasan academic year 2015/2016, it is known that there was 1 item (2%) was very difficult item, 26 items (52%) were adequate, 8 items were easy and 10 items (20%) were very easy.

e. Based on distractor efficiency, the results of the test item analysis of 1st semester final examination on economics subject in grade XII IPS SMA N 1 Kalasan academic year 2015/2016, 9 items (18%) were having very good quality, 12 items (24%) were having good quality, 9 items (18%) were having deificient quality, 6 items (12%) were having poor quality.

Overall the quality of the test item analysis of 1st semester final examination on economics subject in grade XII IPS SMA N

1 Kalasan academic year 2015/2016 categorised ass 14 items or 28% were good questions which meet the four criteria, seven items or 14% were adequate question which meet three criteria, 29 items or 58% were poor question that meet less than equal to two criteria.

SUGGESTIONS

According to the items analysis in terms of validity, reliability, discriminating power, difficulty level, and distractor index the test item analysis of 1st semester final examination on economics subject in grade XII IPS SMA N 1 Kalasan academic year 2015/2016, the suggestions can be proposed are as follows :

a. Good items can be collected in Questions bank and then reused with modifications in the semester exams next year. Poor items can be revised according to the indicator. Very poor items are unuseable.

b. Tim maker should increase their capability in compiling test item. This was because the test instrument consisted of items that are categorised as good items as many as 14 items or 28%, 7 items or 14 % were adequate with meet three criteria, 29 items or 58% were poor items that meet less than equal to two criteria.

c. The principal needed to provided training to the teachers so that teachers can do good analyse to items going to be used in the test.

d. The developer of Anates program version 4.09 needed to developed the program so that the existing limitations in the program may be eliminated.

REFERENCE

Anas Sudijono. (2012). *Pengantar Evaluasi Pendidikan.* Jakarta: PT. Raja Grafindo Persada.

Daryanto. (2008). *Evaluasi Pendidikan.* Jakarta: PT. Rineka Cipta.

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

Karno To. (2003). *Menganal Analisis Tes Pengantar ke Program Komputer Anates.* Bandung: FIP UPI.

Mimin Haryati. (2007). *Model dan Teknik Penilaian pada Tingkat Satuan Pendidikan.* Jakarta: Gaung Persada.

Muslikah Purwanti. (2014). Anaslisis Butir Soal Ujian Akhir Semester Gasal Mata Pelajaran Akuntansi Keuangan Kelas XI Akuntansi Menggunakan Microsoft Office Excel 2012 di SMK Negeri 1 Yogyakarta. *Skripsi*. Yogyakarta: FE UNY.

Nana Sudjana. (2013). *Penilaian Hasil Proses Belajar Mengajar.* Bandung: PT. Remaja Rosdakarya.

Nanang Martono. (2010). *Metode Penelitian Kuantitatif*. Jakarta: PT Raja Grafindo Persada.

Ngalim Purwanto. (2009). *Prinsip-prinsip dan Teknik Evaluasi Pengajaran*. Bandung: PT Remaja Rosdakarya.

Riska Kurnia Sari. (2014). Analisis Kualitas Butir Soal Ulangan Akhir Semester Gasal Mata Pelajaran Akuntansi Keuangan Kelas XI Akuntansi SMK Muhammadiyah Cangkringan Tahun Ajaran 2014/2015. *Skripsi*. Yogyakarta: FE UNY.

Sony Irawan. (2014). Analisis Butir Soal Ujian Akhir Semester Gasal Mata Pelajaran Komputer Akuntansi Kelas XI SMK YPKK 2 Sleman Tahun Ajaran 2014/2015. *Skripsi.*Yogyakarta: FE UNY.

Sudaryono. (2012). *Dasar-Dasar Evaluasi Pembelajaran*. Yogyakarta: Graha Ilmu.

Suharsimi Arikunto. (2013). *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.

Sukardi. (2011). *Evaluasi Pendidikan Prinsip dan Operasionalnya*. Jakarta: Bumi Akasara.

Sukiman. (2012). *Pengembangan Sistem Evaluasi.* Yogyakarta: Insan Madani.

Undang-Undang Republik Indonesia. (2003). *Undang-Undang Republik Indonesia, Nomor 20, Tahun 2003, tentang Pendidikan.*

Zainal Arifin. (2013). *Evaluasi Pembelajaran*. Bandung: PT Remaja Rosda Karya