ELT IN A GLOBALIZED WORLD: THE BOUNDARIES AND BEYOND

Item Analysis of English Test of National-Based School Examination (USBN) in Serang Regency

Arnilah

SMP SMA Islam Nurul Fikri Boarding School Serang, Indonesia

Abstract

Abstract: This study aimed to analyze the English Test of the National-Based School Examination (USBN) in Serang Regency held in 2021. The analysis consisted of Difficulty Level, Discrimination Power and Distracter Effectiveness. The intention is to know the quality of test items and the whole test as well. The method used is a descriptive quantitative approach. The participants were 128 students of SMP Islam Nurul Fikri Boarding School. From the research results, it is found that the level of difficulty of the test is not balanced. The easy level is 56%, the moderate is 28%, and the high is 16%. Meanwhile, From the Discrimination index found that 64% are reasonably good, 16% are marginal, and 20% are rejected. From the distracters' point, 55% of options work well, while 44% do not function, and 100% of key answers work correctly. It is concluded that the test needs some revisions and improvements.

*Keyword*s: Item analysis, Item Difficulty, Discrimination Power, Distracter Effectiveness, National-Based School Examination

1. Introduction

Before 2020, the National Examination (UN) and National-Based School Examination (USBN) were used to decide the students graduated from the level of Elementary School (SD), Junior High School (SMP), and Senior High School (SMA). However, the government abolished those two tests by the government policy or Permendikbud 43 the year 2019. The students' graduation is determined by each school.

Thus, each school has its right to decide the criteria of the graduation. There are three main essential criteria of graduation usually used: characteristics, following the learning program at school, and passing the exam held by the school. However, the policymakers in Serang Regency thought that they needed to hold the National-Based School Examination (USBN) for the whole schools in in its region.

The intention of applying that standardized test is to determine the quality of education. The quality of the examination or the test held gives feedbacks about the condition of education. The quality of any assessment in any educational setting results from the "quality of the instruments" the test administrators use as a basis of decision making (Anderson & Morgan 2008, as cited in (Fiktorius, 2014)

Moreover, Butterfield stated that examination fulfills various functions, including the measurement of attainment, accountability (of an institution, or region, curricular definition, and student motivation as cited in (Mc Alpine, 2002). This condition aligns with the concept of evaluation. Evaluation is the process of making decisions based on assessment results as stated in the Minister of Education Policy, Permendiknas 2014.

Regarding that purpose, the government of Serang Regency, through English Teacher Working Group (MGMP), chose a group of qualified teachers to design the English test for the National-Based School Examination (USBN). They are considered senior teachers, certified and have attained the training of designing tests. They made 50 Multiple Choices Questions (MCQ)

According to Swanson et al. (2006), multiple-choice questions are globally the most utilized among the different types of students' learning achievements and progress, as cited in (Anna Siri, 2011). The item analysis of multiple choice questions (MCQs) is an essential tool that can provide input on validity and reliability of items. It helps to identify items which can be revised or discarded, thus building a quality MCQ bank.

Considering the importance of the test in the teaching process, teachers have to make sure that they construct a good quality test. Teachers or test makers have to ensure that test results scores accurately reflect an examinee's ability in a specific area (Weir, 2005). One of the ways of ensuring the quality of a test is by conducting an item analysis.

Item analysis examines students' responses to each test item done to measure the quality of the test items. Item analysis is a process of checking and analyzing the quality of each item by sorting out the good items from the weak ones. Thus, revised to become better ones. Item analysis as a process based on specific procedures and steps to identify which test items are practical and have good quality to use as a tool of assessment (Brown, 1996) and (Musial, 2009)

There were a lot of researches done in item analysis. Some of them were conducted by Hartati (2019), AV Maharani (2020) and R Jannah (2021). Most studies conducted the school level test, except for ES Nurbaeti (2019) that analyzed the national examination. Those studies reveal that the Indonesian teachers are rarely did item analysis on their test. However, this present study is about to re-raise awareness of teachers on the importance of conducting item analysis.

This research intends to conduct an item analysis of the English test of National-Based School Examination (USBN) in Serang Regency in 2021. Based on the information although that standardized test has been done for years by government of Serang Regency through the English Teacher Working Group (MGMP), they have never conducted any researches on the item analysis. Therefore, they have not known the quality of the test items

Thus, noticing the importance of the feedback of the test as one way to measure the quality of education in Serang Regency, an item analysis should be conducted. This present research is to examine the quality of the English test of National-Based School Examination (USBN) in Junior High School in

Serang Regency in 2021. The analysis is conducted to get the empirical evidence on the difficulty level, the discriminating power, and the effectiveness of distracters of the National-Based School Examination (USBN).

There is one formulated research questions for this study, that is:

What are the characteristics of the difficulty level, the discriminating power, and the effectiveness of the distractors of the National-Based School Examination (USBN) in Serang Regency?

The procedures of conducting item analysis in this study involved three kinds of analysis; the analysis of the difficulty level, the discriminating power, and the effectiveness of distractors.

The Analysis of Item Facility or Difficulty Level

Item Facility is the proportion of test takers who answer an item correctly (Fulcher & Davidson, 2007). Therefore, a good test should have a varied difficulty index, which consists of easy, moderate, or difficult. Thus, it is suggested that a good test should have a ratio of 1: 2: 1 for its easy, moderate, and difficult items.

Moreover, he stated that items should not be too easy or too difficult for the population for whom the test has been designed. Items with facility values around 0.5 are therefore considered to be ideal, with an acceptable range being from around 0.3 to 0.7 (Henning, 1987: 50), as cited in (Fulcher & Davidson, 2007)

However, we should note that the proportion correct is actually dependent not only on the difficulty of the item itself but on the ability of the test takers who are used in calculating the value. Technically, this is known as the sample dependence of the statistic. Another way to put this is that with a different sample of test takers, the value could be different.

Analysis on Item Discrimination

If the test and an item measure the same ability or competence, we would expect those with a high overall test score to have a high probability of answering the items. We would also expect the opposite: those with low test scores would have a low possibility of correctly answering the item. Item discrimination is a measure of whether an item discriminated between students who knew the answer and students who did not (Arora, 2018)

Thus, a good test item should distinguish between the high level and the low-level students. Usually, two ways of determining the discriminative power of an item are used: the discrimination index and the discrimination coefficient.

Analysis on Distracter Effectiveness

The quality of the item's distracters is essential to know the quality of the test. Nevertheless, neither the item difficulty nor the item discrimination index considers the performance of incorrect response options or distracters. A distracter analysis addresses the interpretation of these incorrect response options. Just as the key, or correct response option, must be definitively correct, the distracters must be false (or not the "best" option). In other words, Arora (2018) stated that Distracter Effectiveness

is Determination of whether distractors (incorrect but plausible answers) tend to be marked by the less able students and not by the intelligent

In addition to being incorrect, the distracters must also be plausible. That is, the distracters should seem likely or reasonable to an examinee who is not sufficiently knowledgeable in the content area. If a distracter appears so unlikely that almost no examinee will select it, it is not contributing to the item's performance.

The presence of one or more implausible distracters in a multiple-choice item can make the item artificially far more accessible than it ought to be. In a simple approach to distracter analysis, the proportion of examinees who selected each response option is examined. For the key, this proportion is equivalent to the item p-value or difficulty. If the proportions are summed, they will add up to 1.0, or 100% of the examinees' selections.

A distracter analysis can also reveal an implausible distracter. Usually, when the item p-values are typically high, the proportions of examinees selecting all the distracters are, as a result, low. Nevertheless, if examinees consistently fail to choose a given distracter, this may be evidence that the distracter is implausible or simply too easy.

Methods

The method used is in this research is the descriptive quantitative method. Thus, statistical analysis was used to describe the condition of research variables. Then, the gained scored function to describe the condition of the National-Based Test Examination (USBN). The description was based on the table of interpretations of each element. The result will be in the form of percentages.

The participants in this study were 128 students, both male and female students of 9 graders of SMP Islam Nurul Fikri Boarding School Serang academic year of 2020/2021.

Arora (2018) suggested that these are the procedures to do item analysis: (1) Arrange the test papers in rank order from the highest score to the lowest score. (2) Select one third of the papers with high score and call them the higher group. (3) Select another one third of the papers with low scores and call them lower group. (4) Set aside the middle third. (This will not be used in analysis) (5) Prepare a frequency table by counting the number of students in the higher group who selected each alternative. Similarly count for the lower group.

Results and Discussion

Result on Item Facility

There were 128 answer sheets from the students as the data source, and the English Test of National-Based examination (USBN) consists of 50 items of objective questions in the form of Multiple-Choice Questions (MCQ). Based on the data source, the writer took the scores of the students and analyzed difficulty level or item facility (p) by determining the proportion between the number of students who answered correctly and the total number who took the tests.

To categorize the IF, the writer uses a classification of the difficulty level:

Table 1: difficulty level Index

Index	Interpretation
0.0 – 0.30	Difficult
0.31 – 0.70	Moderate
0.71 – 1.00	Easy

Then, the writer did the calculation of IF:

$$p = \frac{N_{correct}}{N_{total}}$$

The item facility of the whole test was found by calculating the average of item facility or level difficulty.

$$P = \frac{\sum p}{N_{total}}$$

Based on the procedure analysis above, the writer got the result as in table 2.

Table 2 : statistical result of item facility

Interpretation	Item Number	Percentage	Question Number													
Difficult	8	16%	3	6	13	31	40	42	47	48						
Moderate	14	28%	4	5	10	12	14	22	35	36	37	39	41	44	46	49
Easy	28	56%	1	2	7	8	9	11	15	16	17	18	19	20	21	23
			24	25	26	27	28	29	30	32	33	34	38	43	45	50
Total	50	100%														

From table 2 above, we found out that the difficulty level is not balanced.

The level of acceptance is shown in table 3.

Table 3: Level of acceptance

Coefficient Range of Levels	Interpretations	Acceptable Percentages
0.01-0.10	More Difficult	10%
0.11-0.29	Difficult	20%
0.30 - 0.70	Moderate	40%
0.71-0.89	Easy	20%
0.90 - 0.99	Easier	10%

Based on table 3, the difficulty level is not balanced. The difficult and more difficult should be 30%. However, the result showed only 16%. Besides that, the moderate level should be 40%.

Nevertheless, the result is 28%. Moreover, the easy and easier should be 30%. Meanwhile, the result is too many that is 56%. The conclusion is that the level of difficulty should be balanced. They should be increased to more difficult and difficult level. Thus, the easy level should be reduced.

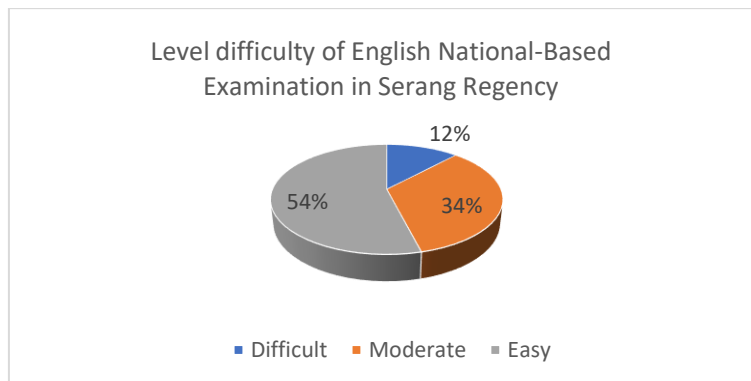From the figure 1 above, we found out that the difficulty level is not balanced.

Level difficulty of English National-Based Examination in Serang Regency

12%

54%

34%

■ Difficult ■ Moderate ■ Easy

Figure 1 : Chart of Item Facility of English Test of National-Based School Examination (USBN)

Item Discrimination/Discrimination Power

The discrimination power refers to a more detailed analysis of the test item the item facility. It shows how the top scorers and the low scorers perform on each item.

Item Discriminations were analyzed statistically through a formula that is

$$D_i = \frac{GA_{correct\ answers} - GB_{correct\ answers}}{N_{largest\ group}}$$

where:

D_i	=	Discrimination index of item i
$GA_{correct\ answers}$	=	Number of correct answers to item i among the 27% of those with highest test scores.
$GB_{correct\ answers}$	=	Number of correct answers to item i among the 27% of those with lowest test scores.
$N_{largest\ group}$	=	Number of person in the largest group (GA or GB)

Table 4 : Item discrimination Index

Coefficient of ID	Interpretation
0.4-1.00	Very good
0.30-0,39	Reasonably good
0,2-0.29	Marginal
Below 0,19	rejected

From the calculation of Item discrimination, we found that :

Table 5 : The result of Item Discrimination

Coefficient of ID	Interpretation	Results	Percentage
0.4-1.00	Very good	0	0%
0.30-0,39	Reasonably good	32	64%
0,2-0.29	Marginal	8	16%
Below 0,2	rejected	10	20%

From the table above, we can summarize that 64% of discrimination power is reasonably good. Meanwhile, 16% are marginal, and 20% are rejected. Unfortunately, none of the test items could be used directly as the test bank in the future, as all the coefficient of discrimination powers are less than 4%. The number of the rejected Discrimination power is quite high, 20%. Thus, the test developer should consider this condition.
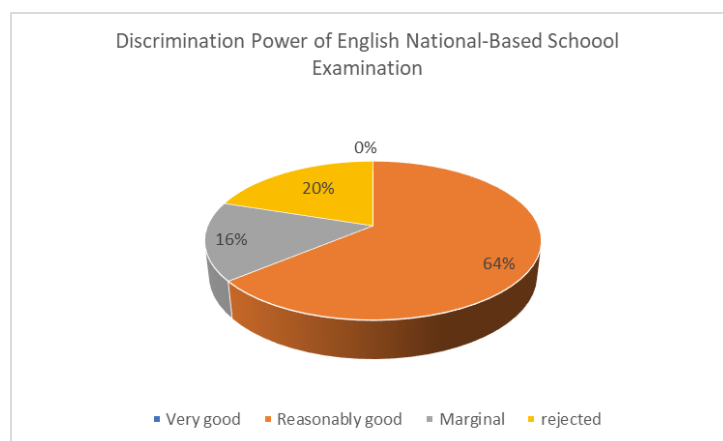

Figure 2 : Discrimination power of English National- Based School Examination

Distracter Analysis

Distracter Analysis aims to investigate which distracters are functioning as they should do ad which is not. It was analyzed by comparing the answering correctly to students into High, Middle, and Low groups. It was analyzed logically based on the number of choices on the options.

Thus, since the test consists of 50 Multiple Choice Questions with four options (a, b, c, d), it meant there were 200 options. Among those, 150 options (75%) were the distracters, while 50 (25%) were the key answer. Thus, we analyze the distracters. Table 6 shows the statistical analysis results all option.

Table 6 : the statistical analysis results all option

-	+	*	*-	Total
67	83	50	0	200
44%	55%	100%		100%

The meaning of the symbols are:

: Distracter is failed or it does not function

+: Distracter functions

* :   Key answer

*- : Failed key answer

When it was separated between the key answers and the distracters analysis, the percentage will be in table 7 and 8.

Table 7 : Statistical analysis result of distracters

-	+	Total
67	83	150
44%	55%	100%

Table 8: statistical analysis results of key answer

*	*-		Total
50	0		50
100%			100%

From Tables 7 and 8 we found that the function distracters were 55%. Meanwhile, the do not function, or 0% students chose 10 or 7%. Distracters less than 5% or less function were 57 or 38%. Thus, we consider the do not function and the less function into do not function distracters were 67 or 44%. Fortunately, 100% of key answers work correctly. We can see the result of distracter analysis in the following figures.
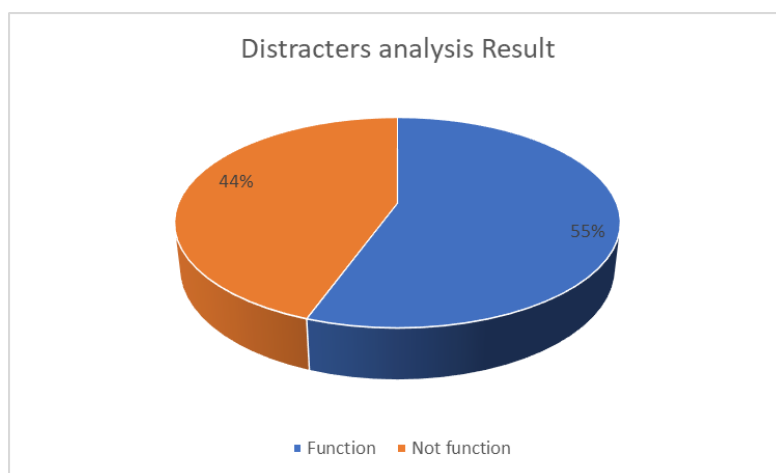

Figure 3 : Distracter Analysis of English National-Based School Examination

[1] Corresponding author.
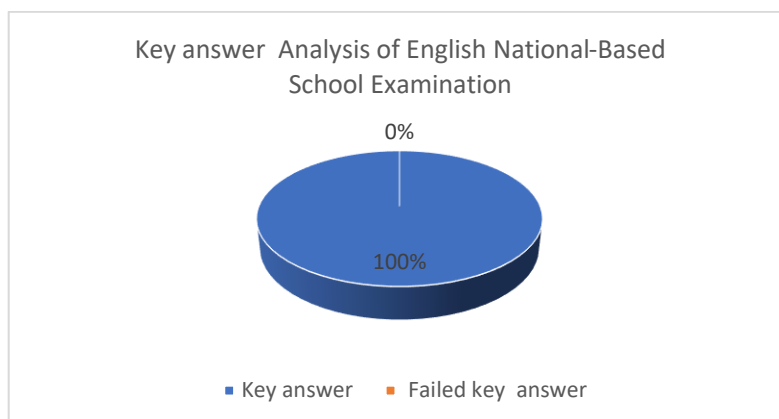*E-mail address*: author@affiliation.xx.xx

Figure 4 Key answer Analysis of English National-Based Schoolt Examination

Conclusion

From the research results, it is found that the level of difficulty of the test is not balanced. The easy level is 56%, the moderate is 28%, and the high is 16%. Meanwhile, From the Discrimination index found that 64% are reasonably good, 16% are marginal, and 20% are rejected.   From the distracters' point, 55% of options work well, while 44% do not function, and 100% of key answers work correctly. Hence, the test developers should consider revising the test of the balance of difficulty level and low item discrimination in the future. Besides that, it is essential to design the distracters so that they work well.

References

Anna Siri, M. (2011). *The Use of Item Analysis for the Improvement of Objective Examination*. Elsevier Journal Procedia Social and Behavioral Science.

Arora, Srimiti. (2018). *Item Analysis*. International journal of Nursing Practice and Research Vol.2 Issue.4

Brown, J. D. (1996). *Testing Language Program*. New Jersey: Prentice-Hall Regent.

Fiktorius, T. (2014). *A Validation Study on National English Examination of Junior High School in Indonesia.* Jurnal Pendidikan dan Pembelajaran Khatulistiwa.

Fulcher, Glen &Davidson, Fred. *Language Testing and Assessmen*t. (2007). New York: Routledge.

McAlpine, M. (2002). *A Summary of Method of item Analysis*. Glasgow: CAA Center.

Musial, D., Nieminien, G., Thomas, J., & Burke, K. (2009). *Foundation of Meaningful Educational Assessment*. New York: McGraw-Hill Higher.

Weir, C. (2005). *Language Testing and Validation: An Evidence-Based Approa*ch (C.N Candlin&D.R Hall, Eds). New York: Palgrave Macmillan.