Analysis of Multiple Choice Tests for the University Entrance Test

Dr. Moedjadi, M.Pd

Faculty Business & Management, Widyatama University
moedijadi@widyatama.ac.id

Alfiah Hasanah, SE.,MEc
Faculty of Economics, Padjadjaran University
Alfiah.hasanah@fe.unpad.ac.id

The multiple-choice test is one typical test formats used in the university entrance test. Multiple-choice (or fixed-response) format allows for a wider sampling of the content because more questions can be given in a testing period. These types of tests also offer greater efficiency and reliability in scoring than an essay type. In university entrance test the multiple choice is effective test as a tool to screen qualified candidates.

This study aimed to conduct an analysis of entrance examin•tion at Widyatama University. The analysis will be done with items about the validity of the test level, the level of reliability, item difficulty level, distinguishing features and distracters analysis. Research will use the answers to the test wave 2 participants university entrance test for academic year 2009/2010.

Keywords: university entrance test, multiple choice tests, validity test

I.. INTRODUCTION

There are three objectives: first tests on student learning achievement measure (measuring Their achievements) is usually carried out at the end of the learning process (end of course), both in the context of class placement (placement) as well as university entrance examination (university entrance) and third in order evaluation of learning (diagnostic test). With the number of students who are many, efficiency in the implementation process becomes increasingly important test. Therefore, tests using a multiple-choice test (multiple choice test) to be one option.

1.1. University Entrance Test of Widyatama University

In conjunction with the admission of new students Widyatama university carry out entrance test either independently or jointly with other universities. In each academic year the University held an entrance Test with includes two functions: as a prospective student recruitment program and also to observe the academic ability of new and prospective students.

Test entrance examination held in the form of a written test conducted simultaneously (onsite) as many as three waves in each academic year. Form test includes testing of basic skills math, English and Indonesian languages, and mastery of knowledge that includes the Knowledge Economy & Management, Computer Knowledge and Knowledge Technology & Culture. The total number of questions in each wave are 150 questions with multiple choice test. For prospective students faculty & Visual Communication Design (DKV) coupled with psychometric tests and drawing ability.

Since the 2004/2005 academic year university entrance exam Widyatama administer centrally, meaning that not every faculty conduct their own tests with exam questions are different, as previously done. At first entrance examination held each academic year in two waves, but began the academic year 2006/2007 carried out three waves in each year.

1.2. Multiple Choice Test

Test with multiple choice questions (MC) is the most popular and widely used screening tests especially in university entrance. Multiple choice questions are used to test for learning purposes (variety of learning outcomes) start from testing their ability to remember facts test (recall of facts) to testing the high level of Bloom's cognitive skills (Osterlind, 1998). Form of multiple choice test is widely used from pre-recruitment tests, screening tests in universities, competitive exams, certification tests, quizzes and other. In addition the MC test is also widely used as a tool to test the activities of e-learning, distance learning, Preparatory / coaching classes, and research or market survey. The main reason is because the MC test capable of evaluating

the large number of respondents simultaneously and the test results can be obtain in quick time.

Survey and other research support the use of many multiple choice test format as a tool to test the ability. Siegfried, et al. (1996) reported that the average contribution to the question with a question ESSAY only covers about 5-9% from the end of the study, about 20% short answer and multiple choice form of the rest which is about 2 / 3 from the end of the course. Becker & Watts (1996) found that the form of multiple choice tests used by the dominant eye test economics professor. One of the reasons is because of his multiple choice tests most easily done, especially with large number of test takers.

1.3. Research Background

Evaluation of the ability of candidates to answer the questions of etrance test from academic year 2005/2006 to 2007/2008 is shown in table 1. On average, the ability of test test takers in answer to the question of computer knowledge achieved the highest value which is 47.8% while the lowest average is the math test that is only reached around 34.5%.

Table 1. Average University Entrance Test for period 2005/2006 to 2008/2009 (11 waves)

Test	Number of Questions	Average of right answers	%
Mathematics	30	10.3	34.5
English	25	8.7	34.9
Indonesian	25	10.8	43.1
Economics	25	9.7	38.9
Computer	25	12.8	51.2
Culture & Technology	20	9.6	47.8

However test takers responded to the question actually quite heterogeneous in every test. This can be observed from the range of the largest and smallest values obtained in each test.

Based on data entrance test 2007/2008 academic year, the range between minimum and maximum values for each test point is large. This means that the ability of test takers to answer the question is very varied, as we see in the table student can answer only three of mathematics problems, while others could answer 26 questions correctly. Some reasons could be the causes of the problem are the difference of academic background, high school quality, and the region where the student come from. The ability to answer the questiosn correctly could also be influenced by the level of difficulty of the questions.

1.4. PROBLEM IDENTIFICATION

By looking at facts that the level of response to the question of test takers is very heterogeneous this research will focus on the question of test item analysis. Question item analysis is an activity that works to see how the question can measure how well students learn, and be useful to improve the quality of the question itself. Technical analysis of the question item consists of two types namely, qualitative and quantitative. Qualitative analysis techniques encompassed consideration of content validity, while quantitative analysis encompassed the measurement of the quality of the question in the case of item difficulty levels, different forces (discriminating power), and the function of distracters.

Three main characteristics that should be consist in a content of a test which is validity, Clarity, and reliability. Validity is determined by how well the question can measure the level of knowledge, skills and other capabilities that are expected of the learning process. Clarity measured by the quality of questions that represent material that has been learned. While reliability is seen from the consistency of values and the ability to distinguish between test takers with the ability to have different values (Sevenair & Burkett, 1988). Good multiple choice test is generally valid, it is clear (clear) and reliable than the essay type questions for multiple choice test to take a broader sample material, recognize the level of success and consistency of assessment (Cassels & Johnstone, 1984; Russell et al., 2003).

Based on the description above, this research identifies the particular question concerns the quantitative analysis that is as follows:

How is the validity , reliability, discriminating power, difficulty level and function of distracters of multiple

choice items on the entrance test Widyatama University wave 3 2009/2010 academic year compare to wave 1 2010/2011 academic year.

1.5. Purpose of the research

Based on the discussion above the purpose of the research is :

To investigate the validity , reliability, discriminating power, difficulty level and function of distracters of multiple choice items on the entrance test Widyatama University wave 3 2009/2010 academic year compare to wave 1 2010/2011 academic year.

II. Empirical framework

2.1. Analysis of Question Items

This analysis is carry out to to find out how well the item questions can work. The analysis is generally investigate in two ways, namely qualitative analysis (qualitative control) and quantitative analysis (quantitative control). Qualitative analysis is often named as the validity of the logical (logical validity) and performed before the question is being asked and purposed to investigate how well the question can work. Quantitative analysis is often called empirical. validity.

One of the purposes of question item analysis is to find out and improve the quality of the question, namely whether a question:

1. be accepted because it was supported by sufficient statistical data
2. can be improved, as evidenced some weaknesses, or even
3. not used at all because it is empirically proven not work at all.

2.2. Qualitative analysis

Qualitative analysis is viewed in terms of technical, content and editorial. Technical analysis is intended as a review of the question based on the principles of measurement and the format of writing technique. Content analysis specifically intended as an analysis of knowledge relating to the qualification in question. Editorial analysis is intended as an analysis of the particular related to the overall editorial format and regularity of one question to another.

Qualitative analysis can also be categorized in terms of materials, construction and language. The analysis is intended as a review of materials related to the substance in question in terms of knowledge and level appropriate to the ability of the question. Construction analysis is meant as a general review of technical literature related to the question. Language analysis is intended as a review related to the use of language according to legal and formal langaage.

2.3. Quantitative analysis

Quantitative analysis used to determine the extent to which the question can distinguish between the cognitive ability of the test takers. Question of quantitative analysis emphasizes the analysis of internal characteristics of the test through the data obtained empirically. Internal quantitative characteristics include parameters is the question of the level of distress, its judgment and reliability. Specific multiple-choice questions of two additional parameters that is visible from the opportunity to guess or answer a question correctly and functioning least preferred answer, namely the spread of all the alternative answers from the tested subjects.

2.4. DIFFICULTY LEVEL.

There are several reasons for the state level of difficulty of the question. Difficulty level may be determined by the depth of questions asked, complexity, or other matters related to the ability measured by the question. However, when we examine the depth of the level of difficulty of the question, will be difficult to determine why a question can be more difficult than the other.

In general, according to classical theory, the difficulty level can be expressed in several ways among (1) the proportion answered correctly, (2) difficulty in linear scale, (3) Davis index, and (4) bivariat scale. Proportion of correct answers (p), is the number of test takers who answered correctly compared the total number of test takers is the most widely use of level of difficulty. Essentially, the quality of the test item can be observed from the degree of difficulty or hardship status owned by the respective items of the question. Test items can be expressed as good items if the question is not too difficult nor too easy. The figures give an indication of the level of difficulty of the items identified with the term Difficulty index (item difficulty index number), that the world of evaluation study results are generally denoted by the letter P, which is the abbreviation of the word proportion (proportion = proporsa). Categories of questions based on difficulty level Nitko, 1996 is

Category level of difficulty

Value of P	category
P < 0.3	Difficult
0.3 ≤ p ≤ 0.7	medium
P > 0.7	easy

Follow up after category identification are:

Category	Follow up
Difficult	1. Items discarded or dropped and is not issued again on next tests of learning outcomes 2. Are reviewed, tracked and traced so that we can investigate the reason of why the test takers could not answer the question, whether the sentence is not clear, the instructions is difficult to understand, or whether the matter is there are terms that are not clear, and so on. After repairs, the question items can be issued again in the next test. 3. Items that are basically difficult can be reused in the tests especially for a tight selection test.
Medium	These questions item can be issued again in the next test.
Easy	1. Items discarded or dropped and is not issued again on next tests of learning outcomes 2. Are reviewed, tracked and traced so that we can investigate the reason of why the test takers could not answer the question, whether the sentence is not clear, the instructions is difficult to understand, or whether the matter is there are terms that are not clear, and so on. After repairs, the question items can be issued again in the next test. 3. Items that are basically easy can be reused in the tests especially for a loose selection test.

2.5. Discriminating Power

One objective of quantitative analysis is to determine whether a question item can distinguish the different aspect exist between groups. Indices used in distinguishing between the ability of candidates is item discrimination. This index can distinguish between a highly capable candidates from low capable candidates.

The index value ranges from -1 to +1. The negative sign indicates that the low ability test-takers who can answer correctly while the positive sign showed high-ability test takers who answer incorrectly. Thus the matter of distinguishing negative power index shows reversal of quality test takers. Item discrimination indices are generally given the symbol with the letter D (an abbreviation of the discriminatory power).

Index Discriminatory item (D)	classification	Interpretation
< 0,20	*Poor*	The question items has low discriminatory power, considered do not have discriminatory power
0,20 – 0,40	*Satisfactory*	The question items has medium discriminatory power
0,40 – 0,70	*Good*	The question items has good discriminatory power
0,70 – 1,00	*Excellent*	The question items has very good discriminatory power
Negaive sign(-)	-	The question items has very low discriminatory power

2.6. Distracter Function.

The multiple choice objective tests in tests learning outcomes test takers are provided with several possible answers sometimes referred to as options or alternatives. Options or alternatives t ranged from 3 to 5 pieces, and one of them is the correct answers (answer key) while the rest is the wrong answer. The wrong answers one commonly known as the distracters.

Analyzing the function of distracters often known by other terms, namely: analyzing patterns of distribution of answers. This spread pattern is a pattern that can describe how the testee determine the answer between choices available. A possibility exists when the testee not choose any choices or blank answer. Blank statement is often known as omit and given the symbol with the letter O.

Distracters is said to be good if the disctracters are choose at least by 5% of the test takers.

The distraction that perform its function properly can be reused on the next tests, while the others should be repaired or replaced with another distracters.

III. Research Method and Object of the Research

3.1. RESEARCH METHOD

This study intended to determine whether the test questions used to screen prospective students at the Widyatama University is a good tool. We can define test tools as a good tools when the question set have a high validity, high reliability, have discriminating power, have the level of difficulty in accordance with the purpose of tests and the distracters items function effectively.

3.1.1. Validity

Validity is a measure that indicates whether a measure tool actually measure something to be measured, or in other words, a measure tool said to be valid if the measure is measuring something that is to be measured. If the gauge used to measure has high validity it means the data obtained from the measurements have a high validity. There are several types of validity, the validity of which is the content (content validity) and construct validity (construct validity) (Bollen, 1989).

3.1.2. RELIABILITY

Reliability is an index (coefficient) which indicates the extent to which a measuring instrument is reliable or unreliable (Singarimbun, 1995:140). That is, if the gauge is used twice to measure the same symptoms and results of measurements obtained relatively consistent, the gauge is reliable. Understanding that research must be reliable measuring instrument actually means that the instrument good enough to be able to uncover reliable data.

Empirically, high and low reliability shown by a number called the coefficient of reliability. Although theoretically the usual reliability coefficient ranged from 0.00 to 1.00, but in fact the coefficient of 1.00 was never achieved in the measurement. In addition, although the correlation coefficient is positive (+) or negative (-), but in terms of reliability negative coefficient has no meaning, because it always refers to the interpretation of reliability coefficients are positive.

3.1.3. Difficulty level

Question set is said to be good if the question is not too easy or too difficult. Questions that is too easy does not stimulate students to enhance efforts to solve it. Conversely, question that is too difficult can cause students to become desperate and do not have the spirit to try again beyond his reach. Value that indicates difficulty level is called difficulty index. The indext ranges from 0.00 to 1.00. Question with difficulty index of 0.00 indicates too difficult meanwhile difficulty index 1.00 means the question is too easy.

The formula to calculate difficulty index is : $P = \frac{B}{J_s}$

Where:
P = difficulty index
B = number of test takers who answered that question correctly
Js = total number of test takers
In this study, difficulty index is classified following (Arikunto, 1997):
question with P = 0.00 to 0.30 is difficult
Problem with P = 0.30 to 0.70 is medium about being
Problem with P = 0.70 to 1.00 is easy

Questions that are too easy or too difficult does not mean unusable. It will depend on the purpose of the test and the circumstances of the test participants. If we want to have many participants who passed the test, we will used easy question, whereas if we want only selected people to pass then we will use a difficult type question.

Selecting an difficulty level of question item should also consider the answers to guess (guessing). For items with five alternative choices, the guess is 1.00: 5 = 0.20. Therefore, this item should have a level of difficulty of about 0.25 + (1,00-0,20) / 2 = 0.65. Items with difficulty levels between 0.50 -0.90 is a good item. Items that have difficulty approaching the level of 1.00 or below the level of guessing (under 0.20) need to be rewritten or discarded. If possible we should put easier question in front of question and followed by more difficult questions.

3.1.4. DISCRIMINATING POWER

Discriminating power is the ability of question to distinguish between test takers that have low ability and high ability. The index is callled discriminating power (D) whose value is between -1.00 to +1.00.
To calculate the discrimination index, we sort the score of the test takers from the highest to the lowest. We divide the value into two groups. First group is the group that received high test scores (JA), while the bottom group (JB) is the group that received low test scores. The formula used to calculate D following Arikunto, 1997:

$$D = \frac{B_A}{J_A} - \frac{B_B}{J_B} = P_A - P_B$$

Where:
J = Numbers of test takers
J_A = Number of test takers with high score
J_B = Number of test takers with low score
B_A = Number of test takers with high score that answer the question correctly
B_B = Number of test takers with high score that answer the question not correctly
Arikunto classified discriminating power as followed :
D = 0,00-0,20 : poor
D = 0,20-0,40 : medium
D = 0,40-0,70 : good
D = 0,70-1,00 :very good
D = negative : not good. All question item that has negative value should be dropped or discarded.

The investigation will use the software Anatest Version 4 to test the validity, reliability, discriminating power, and difficulty level.

3.2. Object of the Research

The investigation will compare the analysis of the test tool of two waves of the entrance test which is wave 3 of 2009 /2010 academic year and wave 1 of 2010/2011 academic year and.
The investigation will analyze the validity, reliability, discriminating power, difficulty level and function of distracters of multiple choice items on the entrance test Widyatama University wave 3 2009/2010 academic year compare to wave 1 2010/2011 academic year.

IV. Result and Discussion

4.1. Result for Entrance Test wave 1 2010/2011

There are 167 respondent test takers in entrance test wave 1 2010/2011.The total number of questions are 150 consist of 30 questions for mathematics, 25 questions for English, 25 questions for Indonesian, 25 questions for Economics, 25 questions for Computer and 20 questions for culture & technology.

4.1.1. VALIDITY & RELIABILITY

	Mathematic	English	Indonesia	Economics	Computer	Culture & technology
Total Number of question	30	25	25	25	25	20
Number of item valid	28	24	18	25	22	20
Number of item not valid	2	1	7	0	3	0
Number of item reliable	30	25	25	25	25	20

4.1.2.Difficulty level

	Mathematic	English	Indonesia	Economics	Computer	Culture & technology
Number of questions	30	25	25	25	25	20
Items difficult ; P(0,0-0,30)	19	10	9	15	11	5
Items medium ; P(0,30-0,70)	9	13	12	9	9	6
Items easy ; P(0,70-1,0)	2	2	4	1	5	9

4.1.3. DISCRIMINATING POWER

	Mathematic	English	Indonesia	Economics	Computer	Culture & technology
Total number of questions	30	25	25	25	25	20
Items poor ; D(0,0-0,20)	15	4	16	13	12	12
Items medium ; D(0,20-0,40)	11	3	6	10	11	7
Items Good ; D(0,40-0,70)	0	1	1	0	2	1
Items very good ; D(0,70-1,0)	0	0	0	0	0	0
Items not good ; (D negative)	4	17	2	2	0	0

4.1.4. Distracters Analysis

	Mathematic	English	Indonesia	Economics	Computer	Culture & technology
Total number of question	30	25	25	25	25	20
Answer key chosed >80%	0	1	1	1	1	1

Distracter chosed <5%	120	100	95	99	95	78
Distracter not chosed	0	0	3	0	4	2
Answer key chosed < other distracter	38	80	76	79	79	72

4.2. Result for Entrance Test wave 3 2009/2010

There are 109 respondent test takers in entrance test wave 3 2009/2010.The total number of questions are 150 questions consist of 30 questions for mathematics, 25 questions for English, 25 questions for Indonesian, 25 questions for Economics, 25 questions for Computer and 20 questions for culture & technology.

4.2.1. UJI VALIDITY & RELIABILITY

	Math	English	Indonesia	Economics	Computer	Culture & technology
Total Number of question	30	25	25	25	25	20
Number of item valid	18	16	16	14	17	15
Number of item not valid	12	9	9	11	8	5
Number of item reliable	30	25	25	25	25	20

4.1.5. Difficulty level

	Math	English	Indonesia	Economics	Computer	Culture & technology
Number of questions	30	25	25	25	25	20
Items difficult ; P(0,0-0,30)	17	16	9	9	9	7
Items medium ; P(0,30-0,70)	13	9	15	15	10	5
Items easy ; P(0,70-1,0)	0	0	1	1	6	8

4.1.6. DISCRIMINATING POWER

	Math	English	Indonesia	EconS	Computer	Culture & tech
Total number of questions	30	25	25	25	25	20
Items poor ; D(0.0-0.20)	5	4	5	7	9	4
Items medium ; D(0.2-0.4)	15	16	11	13	12	13
Items Good ; D(0.4-0.7)	5	4	5	1	3	2
Items very good ; D(0.7-1.0)	0	0	0	0	0	0
Items not good ; (D negative)	5	1	4	4	1	1

4.1.7. Distracters Analysis

	Math	English	Indonesia	EconS	Computer	Culture & technology
Total number of question	30	25	25	25	25	20
Answer key chosed >80%	0	1	1	1	1	1
Distracter chosed <5%	83	94	96	89	92	82
Distracter not chosed	0	2	2	2	3	2
Answer key chosed < other distracter	40	32	36	47	68	18

4.2. Analysis

Based on the comparation of question test item analysis between two period of entrance test at

4.2. Analysis

Based on the comparation of question test item analysis between two period of entrance test at widyatama university it can be observed that most of the question item is performed well although they have high difficulty level, this can be observed from the validity, reliability . The distracter analysis between those two period not performed well since most of them have value more than 80% choseed <5%. In terms of discriminating power the question item should also be evaluate since the findings showed that some of the in medium level. question item also performed well.

V. Conclusion

The Question test item at widyatama university entrance test showed a good test on overall but its should be evaluate regarding the purpose of the test which is to do the selection process and not to evaluate any learning process.

VI. References

Arikunto, S., (1998).*Prosedur Penelitian. Suatu Pendekatan Pr*aktek, Jakarta: Rineka Cipta.

Bolen, K.A., (1989). *Structural Equation with Latent Variables*, New York: John Wiley & Sons.

Burton Steven J, Sudweeks Richard R, Merrill Paul F, Wood Bud , (1991). *How to Prepare Better Multiple-Choice Test Items:Guidelines for University Faculty*, Brigham Young University Testing services and The Department of Instructional Science
Christina Ballantyne , *Multiple Choice Tests: Test scoring and analysis* ,

www.tlc.murdoch.edu.au/eddev/evaluation/mcq/score.html. Diakses tgl 15 mei 2010

Inmaculada Sanz and Fernández Miguel, (2005). *The University Entrance Exam In Spain: Present Situation And Possible Solutions*, EALTA Conference.

Ebel, R. L., & Frisbie, D. A. (1986). *Essentials of educational measurement* (4th ed.). Englewood, Cliffs, NJ: Prentice-Hall.

Frandsen Gudmund S and Schwartzbach Michael I, *A Singular Choice for Multiple Choice*, http://www.brics.dk/Multiple, diakses tgl 12 mei 2010

Grounlund, N. E. (1982). *Constructing achievement tests* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.
Hopkins, C. D., & Antes, R. L. (1979). *Classroom testing: construction*. Itasca, IL: F. E. Peacock.

Kehoe, Jerard (1995). *Basic item analysis for multiple-choice tests. Practical Assessment, Research & Evaluation*, 4(10). Retrieved February 18, 2010 from http://PAREonline.net/getvn.asp?v=4&n=10 . This paper has been viewed 69,467 times since 11/13/1999.

Moedjadi dan Amelda Banon, *Analisis Soal USM Universitas Widyatama*, 2006, BPPM Universitas Widyatama 2006.

Nitko, Anthony J. (1996). Educational Assessment of Students, Second Edition. Ohio: Merrill an imprint of Prentice Hall Englewood Cliffs.

S. Ramesh & * S. Manjit Sidhu & **G. K. Watugala, *Exploring the Potential of Multiple Choice Questions in Computer-Based Assessment of Student Learning* , Malaysian Online Journal of Instructional Technology ISSN: 1823-1144 Vol. 2, No. 1, April 2005
Slavin, R.E., (2003). *Educational Psychology, Theory and Practice*, 7th. Ed., New York: Pearson Education. Inc.