

THE ANALYSIS OF QUESTIONS ITEM QUALITY IN MATARAM UNIVERSITY

Memy Wardani Elthia ¹⁾, Bidari Andaru Widhi ²⁾
Universitas Bumigora

SUBMISSION TRACK

Submitted : 24 April 2023
Accepted : 10 May 2023
Published : 15 May 2023

KEYWORDS

Test Item Analysis, Level of Difficulty, Distinguishing Power, and Effectiveness of Distractor

CORRESPONDENCE

wardanimemy@gmail.com

ABSTRACT

This study aimed to determine the difficulty level, distinguishing power, and distractor effectiveness of the Final Examination Questions in Grammer Subjects in the third semester of Mataram University Academic Year 2020/2021.

It used descriptive quantitative. The subject of the research is the third semester of the An English Education Department of Mataram University, which are 30 students. The data was collected using the questionnaire method. Then, the data analysis was analyzed by Mrs. Excell.

The analysis shows that: Difficulty levels, the number of multiple choice questions classified as easy questions amounted to 1 item, medium category 25 items, and difficult category 24 items. Then, for distinguishing power included of poor categories are 30 items multiple choices questions, the satisfactory categories are 10 items, and the good category is 0, none categories are 4 items, then negative categories are 5 items. While, for the effectiveness of distractors, 120 distractors are accepted or functioning. Then, 32 distractors are revised or not functioning. Therefore, according to the explanation that has been explained, it can be concluded that the Final Examination for the Grammer subject at University of Mataram in the 2020/2021 academic year is including the quality of the unfavorable questions.

Introduction

The transformation in students is known from the evaluation (assessment) of the learning process. Evaluation is a systematic process to decide the level of success and efficiency of the learning concerned. Evaluation is also the main measuring tool used to determine the ability of students or to absorb the material that has been delivered, whether it is following the objectives curriculum that has been set or not at all. Besides that, Evaluation is a data collection process to determine the extent to which educational objectives have been achieved and how to achieve them (Arikunto, 2021). While, to make a decision about the object to be assessed, an evaluation is needed. According to (Arikunto & Jabar, 2014), the purpose of the measure is to provide a quantitative foundation upon which various examples may be evaluated. Despite the fact that the conclusion that may be formed from an assessment can only be either wonderful or dreadful, brilliant or stupid, high or low (Djaali & Pudjie, 2008) etc. Assessment, as defined by (Mahmudi et al., 2022), is a method by which educators are able to evaluate how effectively their students comprehend a subject while it is being taught to them. An evaluation procedure might be used to compile information on the knowledge, talents, and points of view. On the other hand, assessment refers to the process of utilizing information on real student performance to improve the own teaching (Allen, 2004). Then, any sort of evaluation should be used for the goal of shedding light on the instructional process and providing assistance in the growth of student learning.

Evaluation activities have great benefits in the world of education, as well as in learning activities. Because through evaluation it can be seen the results of the learning activities that have been done, and from these results, it can be determined the follow-up

to be done. Whereas Assessment may be a orderly prepare for deciding or making choices to what degree educating objectives have been accomplished by understudies. Whereas, evaluation is decided as the movement of disentangling the estimation information in step with beyond any doubt criteria or rules (Widiyoko, 2013). Evaluation incorporates all the ways utilized in surveying person execution (Mardapi, 2008). Other than that, (Supardi, 2015) states that evaluation is taking a choice on one thing by relating to certain measures like great or terrible, good or not good, high or low. A strategy commonly utilized for assessment in the test.

A test is a tool used as a means of determining value. The test usually serves to measure and assess students, especially cognitive learning outcomes related to mastery of appropriate learning material. The test is an educational evaluation tool that has a role important in measuring student achievement. In arrange to survey the questions' quality, we must know their legitimacy, reability of the questions, the degree of difficulty of the items, the distinguishing power of the items, the distractor function, and the suitability of the questions with the competency standards and how well they adjust with the competency measures and the candidates' foundational aptitudes. Besides that, the validity, reliability, and objectivity are fair a number of the necessities that must be met some time recently any evaluation may give valuable comes about (Anderson, 2003). On the other hand, the exam is utilized to decide the thing investigation quality of the questions. Checking the quality of enquiries is the objective of thing examination, which is utilized to decide whether things ought to be kept, rejected, or revised. Based on the questions' degree of trouble, separating work, and distractor adequacy, this investigation gives data almost the questions' quality (Muhson, 2015). Finding out which questions are great, destitute, and revolting is what thing appraisal is all approximately (Adawiah, 2020)

The efficiency of the current educational system may be evaluated with the use of a model study that examines the outcomes of tests that were carried out in line with the requirements of the applicable curriculum. The process of evaluating is very important. As a result, the assessment of students' overall performance is a helpful indication of the all quality of the educational experience. If students' accomplishments are adequately assessed, only then will there be room for improvement in the quality of higher education. The knowledge and abilities of the students should be evaluated using a method that can be trusted, and teachers should make the effort to do so. A well-designed test may provide a reasonably accurate reflection of students in all performance. It is possible to determine whether or not a test is of high quality by determining whether or not meet a variety of criteria, both statistical and otherwise. By presenting both aspects, we can then see the quality of the test to decide whether the test is good enough to be used or not. If the manufacturers of testing equipment find that their goods do not meet the criteria set by the industry, they may need to re-evaluate the processes that they use.

In the present study, the researchers want to investigate the questions of item quality at the University level. On the other hand, this research deviates from the standard practice in that it does not restrict its focus to items aimed at middle schools. In light of the aforementioned, I contend that my topic is one that should be studied since it makes use of a variety of research methods and techniques, points of view, and objects of examination.

Based on the background study above, the researcher is interested to formulate the problem of this topic by making some of the research questions, below:

1. How does the questions item quality in the university of Mataram?
2. How do the difficulties of the questions used at university of Mataram?

3. How does the effectiveness of the distractor used at university of Mataram?

Research Method

The subject of the study

The subjects were thirty students of Mataram University in the third Semester of the English Education Program. They consist of 10 males and 20 females.

Research Instrument

The researchers use a questionnaire for data collection. Because this research uses quantitative approaches which are to collect the data from the University of Mataram Students. While, this research used a quantitative because the data will be describes as a figure and analyzed using Microsoft excel to find out the item facility, discriminating power, and the effectiveness of the distractor. The emphasis of the research is on the 50 multiple-choice questions that were on the test given at the conclusion of the semester, and there are 30 students serving as participants. Because it enabled a more comprehensive dissemination of surveys to participants. Google Forms was utilized as the survey tool instead of another option. The emphasis of the research is on the 50 multiple-choice questions that were on the test given at the last of the semester, and there are 30 students serving as participants. The sources of the data are the question papers from the final exam, the answer key sheets, and the student response sheets. Data analysis was conducted on the final test at Mataram University, Academic Year 2020/2021. By finding the item facility, discriminating power, and the effectiveness of the distractor. The questions used in the final test at the English Department of Mataram University are multiple choice questions or objective analyses which are analyzed using test item analysis. The thing that was done first before being analyzed was scoring for each student's answer. The data then analyzed include :

Measuring Item Facility

A number that shows the difficulty or easiness of test items is known as the difficulty index or level of difficulty. The formula used to measure it is:

The Formula:

Which means:

IF = Item Facility (Level of difficulty)

B = number of test-takers answering the item incorrectly

JS = number of test-takers responding to that item

$$IF = \frac{B}{JS}$$

(Brown, 2004:59)

Which means:

IF = Item Facility (Level of difficulty)

B = number of test-takers answering the item incorrectly

JS = number of test-takers responding to that item.

a. Measuring Discrimination Power

Although there is no hard and fast criterion for deciding whether a certain P value indicates that an item should be retained in its current form, altered, or discarded entirely from the test, P values between 0.15 and 0.85 are typical for valid test items (Brown, 2004)

Multiple-choice test questions' ability to differentiate between candidates may be evaluated using the following formula:

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

(Brown, 2004)

Which means:

ID = Item Discrimination (Discrimination Power)

BA = number of top test takers that have correct answer

BB = number of bottom test takers that have correct answer

JA = total participant of top test-takers

JB = total participant of bottom test takers

b. Distractor Efficiency

The distractor efficiency of a multiple-choice question is, together with item discrimination, an important indicator for evaluating the test's overall quality. A good distractor will (a) "lure" a sufficient number of test takers, especially those with less ability, and (b) have answers that are reasonably dispersed among all of the other distractors in the exam. There are ways to determine how effective different types of distractions are.

Data Analysis

After getting the data from data source, the researcher did some steps to analyze the data. In doing the analysis the researcher uses procedures as follow : First, took/make several questions that will be asked to the students in university of Mataram through via Google form. Then, spreading the questions online through the WhatsApp application. After that, collecting and filtering the data that has been filled by students. Then, to measure the Item Facility, Discrimination Power, Distractor Efficiency it were analyzed using Microsoft Excel by following the formulas in each of them.

Result and Discussion

Findings

The results are given in the order of the issues by previous research.

Following is a description of the research's responses to the research questions in previous chapter.

Difficulty Levels

According to the research paper written by (Prihartini et al., 2019) the questions of difficulty level was measured by the number between 0,00 to 1,00. Index 0,00 shows that a question was extremely hard, while index 1,00 shows that a number was extremely easy. This means, the closer a question to number 0 index shows that the question was significantly harder, and the closer it is to number 1, the easier it was.

The interpretation of the difficulty number was:

Less than 0,20: too hard

0,20 – 0,801: Medium

More than 0,80: too easy.

The result of analysis for each question" difficulty level in English Department of Mataram University final test is as follows:

ITEM NO	P	CATEGORY	ITEM NO	P	CATEGORY
1	0,83	Too Easy	26	0,63	Medium

2	0,1	Too Difficult	27	0,26	Medium
3	0,53	Medium	28	0,26	Medium
4	0,33	Medium	29	0,56	Medium
5	0,1	Too Difficult	30	0,73	Medium
6	0,5	Too Difficult	31	0,43	Medium
7	0,7	Too Difficult	32	0,63	Medium
8	0,73	Medium	33	0,6	Too Difficult
9	0,33	Medium	34	0,26	Medium
10	0,76	Medium	35	0,3	Too Difficult
11	0,5	Too Difficult	36	0,2	Too Difficult
12	0,7	Too Difficult	37	0,13	Too Difficult
13	0,66	Medium	38	0,13	Too Difficult
14	0,53	Medium	39	0,7	Too Difficult
15	0,8	Too Difficult	40	0,16	Too Difficult
16	0,73	Medium	41	0,46	Medium
17	0,63	Medium	42	0,53	Medium
18	0,13	Too Difficult	43	0,73	Medium
19	0,2	Too Difficult	44	0,3	Too Difficult
20	0,4	Too Difficult	45	0,23	Medium
21	0,4	Too Difficult	46	0,6	Too Difficult
22	0,8	Too Difficult	47	0,43	Medium
23	0,63	Medium	48	0,63	Medium
24	0,4	Too Difficult	49	0,46	Medium
25	0,13	Too Difficult	50	0,5	Too Difficult

From the table above, there is a question item (number 1) that fall to the low difficulty category. Then, there are 25 questions item (number 3,4,8,9,10, 13,14,16,17,23,26,27,28,29,30,31,32,34,42,42,43,45,47,48,49) belongs to Medium Difficulty Category. Besides that, there are 24 questions item (number 2,5,6,7,11,12,15,18,19,20,21,22,24,25,33,35,36,37,38,39,40,44,46,50) belongs to High Difficult Category.

Most of the questions with a moderate/medium difficulty level filled half of the questions with 25 items of the total questions. This can be considered a good thing because the test is balanced. Therefore, It needs to be revised is only the remaining questions with a low difficulty level (1 question) and the questions with a high difficulty level (24 questions).

Item Discrimination:

Division utilized to clarify the whole comes about of recognizing control is 0.00 to 0.19 that are include within the destitute category; 0.20 to 0.39 are included within the palatable category; 0.40 to 0.69 include in good categories; and 0.70 to 1.00 are include within the fabulous category .

NO	Distinguishing Power	Questions Number	Total
1.	0,00-0,19 (Poor)	1,2,3,6,7,8,9,10,14 16,20,22,23,25,27,28,29, 30,31,35,36,39,41,	30

		43,44,45,47,48,49,50	
2.	0,20-0,39 (Satisfactory)	11,12,13,15,17,24,26,32,33, 46	10
3.	0,40-0,69 (Good)	-	-
4.	0,70-1,00 (Excellent)	18	1
5.	0 (None)	4,21,34,42	4
6.	Negative	5,19,37,38,40	5

So, the results indicate the 30 questions of multiple choice with poor distinguishing power, 10 questions with satisfactory, 0 term is a good category, and excellent amount of 1 item, none category is 4 items and negative category is 5 items. Based on the explanation above, the grammatical questions that will be on the final exam for the English Department at Mataram University for the third semester of the 2020/21 academic year may be split into two categories: unacceptable and acceptable. There are a total of fifty questions with a multiple-choice format, and thirty percent of those questions belong to the "unfavorable" category, while just ten percent belong to the "satisfactory" group. According to this, it seems that the multiple-choice questions are unable to differentiate between students with high and poor levels of achievement.

According to (Sudjiono, 2011), in a theoretical study, defines "distinguishing power" as the capability of an achievement test item to discriminate between high-ability and low-ability test takers. The results of this research corroborate the conclusions of Anas Sudijono's theoretical review.

The final exam for the third semester of the 2020/21 academic year offered by the English Department of Mataram University has awful, adequate, and negative categories of multiple-choice questions. As a result, the test has a powerful discriminating power. As a consequence of this things that are bad items need to be discarded, while those that are not excellent yet need to be looked at again and maybe have the factors that led to their development investigated.

Effectiveness of Distractor

NO	Key Answer	DistractorEfficiency	
		ACC	REV
1	B	C	A,D
2	B	C	A,D
3	A	B,D	C
4	A	B,C,D	-
5	D	A,C	B
6	C	A,B,D	-
7	D	B,C	A
8	A	B,C	D
9	A	B,C,D	-
10	D	A,C	B

11	C	A,D	B
12	A	B,C,D	-
13	A	B	C,D
14	A	B,C,D	-
15	C	A,B	D
16	C	A,B	D
17	D	A,B,C	-
18	D	B,C	A
19	A	B,D	C
20	D	A,B,C	-
21	D	A,B,C	-
22	C	A	B,D
23	A	B,D	C
24	A	B,C	D
25	B	A,D	C
26	A	B,C	D
27	C	A,B,D	-
28	B	A,C,D	-
29	A	B,C,D	-
30	B	A,D	C
31	B	A,C,D	-
32	B	C,D	A
33	B	A,C,D	-
34	B	C,D	A
35	B	A,C,D	-
36	B	A,C,D	-
37	B	A,C,D	-
38	B	A,C,D	-
39	A	B,C	D
40	A	B,C,D	-
41	C	B,D	A
42	B	A,C	D
43	A	B,C,D	-
44	C	A,B	D
45	A	B,C,D	-
46	B	A,D	C
47	C	A,D	B
48	D	A,B	C
49	A	B,C,D	-
50	B	A,C,D	-

The data above shows that from 152 distractor, there are 120 distractor that accepted or functioning. Then, there are 32 distractor that revised or not functioning.

Discussion

This part provides an explanation of the results and analyses of the research that were provided in the section that came before this one. In addition, the research problems that were discussed in the previous chapter that will addressed in this chapter.

The order of the degrees of difficulty is as follows. First, statistic that is used to evaluate the difficulty of the question is the number of correct replies expressed as a percentage of the total number of test participants. The good rated of the students if they are neither too simple nor easy, but good subjects are then rated to be average.

The easiest questions do not encourage students to decide the questions. While, the difficultiest questions make students despair and then they do not have the strength to do it again because they are beyond the reach of articles that are regarded as having a medium quality that are given ratings ranging from 0.31 to 0.70.

According to the findings of the research, the number with multiple choice questions had a total of 24 questions with a tough level of difficulty, 25 questions with a medium category items, and 1 question with an easy level of difficulty. It will then be possible to utilize the question bank as a tool for assessment in the future, and items from the medium category will be able to be added to it. If a question is determined to be either too easy or too difficult, it may be evaluated to see why it was placed in that category. If necessary, the question may then be revised before being re-tested.

Distinguishing power. The Determination Index measures the degree to which different participants in the study may be differentiated from one another on the basis of the outcomes that they achieved on these questions, whether from the highest group to the lowest. Scores between 0.00 and 0.19 are considered to be in the poor category; scores between 0.20 and 0.39 are seen as being in the fair category; scores between 0.40 and 0.69 are regarded as being in the good category; and scores between 0.70 and 1.00 are regarded as being in the excellent category.

According to the findings of the research, thirty (30) items of the multiple-choice questions had bad categories, ten (10) items had fair categories, zero (0) item had good categories, and five (5) items had negative categories.

Therefore, based on the explanation above that the grammar subject of final exam in the third semester of the 2020/21 academic year is rated as low and adequate. The quality of the answers to 30 of the total of 50 multiple-choice questions is assessed as "low," "inadequate," or "very poor," while the quality of the answers to the remaining 10 questions is rated as "adequate" or "high." This suggests that the multiple-choice questions are not particularly successful at sorting students into high achievers and poor achievers because the multiple choice questions it's still in the low category

Last, Effectiveness of Distractor. The result shows showed that, there are 120 distractor that accepted or functioning. Then, there are 32 distractor that revised or not functioning.

The function of the distractor, according to (Sudijono, 2009) that a distractor that has been able to perform its function properly can be used again in future tests, while a distractor that has not been able to function properly should be repaired or replaced with another distractor.

Based on the description above it concluded that the Final Examination for Grammar subject at University of Mataram in the 2020/2021 academic year is including the unfavorable questions.

Conclusion

Based on data analysis from the results of the research carried out, it can be concluded that the odd semester exam questions in Grammar subject at Mataram University for the 2020/2021 academic year:

1. Seen from the level of difficulty, overall it is categorized as moderate questions because the highest percentage of questions is in the medium question category.

2. When viewed from the index of distinguishing power of the questions, overall it is categorized as poor or not good because the highest percentage of questions is in the category of poor questions.
3. In terms of the distractor function, as a whole, seen from the results of the students' answers, it is categorized as having a distractor that has been functioning properly.

Suggestions:

Based on the overall results of the research studies that have been carried out, the researcher convey some suggestions to the parties concerned as follows:

1. For the university
It is advisable to hold a workshop for lecturers with the theme of arranging questions according to the rules of good questions, so that they can add insight to the lecturers in making questions and applying them in making the next question.
2. For Lecturer
The products' quality is up to question. The questions that don't fall into the category will either be removed or rephrased at this point. In addition, it has to get rid of anything that has a chance of failing the subsequent examination..
While, It is necessary to coordinate together in the preparation of question grids, question assembly and question analysis in order to achieve question quality. Besides that, the lecturer should validate every question that they made before the questions are given to the students.
3. For students
They have to study harder, because there are still some materials that need to be improved in understanding.
4. For Further Researcher
It has been suggested that this study might be used as a reference for future researchers, especially those individuals who are experiencing the same difficulties and are enthusiastic about doing research. Changes in both subject matter and methodology might help researchers dig deeper into their topics. The author of the study has expressed confidence that it will be of service to researchers in the years to come.

References

The bibliography can be sourced from various references in textbooks, textbooks summarized by editors, translation books, thesis/thesis/dissertation, journals, proceedings, and newspapers. The bibliography is written by following the style of Journey, as stated in the Guidelines of this journal (which ratified APA IV Edition).

References are written in a single space, between libraries given an area of 1 room, as an example of how to write references/references in the Bibliography, given below.

a. Example from book citation

Gronlund, N.E. & Linn, R.L. (1990). *Measurement and evaluation in teaching*(6thed.). New York: Macmillan.

b. From textbooks summarized by the editor:

Effendi, S. (1982). Unsur-unsur Penelitian Ilmiah. Dalam Masri Singarimbun (Ed.). *Metode Penelitian Survei*. Jakarta: LP3ES.

c. From translation book:

Daniel, W.W. (1980). *Statistika Nonparametrik Terapan*. (Terjemahan Tri Kuntjoro). Jakarta: Gramedia.

d. From skripsi/thesis/dissertation:

Suyanto, S. (2009). Keberhasilan Sekolah dalam Ujian Nasional Ditinjau dari Organisasi Belajar. *Disertasi*, tidak dipublikasikan. Universitas Negeri Jakarta.

e. From journal article:

Grandey, A.A., Fisk, G.M., & Steiner, D.D. (2005). Must “Service With a Smile” Be Successful? The Moderating Role of Personal Control for American and French employees. *Journal of Applied Psychology*, 90 (1), 893-904.

f. From providing article:

Paidi. (2008). Urgensi Pengembangan Kemampuan Pemecahan Masalah dan Metakognitif Siswa SMA Melalui Pembelajaran Biologi. *Prosiding, Seminar dan Musyawarah Nasional MIPA yang Diselenggarakan Oleh FMIPA UNY, Tanggal 30 Mei 2008*. Yogyakarta: Universitas Negeri Yogyakarta.

g. From the news article:

Hartiningsih, M. (2015). Indikator Kemajuan: Mengejar “Kebahagiaan”. *Kompas*, 23 Februari 2015.