Research Article

The Nature of an English Test: Dificulty Index and Distractor Analysis

*1Rusma Setiyana, 2Endah Anisa Rahma. 3Rina Syafitri

*1Universitas Teuku Umar, Indonesia, Email: rusmasetiyana@utu.ac.id
2Universitas Teuku Umar, Indonesia, Email: endahanisarahma@utu.ac.id
2Universitas Teuku Umar, Indonesia, Email: rinasafitri@utu.ac.id

How to cite this article: Setiyana, S., Rahma, E.A., & Syafitri, R. (2020). The nature of English test: Difficulty index and distractor analysis. *IJELR: International Journal of Education, Language, and Religion, 2*(2), 77-84. http://doi.org/10.35308/ijelr.v2i2.2781

Abstract
Language assessment plays such a vital role in English teaching and learning process to measure students' ability. Doing mistake in creating test items leads the failure in finding out presumption of the student's achievement. This research was expected to break down the nature of English test including the difficulty index and the adequacy of distractors as a part of an English language assessment. Content analysis was utilized in this study by using two different techniques, namely check list and document analysis to collect the information data. The information from checklist were investigated by utilizing a factual method in percentage and the information from archive examination were analyzed by using a test item analysis software. The result demonstrates that in the term of difficulty index, the English item test met 48% to be acknowledgment and the percentage of distractor effectiveness of English test was 41%. It means that more than 50% of the test should be revised and removed.

Keywords
English test; difficulty index; distractor; language assessment

Introduction

Regarding the fact that English is seen as a device of overall correspondence in revealing numerous communicative expressions, lecturers are the ones who significantly have the greatest commitment and obligation in teaching it, along with the results and evaluation. This is considered in such a way since English is one of the mandatory subjects in the Indonesian national curriculum despite its status as a foreign language in this nation. The real point of English educating in Indonesia is to create relational abilities, both oral and written skills such as listening, speaking, reading, and writing. So, the teaching process of this school subject is shouldered to the teachers at school without excluding the assessment process which measures the students' dominance in English.

Evaluation has been assumed as an essential part in the instructing and learning process. It is utilized to recognize how students learn and understand the materials given at their class. Numerous sorts of appraisal are utilized by instructors in the university as an unquestionable requirement to be constantly used in measuring their accomplishment in achieving the teaching objectives. It is as well as whether the students are able to continue the following stage to achieve next destinations of learning. Moreover, during pandemy Covid-19 all educators needs to think about the nature of a test while making a

compelling test so that students can also enhance their ability in absorbing the materials that have been taught in their online classes.

Multiple choices test is probably one of the test used in an online class. To cerate this test, sometimes it takes to long time because a conductor should master how to create effective distractors. Eventhough this test is difficult to make, many departments or institutions choose this type of test as it permits direct measurement. It may be used for multi-disiplines study and to measure wide ranges of content domain in a short time (Haladyna & Rodriguez, 2013; Rodriguez, 2016; Gierl et al, 2017). Furthermore, multiple choice test usually provides three to five options to be picked by test-takers as an answer. One option is called as a key and others are distractors. To create this test is nor easy and difficult. For instance, instructor must have an ability to investigate the nature of the distractors of the test to assess the viability of the test. This is in line with Brown (2004) who states that utilizing distractors is one of the approaches to gauge the viability of different decision tests. The distractors function as a bait choice to influence a test taker to pick the wrong decision. Eventually, Burton et al (1991) urge that it is essential to be sure about the significance of distractors to show up as conceivable alternative for the students who have not accomplished the goal being estimated. Therefore, distractors are assumed as such a crucial part to guarantee the nature of the test.

Unfortunately, not all instructors are aware of importance of these distractors form and understand how to create a good - multiple choices - test. Some of them only create the test regardless of whether the distractors work. Whereas, if distractors do not worked effectively, the test becomes futile and it cannot differentiate the test-takers to choose correct answer anymore. Besides, Gierl et al (2017) state that distractors are able to influence the quality of the test and learning outcomes. Briefly, it will not have the capacity to recognize higher group and lower group of students. Again, great distractors will likely draw the attention of the lower group more than the upper group. The multiple-choice test with poor distractors is responsible for the flaw made toward the test accuracy.

Analysis of a test is vital to the modification of the items included into the tests. From this position, an educator or test-designer can perceive how compelling is the test for their students, and it also becomes their own success measurement in teaching. In the event that there are an excessive number of easy or difficult items, the items should be changed or tossed out. In spite of the fact that a test is merely built as a purpose for final test, the items should be built deliberately on some proper theoretical considerations so that the items are not only score identifiers, they can go about as an inspiration promoter for students and as a good assessment for instructors themselves. From this stance, the teachers can decide whether to utilize a similar teaching method or substitute it to another method. The impact of assessment is for sure moderately prodigious for all parties in teaching and learning environment (Fulcher & Davidson, 2007).

There are many steps in writing good distractors. First, the statement that are taken from the students mistakes and confusions frequently make solid distractors. Second, the wrong option should be homogeneous, independent, familiar but incorrect phrases. Third, true explanations that don't answer the inquiry regularly will make great distractors. Fourth, absolute statements (e.g. never, all, etc) are best maintained as a strategic subtace in distractors. Then, differentiate the distractors adequately to the key answer by not only rearranging the words, but also write them in logical or numerical order. Next, ensure that the choices are autonomous and totally unrelated, and also avoid preposterous, jokey and peculiar distractors because they can be quickly seen. Finally, make each distractor linguistically alike with the key answer (Jackson, 2003; Haladyna & Rodriguez, 2013; Moreno, 2015).

From the recommendations of previously mentioned composing distractors, it appears that making distractors is likewise a muddled errand to perform. An instructor, as a test-producer, should draw in their abilities and information with reference to contribute to the well-developed distractors. Despite non-working distractors, those are marked by existing examinees who select those around at least 5%

(DiBattista and Kurzawa: 2011; Mkrtchyan: 2011). Additionally, Linn and Gronlund (2000) recommend that if distractors are not favored by any examinees, they should be removed or replaced. Apart from distractors, difficulty index is also needed. The difficulty index can be achieved to by gathering all test-takers who answer the items accurately. Asaad and Hailaya (2004) state that difficulty index depicts the extent of the item quantity of upper group and lower group who answers the item accurately. Besides, difficulty index of test item is expected to be rechecked whether the test is easy or excessively difficult, making it impossible to answer with the goal that the educator can assume whether the item should be reconsidered or wiped out (Brown: 2004; Khalifa and Weir: 2009).

There are recent studies in distractor analysis such as the study conducted by Sajjad et al (2020) by entiltled "Nonfunctional Distractor Analysis: An Indicator for Quality of Multiple Choices Question". This study is intended to analyze multiple choices in term of both poor and moderate distractors effeciency. The results of this study show that the persentages of the test was categorized as high, moderate and low respectively were 38%, 42%, and 20%. The second study, "An Analysis of Students Reading Final Examination by Using Item Analysis Program on Eleventh Grade of SMA Negeri 8 Medan", was done by Manalu (2019). This study aims to measure the quality of final examination of a reading test. Particularly at the level of distractors, it reports that there were four categories results including (1) poor (48%), (2) average (8%), (3) good (4%), and (4) excellent items (32%). The third study was done by Herawati (2012). She found that in the terms of index of difficulty, five items of stated detail question had a decent trouble, one implicit detail question was extremely troublesome, four references had a decent index of difficulty, one grammar question was very easy, seven items of vocabulary questions were exceptionally difficult, and three out of ten language focus questions were very easy. Regarding the distracters adequacy, one item must be reexamined. Taking these facts into account, the test items were generally good, despite the fact that there are a few things must be reconsidered and discarded. Subsequently, the English educator must be able modify it.

Similar to two researches mentioned above, this study discussed the same issue about analysis of a test. However, it focussed on the difficulty index and distractor analysis of multiple choice test at the summative test made by an english teacher. This study was considered important to be made to assess whether both difficulty index and the distractors created individually by the teacher meet the effective one. Moreover, as stated before, distractors play such a vital role in multiple choices test to ensure the quality of the test itself. The multiple-choice test that belongs to poor distractors is questionable concerning its accuracy of the test. On the other hands, as well as any question or stem is made, if distractors are not functioned effectively, then the test is useless now that it is most likely to lead test-takers to select true answer and, obviously, it will not be able to distinguish between high group and low group of students.

This study meant to examine the viability of English test regarding the difficult index and distractors of the English test taken by the students in a high school in Meulaboh. Based on the clarification above, the author formulated the following research questions: (1)What is the difficulty index of the English test? and (2) How is the viability of distractors of the English test?

In addition, this examination is relied upon to give additional data to educators about a decent quality test. Furthermore, not only exclusively would this examination give an important commitment to the university, but also the consequence of this investigation would likewise be beneficial to educators. They must be more mindful in making a test precisely, particularly in making formative as well as summative test so that test would have a good difficulty index and effective distractors. Moreover, this examination would empower and urge different specialists to direct researches of a similar issue.

Method

The techniques utilized as a part of this investigation were content analysis created by Laswell. As indicated by Stemler (2001), content analysis is important for taking a gander at designs and inspecting patterns in records. Churchill (2013) characterizes content examination as a versatile strategy that can be utilized to look at writings or information objects. Krippendorff (2003) adds that content analysis alludes to an exploration system to make replicable and substantial conclusions from instant messages (or other noteworthy issues) to the circumstances of their utilization. Briefly, it underlines a perspective of writings which incorporate word, sentence, and so on. The point of undertaking a quantitative content analysis was to break down information regarding the fact that the goal to research test items of the English summative test whether they met a decent difficulty index and successful distractors. Statistics was utilized in this investigation to discover the consequences of concentrate by utilizing Anates Software version 4 as stated in the following table.

Table 1. Criteria of difficulty index based on Anates Sofware

0 – 15%	Very difficult
16% - 30%	Difficult
31% - 70%	Moderate
71 % - 85%	Easy
86% - 100%	Very easy

Regarding the difficulty index, the higher the difficulty index record score, the simpler the test item is. Hence, the lower the difficulty index score, the more troublesome the test item is. Accordingly, either the most troublesome test item or the simplest test item must be revised. Further, the data were analyzed by utilizing the criteria of the adequacy of distractors as proposed by Sunarya (2003) as shown below:

Table 2. Criteria of the effectiveness distractors

Criteria	Interpretation
76% - 125%	Very good
51% - 75% or 126% - 150%	Good
26% - 50% or 151% - 175%	Fair
0% - 25% or 176% - 200%	Poor
$\geq 200\%$	Very poor

The sample of the research was an english formative test consisting 25 test items. In detail, each test item includes three answer choices. It means that these test items have 100 answer choices to be selected as the best answers by university students. This online test was taken by university students who were in the middle term of teaching and learning process. Therefore to analyze the data, two techniques were used, namely check list which was used to check the test item one by one and document analysis--in this second technique, the test items of midterm test were collected. The data were analyzed in detail beyond how the nature of the test. Concerning the criteria of the effectiveness distractors, it was calculated by using the percentage formula. The formula is as shown below:

$$P = \frac{F}{N} \times 100\%$$

To clarify the symbols above, *P* is the data percentage, *F* is frequency of item, and *N* is number of sample in the whole data with the criteria of percentage suggested by Arikunto (2009) in the following:

Criteria	Interpretation
76 - 100%	Good
56 - 75%	Sufficient
40 - 55%	Fair
< 40%	Bad

Results

Difficulty Index

By dissecting the difficulty index of the English summative test, the test producers or educators would know which items of the English summative test items are classified into three categories as easy, moderate, and difficult. The result is as shown below:

Table 4. Difficulty index of the second-grade summative test

Index of Difficulty	Interpretation	Total	%
0 – 30%	Difficult	0	0%
31% - 70%	Moderate	12	48%
71% - 100%	Very easy	13	52%

The information above demonstrates that there is no difficult or exceptionally difficult thing in the second grade English summative test items. Contrarily, the test items are arranged into three categories which include moderate, easy, and very easy test items. To be specific, there are six easy items, seven very easy items, and twelve moderate test items. From all classifications, moderate test things achieve a higher rate, specifically 48%.

Distractors

Distractors allude to the appropriate response alternatives gave in numerous decision tests. It allows us to draw in test takers who have not increased their ability. To know whether the distractors worked, all choices of each test item consisting of 100 answers would be embedded into tables with the goal that it could be watched the students who neglected them to pick the right answers. The adequacy of distractors was likewise classified factually by applying Anates V4. The outcome is as following:

Table 2. Distractors effectiveness

Category	Good	Fair	Poor	Total
Persentage	41	26	33	100

The table above shows that there is 26% of the distractors which are reasonable. It implies that the adequacy of distractors list extends between 26% - 50% or 151% - 175%. In this way, it should be changed. Moreover, there are 33 incapable distractors which must be disposed. Then, 41 viable distractors can be kept in the test.

Discussion

As mentioned before, there were 25 of 100 answer choices to be selected by 62 students to answer the questions provided. The result shows that the number of successful distractors is 41 items. It means that

less than 50% distractor was able to create students' confusions. As Jackson (2003) suggests that one of steps in writing good distractors is the statement that are taken from the students mistakes and confusions frequently make solid distractors.

Next, the number of insufficient distractors is 33 items. It means that these test items should be removed. Other distractors classified as reasonable distractors are recommended to be rechecked or reexamined to be better distractors. This statement is in line with Linn and Gronlund (2000) who recommend that if distractors are not favored by any examinees, they should be removed or replaced.

Compare to the recent studies, all these findings are very different with the results of studies conducted by Manalu (2019) and Sajjad (2020). The highest persentage of Manalu's study achieved by poor distractors. Vice versa, the result of this study shows that the highest one is by good distractors. Otherwise, in Sajjad (2020) study, the highest persentage is moderate. However, it would be natural if these finding meets different result due to different test items.

For further information, only 53 items of the distractors were analized, especially for those that met the syllabus requirements. Regarding the analysis of each item, it was discovered a few mistakes in building distractors. They were likely the elements why the distractors turn out to be less powerful or incapable. The errors found were as follows; First, the distractors are not homogenous to the key answer. Five various phrases were made and the distractors are either shorter or longer than others (3 test items). Here is one example of trouble test item.

Table 3. Example of shorther or longer answer

No Item	TS	Question
		WHY THE MOSQUITO BUZZ?
		To begin with the story, let me tell you that a long time ago, mosquitoes didn't buzz, the talked and talked. Then one day when one of them talked to an Iguana and didn't let the Iguana say one word. The Iguana just grumbled and waved her tail until she forgot to say hello to her friend a snake. After that everything was in mess. The snake was so angry that he made a rabbit, monkey, a crow, and an owl get frightened. Finally, the whole jungle was mad at the mosquito and cried for the sun to come up and when it did, the mosquito lost his voice.
1		The text is about……
	5---	a. Mosquito kept silent
	0--	b. Mosquito was loved by its friends
	0--	c. Mosquito talked very much and its friends were happy
	1+	d. Mosquito was the best friend's Iguana
	57**	e. Mosquito lost his voice because sun came up

The length of phrases used in the options above is different with others. Contrast to the other distractors, the key answer and option C use two phrases while others use only one phrase. These both short and long sentences made them inconsistent. This case is regardless of what suggested by Haladyna & Rodriguez (2013) and Moreno et al (2015) to keep distractor homogenous including content and structures. In addition, this test item analysis based on Anates, option D is classified as a good and effective distractor. Otherwise, one option is included as an extremely poor distractor and two of four options are classified as ineffective distractor since it is not functioned at all.

Second, the items comprise of inaccurate grammar (1 test item). The test item is shown in this following table:

Table 4. inaccurate grammar

3		Based on the above text, which of following statements is NOT TRUE?
	2++	a. The iguana just grumbled
	54**	b. The snake lost his voice
	0--	c. The iguana waved her tail

6---	d. The mosquito do buzz
1-	e. The rabbit gets frighneted

This test item has an extremely effective distractor, namely option A. The distractor in option C, D, and E should be replaced or removed for it was selected by less than 5% of test takers as suggested by DiBattista and Kurzawa (2011) and Mkrtchyan (2011). Apart from that, there is a mistake in distractor D and E which consisted of a wrong grammar (word "do" should be replaced by "does") and wrongly typewritten (word "frighneted" should be replaced by "frightened").

Third, some key answers were directly incorporated (2 test items). In this case, the phrases were precisely shown as it is in the text in order that the test takers could answer the items effortlessly. Finally, a few stems or choices additionally provide the hints to the key answer i.e. 3 test items. Aside from that, all options incorporated into the English test are started with a capital letter. It demonstrates that there is no separation among questions and explanations shaped in the stem. Though as Gronlund and Waugh (2009) urge that if the stem is being referred to as a question, every option should start with a capital. Further, a lowercase letter is utilized as a part of every option when the stem is a deficient articulation. Moreover, there are numerous mix-ups in typewriting either in the stems or choices. Incorrectly spelled words can be found in 3 items.

Conclusion

Considering the results above in relation to the viability of distractors, it demonstrates that 41% of English formative test is acknowledged. Moreover, the total of poor distractors is 30%. It implies that the nature of the English summative test items is averagely great in term of the viability of distractors. Moreover, either easy or difficult items could be influenced by the adequacy of distractors. So that, the instructor additionally should focus on the nature of distractors.

The study of test analysis is important in order that similar study related to nature of the test is needed. The weakness of this study was using the small sample of subject and object of study that is less than 100 distractors and also students. Therefore, it is better for future research to take more than 100 test takers and distractors.

References

Arikunto, S. (2009). *Dasar-dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.

Brown, H Douglas. (2000). *Language Assessment: Principles and Classroom Practices.* White Plains, NY: Pearson Education.

Burton, S.J, Richard R. Sudweeks, Paul F. Merril, Bud W. (1991). How to Prepare Better Multiple Choice Tests: Guidelines for University Faculty. *Retrieved from* https://testing.byu.edu/handbooks/betteritems.pdf

Churchill, B. (2013). *Social Research Methods*. Melbourne: Oxford University Press

DiBattista, David & Laura Kurzawa. (2011). Examination of the Quality of Multiple-choice Items on Classroom Tests: *the Canadian Journal for the Scholarship of Teaching and Learning: Vol. 2: Iss. 2, Article 4,* 1-23.

Fulcher, G, & Fred Davidson. (2007). *Language Testing and Assessment*. New York: Routlegde.

Gierl, M.J. (2017). Developing, Analyzing, and Using Distractors for Multiple-Choice Tests in Education: A Comprehensive Review. AERA: 87(6), 1082-1116. http://rer.aera.net

Haladyna, T. M. (2016). Item analysis for selected-response test items. New York, NY: Routledge.

Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating test items*. New York, NY: Routledge.

Jackson, K. (2003). Guidelines for Writing Multiple Choice Questions. *Retrieved from* http://www.education.vic.gov.au/languagesonline/games/comprehension/docs/multiple%20choice%20questions.pdf

Khalifa, H & Cyril J.W. (2009). *Studies in Language Testing.* Cambridge: Cambridge University Press.

Krippendorff, K. (2003). *Content Analysis: An Introduction to Its Methodology 2nd Edition*. Thousand Oaks: SAGE Publications.

Linn, R. L & Gronlund. (2002). Measurement and Assessment in Teaching. Upper Saddle River, NJ: Prentice Hall.

Manalu, D., Sipayung, K.T, & Lestari, F.D. (2019). An Analysis of Students Reading Final Examination by Using Item Analysis Program on Eleventh Grade of SMA Negeri 8 Medan. *Journal of Eenglish Teaching & Applied Linguistic*, 1(1), 13-19. https://doi.org/10.36655/jetal.v1i1.98

Mkrtchyan, A. (2011). Disctractor Quality Analyze in Multiple Choice Question Based on Information Retrieval Model. *Retrieved from* http://www.eqanie.eu/media/Conference%20Vienna%202011/Papers/Paper%20Mkrtchyan_Short.pdf

Moreno, R., Martínez, R. J., & Muñiz, J. (2015). Guidelines based on validity criteria for the development of multiple choice items. *Psicothema*, *27*, 388–394. doi:10.7334/psicothema2015.110.

Sajjad, M., Iltaf, S. & Khan, R.A. (2020). Nonfunctional distractor analysis: An indicator for quality of Multiple choice questions. Pak J Med Sci. 2020;36(5):982-986. doi: https://doi.org/10.12669/pjms.36.5.2439

Stemler, S. (2001). An overview of content analysis. Practical Assessment, Research & Evaluation, 7(17). Retrieved from http://PAREonline.net/getvn.asp?v=7&n=17

Sunarya. (2003). Panduan Anates Kr-2003. Retrieved from http://file.upi.edu/Direktori/FIP/JUR._PSIKOLOGI_PEND_DAN_BIMBINGAN/195911301987031-YAYA_SUNARYA/BAHAN_EVALUASI-ASESMEN/ANALISIS.pdf.