

## THE QUALITY OF TEACHER-MADE MULTIPLE-CHOICE TEST USED AS SUMMATIVE ASSESSMENT FOR ENGLISH SUBJECT

Ni Putu Liana Santy<sup>1</sup>, Ni Luh Putu Eka Sulistia Dewi<sup>2</sup>, A.A. Gede Yudha Paramartha<sup>3</sup>,  
Program Studi Pendidikan Bahasa Inggris  
Universitas Pendidikan Ganesha, Singaraja, Indonesia  
e-mail: [putu.liana.santy@undiksha.ac.id](mailto:putu.liana.santy@undiksha.ac.id)<sup>1</sup>, [sulistia.dewi@undiksha.ac.id](mailto:sulistia.dewi@undiksha.ac.id)<sup>2</sup>,  
[yudha.paramartha@undiksha.ac.id](mailto:yudha.paramartha@undiksha.ac.id)<sup>3</sup>



This is an open-access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

Copyright © 2020 by Author. Published by Universitas Pendidikan Ganesha.

Received : June, 2020

Accepted : July, 2020

Published : December, 2020

### ABSTRACT

*This study analyzed the quality of multiple-choice test used as midsemester test made by the English teachers in one school in Singaraja. This study was essential to be conducted since the items of the multiple-choice test must have good quality to be used to assess the students' achievement levels. This study used content analysis method in analyzing 100 items from 3 different instruments. In collecting the data, the checklist analysis form was used to compare the items of teacher-made multiple-choice test with the norms as one of the standards in making a good multiple-choice test, then clarified through interview. Based on the data that have been analyzed, the multiple-choice tests made by English teachers have very good quality. There is 1% of the items has sufficient quality, 8% of the items have good quality, and the rest of the items have very good quality. There are 5 norms fulfilled perfectly, such as the norms about reflecting the basic competencies, opinion-based items, spelling, double negatives, and absolute options. However, the most common mistake found is in the norms of punctuation and capitalization. It can be concluded that the teachers already follow the norms of making a multiple-choice test, and it is indicated that the quality of the multiple-choice test is not the only factor that affects students' achievement levels.*

**Keywords:** *instrument quality, norms, teacher-made multiple-choice test*

### BACKGROUND

The teachers use the multiple-choice test as a summative assessment to assess the students' achievement levels precisely (Omoruan, 2018). Even though it is a traditional assessment, the multiple-choice test is credible and relevant to be used. Effendi and Suyudi (2017) stated that multiple-choice test had been used as the National Examination in Indonesia since 1950s for assessing particular subjects and one of the subjects is English subject. Lebagi et al. (2017) argued that multiple-choice tests have been designed by teachers could help the students to be ready in facing further examination. It means that the quality of teacher-made multiple-choice test designed by the teachers is expected to have good quality since it is used as a summative assessment to assess the students' achievement levels.

In constructing the multiple-choice test, teachers as the test makers need to follow some standards (Haladyna, 2004). One of the standards of making a good multiple-choice test is norms. It is the starting point in the guideline of making a good multiple-choice

test besides the validity and reliability (Burton et al., 1991; Haladyna, 2004; Hall & Marshal, 2013; Zimmaro, 2016). It is supported with the results of the previous studies which show the essential role of norms in making multiple-choice test. Patil et al. (2016) found that the accuracy of the multiple-choice test based on the norms can be the factor that affects the difficulty index, discrimination index, and distractor efficiency of the test. In addition, Crockett and Churches (2016) stated that one of the requirements to conduct a good assessment practice is following the norms. Furthermore, Martinez et al. (2009) argued that assessment practice done by the teachers influences the students' achievement levels. Furthermore, according to Lebagi et al. (2017), the quality of the multiple-choice test is needed to be analyzed since it can influence the students' achievement levels and it is still found mistakes that exist in a multiple-choice test. It can be concluded that norms take an essential role in making the multiple-choice test since it is one of the standards that exist in the guideline. It can influence the students' achievement levels indirectly.

Based on the preliminary observation in one school in Singaraja, Buleleng Regency, multiple-choice test is used to assess the students' achievement levels for English subject in the midsemester and final examinations. The development of items was based on the syllabus, and the English teachers validated the items of multiple-choice tests before they were given to the students. It is done by the teachers to make a good quality multiple-choice test. However, based on the results of the National Examination in academic year 2018/2019 obtained by students in the school, the students could only achieve an average score of 53.31 from the standard minimum scores of the National Examination in English subject that is 55.00, (Puspendik Kemendikbud, 2018). Furthermore, there are 72% of students obtained scores less than 60, which is the standard minimum scores that the students should achieve are above scores 70. It indicates that there is a problem with the construction of the teacher-made multiple-choice test in this school and the unidentified quality of multiple-choice test based on the norms can be the factor that also causes the problem.

This study was conducted to investigate the quality of multiple-choice tests used in the midsemester test, which were made by the English teachers. The focus of this study was to analyze the quality of multiple-choice tests based on the norms of making a good multiple-choice test. This study emphasized more on the analysis of the congruity of the items of the multiple-choice tests with the norms since the previous studies only focused on the difficulty index, discrimination index, and distractor efficiency. It is also needed to analyze the quality of multiple-choice test since there are still found the multiple-choice tests that have not identified its' quality yet. The norms used in this study have been synthesized from Haladyna (2004), Hall and Marshal (2013), and Puspendik Kemendikbud (2019). Then, from the results of the analysis, it can be seen how the quality of the multiple-choice test made by the teachers.

## **METHOD**

The content analysis method was used in this study to analyze the quality of multiple-choice test made by the English teachers. It is a research method applied to written or visual materials to identify the specified characteristics of the materials (Ary et al., 2010). It is also stated that content analysis is widely used in education, in which some of the purposes in educational researches are to identify bias or prejudice, to analyze the

errors and to discovers specific topics in particular documents, (Ary et al., 2010). Thus, content analysis was appropriate to be used in this study.

This study was conducted in SMPN 6 Singaraja, which is located at Bisma street No. 3, Banjar Tegal, Singaraja, Buleleng, Bali. There were 100 items of 3 multiple-choice tests made by the English teachers analyzed. The items in multiple-choice test grade seven were 40 items, and the items in multiple-choice test grades eight and nine were 30 items for each grade. Besides, the quality of those items was the object of study. There were document analysis and interview as the methods of data collection. The instrument used in document analysis was the checklist analysis form. The items of the multiple-choice tests were analyzed its congruity with the norms of making a good multiple-choice tests. The results of document analysis were reconfirmed by conducting interview section. In analyzing the results of the interview, there were three steps that have been passed, such as data reduction, data display, and data conclusion.

To find out the quality of multiple-choice test made by the English teachers, the items of the multiple-choice test were compared with the norms that have been synthesized from the norms suggested by Haladyna (2004), Hall and Marshall (2013), and Puspendik Kemendikbud (2019). There were 18 norms used to analyze the items and 2 norms used to analyze the instrument of multiple-choice tests. Those 18 norms used to analyze the items are about reflecting basic competencies, the independence of item with the previous item, the clarity of the item, opinion-based item, the clue to the correct answer, grammar, spelling, format of the options, punctuation and capitalization, double negatives, the homogenous of the options, number of the correct answer, the length of options, the placement of the options, repetition of words or phrases, the plausibility of the distractors, and the using of “none of the above” or “all of the above”. The norms used to analyze the instrument are the clarity of the instructions and the variated locations of the correct options.

First, the data of the items were fulfilled in the checklist whether the items have fulfilled the norms or not. Then, the data was inserted in MS. Excel by giving codes “1” for the norm fulfilled and “0” for norm unfulfilled. This method was used to find out the percentages of norms fulfilled by the items, the percentages of norms fulfilled, and then categorized them by judging the items based on the formula stated by Nurkencana and Sunartana (1992). There were 5 classifications based on the percentages of each item. These categorizations can be seen in Table 1.

**Table 1. Classifications of the teacher-made MCTs’ quality**

Interval	Criteria
$75\% \leq x$	Very Good
$58\% \leq x < 75\%$	Good
$42\% \leq x < 58\%$	Sufficient
$25\% \leq x < 42\%$	Poor
$x < 25\%$	Very Poor

The item is classified very good quality if the percentage is more than or equal to 75%. If the percentage is more than or equal to 58% but less than 75%, the item is considered to have good quality. The item is categorized sufficient if the percentage is more than or equal to 42% but less than 58%. The item has poor quality if the percentage

is more than or equal to 25% but less than 42% and is very poor if the percentage is less than 25%.

The interview was conducted to get additional information to support the results of the data analysis. This study used open-ended interview based on the results of the data analysis. The English teachers were interviewed based on the results of the analysis through the checklist analysis form.

**FINDING AND DISCUSSION**

The data of this study were obtained from three different multiple-choice tests. The total numbers of the items from those multiple-choice tests that had been compiled and analyzed were 100 items. The results of the analysis show the total of norms fulfilled and unfulfilled by the items. The percentages of the norms fulfilled by items in grade seven, eight, and nine can be seen in Table 2.

Table 2. The Percentages of the Norms Fulfilled by Items

Item Number	Grade Seven Norms Fulfilled (%)	Grade Eight Norms Fulfilled (%)	Grade Nine Norms Fulfilled (%)
1	89%	83%	89%
2	78%	78%	78%
3	78%	83%	89%
4	89%	83%	89%
5	72%	67%	89%
6	94%	83%	89%
7	100%	83%	78%
8	89%	89%	94%
9	83%	67%	89%
10	67%	67%	67%
11	83%	78%	67%
12	78%	83%	83%
13	72%	83%	89%
14	78%	78%	83%
15	100%	89%	94%
16	94%	83%	83%
17	100%	83%	89%
18	83%	83%	89%
19	100%	94%	94%
20	94%	56%	89%
21	83%	89%	89%
22	89%	89%	100%
23	78%	89%	94%
24	94%	89%	89%
25	83%	83%	89%
26	83%	94%	89%
27	94%	89%	89%
28	100%	89%	83%
29	94%	94%	89%

30	100%	94%	94%
31	100%		
32	94%		
33	89%		
34	100%		
35	83%		
36	89%		
37	89%		
38	89%		
39	89%		
40	83%		

Out of 100 items, there are 9 items which perfectly fulfilled the 18 norms. The 16 items have fulfilled more than 90% of norms. It means those items neglected one norm. There are 56 items have fulfilled above 80% of norms and neglected 2 until 3 norms. 12 items neglected 4 until 5 norms and were categorized as above 70% norms fulfilled. 6 items have above 60% of norms fulfilled as a result of neglecting 6 norms. There was only 1 item neglected 8 norms with the lowest percentage, that is above 50% norms fulfilled. It can be concluded that the items of multiple-choice tests have different numbers of norms unfulfilled in ranges of 0-7 norms.

Based on the results of the analysis, it was found that 9 items that perfectly followed the 18 norms of making a good multiple-choice test. There was an item neglected 8 norms causing the item only fulfilled above 50% norms. Besides, the biggest number of the items, that is 56 items neglected only 2-3 norms, and they fulfilled above 80% norms. The most common mistake found in the items is the error in punctuation and capitalization, which makes there were 69% of items not following the norm. Besides, the rest of the norms cover 72%-100% norms following them. The complete information about the percentages of fulfilled and unfulfilled norms can be seen in Table 3.

Table 3. The Percentages of Fulfilled and Unfulfilled Norms

Norms Number	Norms' Description	Percentages of Fulfilled Norm Frequencies (%)	Percentages of Unfulfilled Norm Frequencies (%)
1	Reflecting basic competencies	100%	0%
2	Not depending on the previous option	92%	8%
3	Giving a clear focus	98%	2%
4	Avoiding opinion-based items	100%	0%
5	Not giving a clue to the correct answer	87%	13%
6	Being grammatically correct	72%	28%
7	Having correct spelling	100%	0%
8	Options are formatted vertically	43%	57%

9	Taking concern on the use of the punctuation and capitalization	31%	69%
10	Not containing double negatives	100%	0%
11	Options are homogenous	83%	17%
12	Having one correct answer	99%	1%
13	Options have about the same length	90%	10%
14	Options are placed in logical and numerical order	94%	6%
15	Options do not repeat the same words or phrases	99%	1%
16	Options are not overlapping	85%	15%
17	Distractors are plausible	82%	18%
18	Not using “none of the above” or “all of the above.”	100%	0%

The item can be judged to have a good quality of the multiple-choice test if the item has fulfilled 18 norms. Thus, 9 items that covered the highest percentage, that is 100% norms fulfilled. Those items were clear, which was reflected the basic competencies that the teachers wanted to be achieved. The items were independent, did not give any clues to the correct answers, and they were not subjective or opinion-based items. The grammar and spelling in the items were correct. The options of the items were formatted vertically and in a logical order. The lengths of the options in each item were about similar, and they were homogenous. Each item consisted of one correct answer. Therefore, the options did not use “all of the above” or “none of the above.” Each option of the items did not repeat word or similar to other option, the distractors were plausible, and the options were not overlapping. The stems of the items did not contain double negatives, and the punctuations and also the capitalizations of the items were correct. One of the 9 items which have 100% norms fulfilled can be seen in Figure 1.

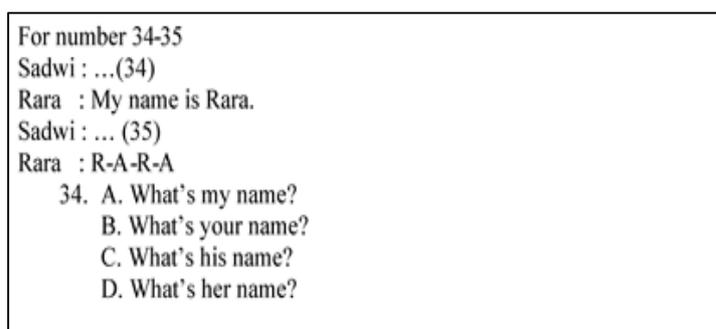


Figure 1. The example of an item with 100% norms fulfilled

The item of the multiple-choice test above given to the students in grade seven. It is about an introduction, the topic that exists in basic competency 3.2. The options were also plausible and homogenous because all options make-sense and in the form of interrogative sentences. The other norms in each dimension were fulfilled, such as the

dimension of content, dimension of style and format, dimension of writing stem, and dimension of writing options. Although there was no direct instruction to complete the dialogue, the item was still understandable by the students.

Besides, there were 56 items have fulfilled above 80% norms, in which the items did not fulfil 2-3 norms. Those 56 items had different norms unfulfilled, or it can be said that the norms unfulfilled were varied. The most common unfulfilled norms from the 56 items were about punctuation and capitalization, and also the format of the options existed in the dimension of style and format.

There are 5 mistakes found in punctuation, such as 1) the options that must not be ended with a full stop, yet there are full stops that existed at the end of some options. 2) The options are in the form of the interrogative sentence, but there is no question mark at the end of the options. 3) The stem consists of more or less than four full stops, and there is blank space at the end of the stem. 4) The stems are in the form of a question, but a full stop changes the question mark. 5) The options must be ended with the full stops, but several options are not ended with a full stop. The results of the interview support it. Two of the teachers stated that they did not have prior knowledge about the norm of punctuation.

Meanwhile, there are 5 mistakes found in capitalization, such as 1) The initial letter of the options should be not in capital letter, yet, the initial letter of some options is in capital letter. 2) There are options in the form of sentences, but the initial letter of those options is not in a capital letter. 3) When the blank space is at the beginning of the stem, the initial letter of the options is not in capital letter. 4) When the blank space is at the end of the stem, there are several options that the initial letter of the options is capitalized. 5) When the stem is in the form of an interrogative sentence, it is found that the initial letters of the stem and options are not in capital letter. Based on the results of the interview, it is stated that the teachers did not have enough time to revise those mistakes in capitalization. According to Allen (2002), punctuation is an essential part of writing which has a useful purpose, that is making the writing clear and easy to understand. Connelly (2009) argues that if the teachers did mistakes in punctuation, the test will be ambiguous and cannot be understood. It means that the accuracy and correct usage of the punctuation mark is highly needed to give a clear meaning of a text.

The next common mistake found in the items was the format of the options. Most of the options were not formatted vertically, but they were formatted horizontally. From 56% of items, 48 items were not formatted vertically. The teachers pointed out that the reason why most of the options were formatted horizontally was that they want to economize the use of papers. The other mistakes found in the items were the independence of the item, giving a clue, grammar, the number of correct options, homogeneity, length, and placement of the options, overlapping, and also plausibility of the options.

One of the examples showing the independence of the item was the item asks about the weight of the tablet, while the previous item asked about the kind of the drug and one of the options was “tablet”. The next mistake was giving a clue to the correct answer, the stem in an item mentioned the weight of the tablet. It gave a clue to the correct answer of the previous item that asks about the kind of the drug. Then, the grammar in the items was not perfect. For example, there was a conversation between two persons, and one of them said “Really? It’s great, guys”, in which the word “guys” refer to many

people, and it was not suitable used in the conversation between two persons. 2 options of the items were possible to fulfil the appropriate response of the conversation. It was not appropriate to the norm which states only one option in the item as the correct answer. The options were not homogenous; it happened in some items, which the type of option was not similar to the other options. The next mistake is some options were not similar in length found in some items. Option C consisted of 14 words, while others consisted of 7 words which made the option C is different from other options. The options were not placed in a logical order. For example, there was an option which consists of 14 words, while others only consisted of 7 words which made the option was very long. However, the option was placed in the middle of the options (in option C). The overlapping option was the next mistake found in the 56 items. There was an option which has no relation with other options.

Two mistakes found in the plausibility of the options were 1) The option was not suitable to fulfil the blank space because it was incorrect in grammar. 2) The option was not making-sense to fulfil the blank space. The mistakes of the plausibility of the options were supported with the statements from the teachers. The teachers said that they were not careful and did not check the plausibility of each option.

Besides, the norms of homogeneity and plausibility of the options should be paid attention (Mukherjee & Lahiri, 2015; Lebagi et al., 2017; Ingale et al., 2017). Green (1984) found that the homogeneity of the options in the items of the multiple-choice tests gave impact to the item difficulty. It means that the homogeneity of the options is essential since it affected the item difficulty. Meanwhile, Dehnad et al. (2014) stated that the ability to write plausible options in the multiple-choice test was often a highly needed job by the test makers. According to Ebel (1951), it is still found that there were groups of teachers who making poor items of a multiple-choice test. One of the mistakes done by the teachers was the implausible options. It is related to the findings, in which the teachers still did mistakes in writing homogenous and plausible options.

There is only one item which has fulfilled above 50% norms. It is the item in grade eight number 20. The item neglected 8 norms from 18 norms that must be fulfilled. It was found that the item gave a clue to the correct answer. Option D was incorrect in grammar, and the stem was not correct in punctuation. The options of the item were not homogenous. There was a very long option, and it was placed in option B. Option D was overlapping, and the options were not plausible. The item considered to have above 50% norms fulfilled can be seen in Figure 2.

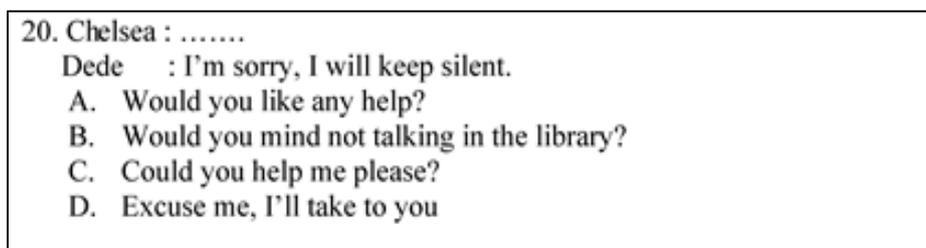


Figure 2. The example of an item with 50% norms fulfilled

Previously, it has been mentioned that the most common mistake of the items was in the norm of punctuation and capitalization, which means the norm of punctuation and

capitalization was the most unfulfilled norm. There were only 31% of the items fulfilling the norm. It is supported by the results of the interview with the three different teachers—only one teacher who knows about the usage of punctuation and capitalization in every type of item.

Different to the norm of punctuation and capitalization as the lowest fulfilled norm, there are 5 norms which were perfectly fulfilled by all items. Those 5 norms were norms about reflecting the basic competencies, opinion-based items, spelling, double negatives, and absolute options, that is the use of “none of the above” or “all of the above” which are prohibited. From those 5 norms, there were some norms that the teachers have known and some of them are not.

The first norm that is perfectly fulfilled is the congruity among the items of multiple-choice tests with the basic competencies. The teachers confirmed that the congruity among the items with the basic competencies is the most important in constructing the multiple-choice test. Thus, it was checked frequently before the teachers gave the tests to the students. It is in harmony with the statement stated by Lebagi et al. (2017). They argued that the test makers have to pay attention to the congruity among the items of multiple-choice tests with the contents that wanted to be assessed. The congruity among the basic competencies with the items has a vital part in assessing students' achievement levels, in which from the results of the test, the students will get the precise information about their knowledge capabilities, (Roy, 2016).

The second norm that is 100% fulfilled is the spelling of items must be correct. Those three teachers argued that they did not make any mistake in spelling because of the aid of their laptops. So, they used MS. Word in making the multiple-choice tests. While they were inserting the words in MS. Word, if the word is misspelt, it would be signed automatically. Then, they changed the word to the correct one.

The other fulfilled norms are avoiding to use “none of the above” or “all of the above” in the options, avoiding the opinion-based items, and avoiding double negatives. The teachers stated that they did not know exactly about those three norms—one of those teachers known this norm indirectly by exploring the previous multiple-choice tests.

From the results of the fulfilled norms and the correlation with the results of the interview, it can be concluded that the most significant number of the unfulfilled norm is the norm of punctuation and capitalization. There are 5 norms which were perfectly fulfilled out of 18 norms. The rest of the norms are considered to have 43%-99% items following them.

Besides the 18 norms as the standards in categorizing whether the multiple-choice tests have good quality or not, two additional norms determine the quality of the multiple-choice tests, such as the locations and number of the correct options and also the clearness of the instructions.

First, locations of the correct options in the multiple-choice tests were varies. It also happens in the number of correct options in each grade. The total number of options in every grade is around 3-13 correct options. It means that the number of options from each multiple-choice test varies. It correlates with the statement from one of the teachers who said that the number of the correct options must be varied.

The second one is about the clarity of the instructions in the multiple-choice tests. Based on the results of the analysis, the instructions in the multiple-choice tests in grade seven, eight, and nine were clear enough. Those three multiple-choice tests were started

with the instruction to choose the best answer between options A, B, C, or D. There are some of the instructions which gave information about the stems for particular numbers, such as “The following text is for questions 15 to 18” (Grade nine.15-18). However, some items must consist of instructions to give clear orders, yet those items did not include them.

In general, the quality of the items of multiple-choice tests made by the teachers has very good quality. There are 91 items considered to have a very good quality, in which they have percentages more than or equal to 75%. Besides, the number of items which has good quality is 8 items. They have percentages of less than 75% and more than or equal to 58%. However, it can be seen that 1 item has percentages of 56 %, in which it is less than 58% and more than or equal to 42%. Thus, the item was categorized as sufficient quality.

Even though the results of the quality analysis and the statements from the teachers show that those multiple-choice tests have good quality, yet it was not coherent with the satisfaction of the teachers with their multiple-choice tests. The teachers were not satisfied with the multiple-choice tests because there are only a few students who got the good scores, there were some mistakes existed in the multiple-choice tests, and they did not know about many norms of making a good multiple-choice test. Even though they have prior knowledge about how to construct the multiple-choice test, they argued that they need more practices in making the multiple-choice tests.

The quality of the teacher-made multiple-choice tests is related to the prior knowledge that the English teachers have as the test maker. Their knowledge about how to construct the multiple-choice test was derived from 4 factors. First, they got the lecture of assessment in the college. Second, two of the teachers have followed the workshop and socialization of making a good multiple-choice test conducted by MGMP. Third, the teachers explored the multiple-choice tests on the internet, it helps the teachers to vary the types of the items. Fourth, the teachers explored the previous multiple-choice tests.

Even though the quality of the multiple-choice tests was excellent, yet it is not aligned with the scores of the National Examination got by students in school in academic year 2018/2019. The students could only achieve an average score of 53.31 from the minimum scores of the National Examination in English subject that is 55.00, (Puspendik Kemendikbud, 2018). Furthermore, most of the students got lower scores than the minimum scores in the midsemester test. It is related to the results of the interview. The teacher stated that most of the students got bad scores in the midsemester test, although the multiple-choice test correlates with the materials that have been taught in the class.

The teachers stated that other factors influenced the students low scores in the midsemester test. The factors come from the students and teachers themselves. Those factors are students’ motivation in learning, students’ self-awareness, fewer vocabularies, and the teachers’ motivation in the process of teaching and learning.

First, students have less motivation in learning. Based on the findings, it was found that the students did not have the initiative to ask the teacher when they had not understood the materials yet. Motivation is a desire or feeling felt by people that encourages them to do something, in which the students who have high motivation would do the best in learning, (Santrock;2004, Pintrich & Date 2008). The students’ achievement levels on the materials were assessed by using an assessment, that is the multiple-choice test providing the materials that have been taught in the class. So, it can be said that

students' motivation gives effect to the students' achievement levels besides the quality of multiple-choice test made by the teachers.

Second, the other possible thing that affected students' achievement level is their self-awareness. Flavian (2016) argues that self-awareness is defined as a concept describing people's description of themselves. According to Arabsarhangi and Noroozi (2014), students' self-awareness affects their performance and achievement levels in tests. Therefore, according to Rinkeviciene and Zdanyte (2002), self-awareness is important to be taught to the students to help them to find out what their needs and develop their motivation to achieve their goals. Relating to the findings, the students did not ask the materials that they have not understood yet, because they did not have self-awareness which means that they did not know whether they have understood the materials or not. Meanwhile, because there was no question about the material, the teacher continued the other material. It is indicated that the multiple-choice test is not the only factor that affects students' scores in the midsemester test.

Third, the other factor that affected the students' performances and achievement is students' difficulties in mastering the vocabularies. Ur (1996) stated that vocabulary takes an important role to be taught in the process of teaching and learning English, and without mastering the vocabulary, nothing can be conveyed. In addition, it is supported by Richards and Renandya (2002), who argued that vocabulary takes place in a core component of language proficiency. Relating to the findings, the teachers stated that the students only have limited vocabularies. Thus, they were difficult to understand what is being asked in the items of multiple-choice tests. It means that the vocabulary mastery of the students is needed to be improved, and it signifies that not only the multiple-choice test affects students' achievement levels.

Fourth, the factor which could give impact to the students' achievement level is teachers' motivation in the class. Iliya and Ifeoma (2015) argued that teachers' motivation in the class directly connected to the instructors' desire to take part in the teaching and learning process and share the knowledge that they have to the students. It is supported by the statement of Tastan et al. (2018), who found that teachers motivation gives a significant impact on the students' achievements. Based on the interview, one of the teachers stated that she has a small voice that can be her weakness in teaching English. It affected the process of delivering materials. Besides, the other teachers did not know about making multiple-choice test as well as the teacher who has a small voice. It means that the teachers need to improve their motivation in teaching. They should find the strategies to cover their weaknesses, such as by moving around when the voice cannot be heard by the students and exploring many examples of items or the guideline of making a good multiple-choice test.

From the description above, it magnifies that the multiple-choice tests made by the English teachers were not the only one factor in influencing the students' achievement level in the midsemester test since the quality of the multiple-choice tests made by the English teachers is very good. In addition, more attention on the norm of punctuation and capitalization is needed, since it is the most common mistake found in the multiple-choice test.

## CONCLUSION

Based on the data above, it can be concluded that the teachers in the schools already applied their knowledge about norms in making multiple-choice tests since the multiple-choice tests made by the English teachers have good quality based on the norms. There are 5 norms which have been fulfilled by all items, such as the norms of reflecting basic competencies, opinion-based items, spelling, double negatives, and the using of “none of the above” or “all of the above”. However, 13 norms were not fulfilled perfectly by all of the items, such as the independence of the option, the clarity of the item, giving a clue to the correct answer, grammar, punctuation and capitalization, the format of the option, homogeneity, number of correct answers, length of options, repetition of the words, overlapping options, and plausibility of the distractors. The most common mistake found in the multiple-choice tests was the norm of punctuation and capitalization. So, the norm is needed to get more attention from the teachers in making the multiple-choice tests. In addition, the knowledge of punctuation and capitalization should be mastered by the teachers. From the data obtained during the interview, teachers need to explore the norms of making a good multiple-choice test because there are some norms that those teachers have not known yet.

## ACKNOWLEDGEMENT

The author would like to express sincere gratitude to Luh Putu Eka Sulistia Dewi and A.A. Gede Yudha Paramartha as the supervisors who have given guidance, encouragement, and valuable comments throughout the writing of this article. The author thanks the English teachers who have patiently cooperated by providing data and confirmation through interview. Also, the author also thanks the editors of PRASI who reviewed and gave comments to make this article is better than before.

## REFERENCES

- Allen, R.E. 2002. *Punctuation*. Oxford: Oxford University Press.
- Arabsarhangi, M. & Noroozi, I. 2014. The relationship between self-awareness and learners' performance on different reading comprehension test types among Iranian EFL elementary learners. *Theory and Practce in Language Studies*, 4(4), 675-685.
- Ary,D., Jacobs, L.C., Sorensen, C, & Razavieh. 2010. *Introduction to research in education*. Canada: Wadsworth, Cengage Learning.
- Burton, S. J., Sudweeks, R. R., Marrill, P.F., & Wood, B. 1991. How to prepare better multiple-choice test items: Guidelines for university faculty. *Brigham Young University Testing Services and The Department of Instructional Science*.
- Connelly, M. 2009. Get writing: Sentences and paragraph. *Cengage Learning*.
- Crockett, L.W. & Churches, A. 2016. *Mindful assessment*. Indiana: Solution Tree Press.

- Dehnad, A., Nasser, H., & Hosseini, A. F. 2014. A comparison between three-and four-option multiple-choice questions. *Procedia Social and Behavioral Sciences*, 98, 398-403.
- Ebel, R. L. 1951. *Writing the test item: In E.F. Linguist (Ed), educational measurement*. Washington, DC: American Council on Education.
- Effendi, T. & Suyudi, I. 2017. The impact of English national examination in Indonesia. *Advances in Social, Science, Education, and Humanities Research*. 82. Retrieved 2nd of September 2019 from <http://creativecommons.org>.
- Flavian, Heidi. 2016. Towards teaching and beyond: Strengthening education by understanding students' self-awareness development. *Power and Education*, 8(1),88-100.
- Haladyna, T.M. 2004. *Developing and validating multiple-choice test items*. New Jersey, London: Lawrence Erlbaum Associates.
- Hall, C. & Marshall, S. 2013. *A guide for developing multiple-choice and other objective style questions*. New Zealand: Centre for Academic Development.
- Hanna, G. S. & Dettmer, P.A. 2004. *Assessment for effective teaching: Using context-adaptive planning*. Boston, MA: Pearson A & B.
- Iliya, A. & Ifeoma, L.G. 2015. Assessment of teacher motivation approaches in less developed countries. *Journal of Education and Practice*, 6(22),10-17.
- Ingale, A.S., Giri, P.A., & Doibale, M.K. 2017. Study on the item and test analysis of multiple-choice questions amongst undergraduate medical students. *International Journal of Community Medicine and Public Health*, 4(5), 1562-1565.
- Lebagi, D., Sumardi, & Sudjoko. 2017. The quality of the teacher-made test in EFL classroom at the elementary school and its Washback in the learning. *Journal of English Education*. 2(2). 97-104.
- Mappiasse, S. S. & Sihes, A.J.B. 2014. Evaluation of English as a foreign language and its curriculum in Indonesia: A review. *English Language Teaching*, 7(10).113-122.
- Martinez, J. F., Steche, B., & Borko, H. 2009. Classroom assessment practice, teacher judgments, and student achievement in mathematics: Evidence from the ECLS. *Educational Assessment*. 14.78-102.
- Mukherjee, P. & Lahiri, S. K. 2015. Analysis of multiple-choice questions (MCQs) : Item and test statistics from an assessment in a medical college of Kolkata, West Bengal. *IQSR Journal of Dental and Medical Sciences*, 14(12), 47-52.

- Nurkencana, W. & Sunartana. 1992. *Evaluasi hasil belajar*. Surabaya: Usaha Nasional.
- Omoruan, B. E. 2018. Valid teacher-made tests: It's implication on students' achievement in business education. *ATBU, Journal of Science, Technology & Education (JOSTE)*. 6(2).
- Patil, R., Palve, S. B., Vell, K., & Boratne, A. V. 2016. Evaluation of multiple-choice questions by item analysis in a medical college at Pondicherry , India. *International Journal of Community Medicine and Public Health*, 3(6), 1612–1616.
- PerMendikbud. 2016. Peraturan Menteri Pendidikan dan Kebudayaan nomor 23 tahun 2016.
- Pintrich, P.R. & Date, H.S. 2008. *Motivation in education theory, research and application*. New Jersey: Pearson Education.
- Reece, I. & Walker, S. 1997. *Teaching, training and learning: A practical guide*. Sunderland: Business Education Publisher.
- Richard, J.C. & Renandya, W.A. 2002. *Methodology in language teaching: An anthology of current practice*. Cambridge: Cambridge University Press.
- Rinkeviciene, I. & Zdanyte, J. 2002. Raising students' awareness in language learning. *Kalbu Studios*, 3.
- Roy, Goutam. 2016. *Competency-based assessment in primary education in Bangladesh- A review*. Bangladesh: Share TA Office.
- Santrock, J. W. 2004. Psikologi Pendidikan. Jakarta: Kencana Prenada Media Group.
- Tastan, S.B., Davoudi, S.M.M., Masalimova, A.R., Bersanov, A.S., Kurbanov, R.A., Boiarchuk, A.V., & Pavlushin, A.A. 2018. The impacts of teacher's efficacy and motivation on student's academic achievement in science education among secondary and high school students. *EURASIA Journal of Mathematics, Science and Technology Education*, 4(6),2353-2366.
- Ur, Penny. 1996. *A course in language teaching: Practice and theory*. Cambridge: Cambridge University Press.
- Zimmaro, D. M. 2016. *Writing good multiple-choice exams*. Austin: University of Texas.