

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

Ni Kadek Dwi Candra Septi¹, A.A. Gede Yudha Paramartha², Luh Gede Eka Wahyuni³

¹²³Jurusan Pendidikan Bahasa Inggris, Universitas Pendidikan Ganesha
Singaraja, Indonesia

Email : kadek.dwi.candra@ undiksha.ac.id, yudha.paramartha undiksha.ac.id, ekawahyuni@undiksha.ac.id

ABSTRACT

Assessment is an important part in teaching and learning process. An instrument that is used to assess the students' level must be high in quality by following certain standard in constructing the instrument itself. This research is a descriptive research that aimed to investigate the quality of teacher-made multiple-choice tests that were used as summative assessment in middle test for English subject at SMP N 4 Singaraja. There are 125 items in total from 4 teacher-made multiple-choice tests. The data were collected by using document study and interview as the method with the assistance of checklist and interview guide as the instrument. The data was analyzed by comparing each item in the multiple-choice tests with a set of norms to find the congruity and further to be classified to determine the quality. The result shows that all of the teacher-made multiple-choice tests have a very good quality where 124 (99%) out of 125 items are qualified as very good and only 1 (1%) item is qualified as good. Some improvement is needed by paying more attention specifically for the unfulfilled norms.

Keyword: summative assessment; teacher-made multiple choice test; instrument quality; norm

INTRODUCTION

The education of English is implemented for students since English is really important to be possessed in this globalization era. To accomplish the learning goal of English education, one of the essential parts that must be applied in the teaching and learning process is assessment According to Tosuncuoglu (2018), assessment is used by teachers to classify and grade their students, give feedback and structure their teaching. In line with this statement, Taras (2005) states that educators can determine the level of skills or knowledge of their students through assessment so that it is accepted as one of the very crucial parts of teaching.

Since assessment is one of the important parts in teaching and learning process, it is

regulated in curriculum 2013. Based on Education and Culture Ministry, Province and District Education Department, the law conducted by Indonesian government, Curriculum 2013 has three aspects of assessment which are knowledge, skills, and attitudes aspects. Based on Regulation of Education and Culture Ministry No. 23 2016 about Educational Assessment Standard (article 9 paragraph 1 item c), which is used as the reference of the assessment standard in 2013 Curriculum, the knowledge aspect of the students can be assessed through written test, oral test, and assignment which depends on the competency that wants to be achieved. Thus, based on this regulation, the teachers can test the students' knowledge through written test.

In Indonesia, the type of written test that is commonly used for assessing the students' knowledge is multiple-choice test. Multiple-choice tests have been used extensively in many years for assessment purposes (Roberts, 2006). According to Toksöz & Ertunç (2017), multiple-choice test can help assessing the four competencies of English that are needed to be mastered by the students. The very common examples of multiple choice tests that have been used extensively are TOEFL, IELTS, and TOEIC. According to Hameed, et al., (2005), besides being used to measure application and analysis, multiple choice tests are good assessment tool for measuring knowledge and comprehension.

Since it is used to assess the students' knowledge, the multiple choice test is expected to be high in quality by following certain standard. The process of developing the items of the instrument should follow the norms of making a good multiple-choice test. According to Burton et al., (1991), the quality of multiple-choice test can be seen from the norms that are used in the process of constructing the test. In line with this statement, Haladyana (2004) states that a set of guidelines or norms should be adopted in writing items of multiple choice test. He also suggests a set of norms which consist of 4 dimensions. Not only by Haladyana (2004), a set of norms is also suggested by Hall and Marshall (2013) and Puspendik Kemendikbud (2019).

This set of norms is expected to be implemented in the assessment in Indonesia. SMPN 4 Singaraja is one of junior high schools in Buleleng regency which use multiple choice test made by the classroom teacher as summative assessment for middle test for English subject. Based on the pre-observation data, the achievement level of the students in SMP N 4 Singaraja, especially in English subject, is considered low. According to Puspendik Kemendikbud (2019) about the national examination result, the average score of national examination of English subject of SMPN 4 Singajara in 2018/2019 academic year is 52.15 which mean that it does not meet the minimum standard score of national examination which is 55.00.

Besides, it shows that the middle test score of seventh grade students in odd semester 2019/2020 is low that most of the students have to conduct remedial test. It can be concluded that the students in SMP Negeri 4 Singaraja has low achievement since they could not pass the standard of the examination that is seen from students' national examination score and their middle test's score.

Besides indicating students' achievement, the national examination scores also indicate to the implementation of assessment practice in SMP Negeri 4 Singajara. According to Black and William (1998a), a good mastery of materials that have been taught in the class is resulted by a good implementation of assessment. Thus, the low achievement level of the students can be caused by assessment practice that needs to be improved. Since the students' English achievement is low, it is presumably that the assessment practice implemented by the teachers in SMP Negeri 4 Singajara needs to be improved.

It is also proven by the pre-observation data which shows that even blueprint is not provided in the process of constructing the instrument. However, there is no further investigation about the quality of the teacher-made multiple-choice test in SMP Negeri 4 Singaraja. Considering the significant roles of the multiple-choice test as the instrument for summative assessment, a study which tries to investigate the quality of the test must be conducted. It is because the norms can be vital in ensuring that the teacher-made MCTs have reflected the learning objectives and have paid attention to the details. Thus, this study tries to investigate the teacher-made multiple-choice tests that are used as summative assessment for English subject at SMP Negeri 4 Singaraja. The study investigates the quality based on a total of 18 norms suggested by Haladyna (2004), Hall and Marshall (2013), and Puspendik Kemendikbud (2019) as guidelines in developing a good MCT. This study aims to investigate whether or not the teacher-made MCTs are high in quality in reference to the norms of making a good MCT.

METHOD

The design of this study is descriptive study. Descriptive research is a research design that is concerned in describing a certain phenomenon and its characteristics to provide more in-depth understanding and examination of the phenomenon itself (Nassaji, 2015). In descriptive research, the data is collected through test, questionnaires, interviews, or observations (Atmowardoyo, 2018).

This study was conducted in SMP Negeri 4 Singaraja and took four multiple-choice tests made by the English teachers as the subject of the study. The object of the study is the quality of the multiple-choice tests that was seen from congruity of the multiple choice tests with the norms in making a good multiple-choice test.

In this study, data were collected by using a set of method which are document study and interview as well as instruments of data collection which are checklist and interview guide. Document study is the method that will be used in investigating the congruity of the items of multiple choice tests with the norms of making a good multiple choice test. Checklist and anecdotal records are used as the instrument. According to Stufflebeam (2000), a checklist is one of instruments that is very useful used not only in planning and guiding but also in assessing

the outcomes. Anecdotal record is used to get more accurate and detailed data. Anecdotal records are more specifically detailed and naturalistic which can give meaningful information (McFarland, 2008).

Interview is done in order to get further information and explanation related to the congruity. According to Berg (2007) interview does not only provide detailed information but also enable interviewee to express their thoughts and feelings. The instrument that is used is interview guide as the guidance to conduct an interview with the English teachers in SMP Negeri 4 Singaraja. During the interview, recorder will be used in order to get the clear data.

The checklist is used in comparing each item of the teacher-made multiple-choice tests with the norms suggested by Haladyna (2004), Hall and Marshall (2013), and Puspendik Kemendikbud (2019) that were synthesized that turned into 18 norms with 4 dimensions which are content guidelines, style and format, writing stem, and writing option.

After the total of the 125 items were compared by using the checklist, the data was then analyzed by formula suggested by using Nurkancana & Sunartana (1992). Then, the results of data were calculated and classified to some classifications. The classifications were determined by using the following formulas:

Tabel 1. Data Classification Formula

Interval	Criteria
$Mi+1.5S \leq x$	Very Good
$Mi+0.5S \leq x < Mi+1.5S$	Good
$Mi-0.5S \leq x < Mi+0.5S$	Sufficient
$Mi-1.5S \leq x < Mi-0.5S$	Poor
$1 \leq x < Mi-1.5S$	Very Poor

The ideal mean (MI) and the standard deviation (SDI) scores are calculated as follow:

$$Mi =$$

$$\frac{\text{max score} + \text{min score}}{2} = \frac{18-0}{2} = 9$$

$$S = \frac{Mi}{3} = \frac{9}{3} = 3$$

RESULTS AND DISCUSSION

Teacher-made multiple-choice tests in SMP N 4 Singaraja are varied in term of the number of item. There are 30 items in MCT grade VII, 40 items in MCT grade VIII, 30 items in MCT grade IX A and 25 items in

MCT grade IX B. All of those 125 items were being analyzed to find the quality by analyzing the congruity between each item and the norms of making a good multiple-choice test. The result shows that the percentage of the norms fulfilled are varied.

Tabel 2. The Number of Items Fulfilling Norms

Percentage	Number of item
100%	34
Above 90%	44
Above 80%	39
Above 70%	8

The items have different number of norms being unfulfilled which ranges from 0-5 norms. Thus, the percentages of norms fulfilled are varied from the highest which is 100% to the lowest which is above 70%. Out of 125 items, there are 34 items (27%) which completely fulfilled all the norms. Neglecting 1 norm, 44 items (35%) are considered to have above 90% norms fulfilled. There are 39 items (31%) that are considered to have above 80% norms fulfilled since they neglect 2-3 out of 18 norms. Neglecting 4-5 norms, there are only 8 items (7%) that are considered to have the lowest percentage of norms fulfilled which is above 70%.

Completely following all the norms in making a good multiple-choice test, 34 items (27%) cover the highest percentage which is 100%. These items follow all norms in each dimension which are content guideline, style and format, writing stem and writing option are also fulfilled. The most common unfulfilled norm is about punctuation and capitalization. The problem is caused by different issues which are: 1) blank space that needs to be filled with a phrase or clause is not consisted of one ellipsis and ended with a full stop in the stem without any period in the options, 2) blank space that needs to be filled with a sentence is not consisted of one ellipsis in the stem and one full stop in each of the options, 3) blank space that needs to be filled with an interrogative sentence should be

consisted of one ellipsis in the stem and one question mark in each of the options, or 4) the options are not capitalized and ended with full stop when it is supposed to be so.

Relating to the norm about plausibility of distracters, the problems happen because there is no relevancy between the options and what being asked is. For instance, what being asked is about one person yet the option refers to two names of person. The problems about homogeneity happen because there is no consistency in the options about the grammatical structure, mostly in the used of part of speech. The problems in overlapping options happen because there are some options in one item which has the same contextual meaning. Relating to the norm about grammar, the problems are mostly caused by the error in using singular and plural form of nouns. Meanwhile, the errors in the placing order of the options happen because the options on items which require the test taker to choose between numbers on the option are not placed orderly from the smallest to the highest and vice versa, or the options with several expressions of daily greeting are not placed orderly from good morning to good night when it is supposed to be so.

Having only 1 norm unfulfilled, there are 44 items (35%) which have above 90% norms fulfilled. These items have different type of norm that is being unfulfilled. The

unfulfilled norm are varied which are about punctuation and capitalization, homogeneity, overlapping, plausibility of distracters, grammar, and placing order of options.

The items which have above 80% norms fulfilled neglect 2-3 out of 18 norms. Although they have the same number of norms fulfilled, they have different issue in term of the types of norm that are being unfulfilled. Those unfulfilled norms are the same with the previous classification of percentage. However, there are other norms that are being unfulfilled by the items in this category of percentage. They are clue, subjectivity, length of the options, and spelling. The problems about clue are caused by 3 different issues which are: 1) the correct answer is the only option which means contextually positive while the other options are negative and vice versa, 2) the correct answer is directly stated in the previous item, or 3) the correct answer is the only option that is given a full stop when it is not supposed to be so.

Relating to norms about subjectivity, there are some items which require the test taker to give his/her opinion about the correct answer. There are also some problems about misspelling such as the word cannot that is spelled incorrectly as can not and apologize as

apologise. The problem about length of options happen because one option has many more words than other options which makes the length becomes much longer than the other options. In some cases, this problem is automatically caused the problem of placed order of the options where it should be placed in option a or d.

Neglecting 4-5 norms, there are 8 items (7%) which are considered to have above 70% norms fulfilled. The unfulfilled norms are the same with the previous classification of percentage. However, in this percentage, there is one more unfulfilled norm which is about item clear focus. The problem happen because there is a dialogue provided in the item yet it is not clear who the speakers are.

Based on the result of item analysis above, it can be concluded that most of the items have problem on the punctuation and capitalization causing this norm covers only 38% items following it. It is the only norm which has such a low number of items following it which is only 48 out of 125 items. The rest of the norms have more than 100 items following it resulting the percentage of norms fulfilled range from 82%-100%. The detail of the percentage of norms fulfilled can be seen in Table 3.

Table 3. The Percentage of Fulfilled Norms

Norms Number	Norms' Description	Number of Item Fulfilling Norm	Frequency (%)
1	Reflecting basic competencies	125	100%
2	Not depending on the previous options	125	100%
3	Giving clear focus	124	99%
4	Avoiding opinion based items	123	98%
5	Being grammatically correct	117	94%
6	Having correct spelling	119	95%
7	Not containing clues	114	91%
8	Options are formatted vertically	125	100%
9	Taking concern on the use of punctuation and capitalization	48	38%
10	Not containing double negatives	125	100%
11	Options are homogeneous in content and grammatical structure.	113	90%
12	Having one correct answer	125	100%

13	Options have about the same length.	114	91%
14	Options are placed in logical and numerical order.	103	82%
15	Options do not repeat the same words or phrase.	125	100%
16	Options are not overlapping.	119	95%
17	Distracters are plausible.	113	90%
18	Not using “ <i>none of the above</i> ” or “ <i>all of the above</i> ”	125	100%

Norm about punctuation and capitalization is the most unfulfilled norm. Truss (2003) states that proper punctuation leads appropriate meaning and understanding. In line with this statement, Connelly (2009) argues that incoherent use of punctuation can cause ambiguity. Based on the interview, it shows that the accuracy in using punctuation needs to be improved. The issue is consistently in the use of ellipsis (...), which signs that there is a part of a sentence that has been omitted, especially about where and when to place it together with a period (.)

Mann (2003) claims that there is a difficulty in applying punctuation marks. This statement is in line with the result of a research done by Kurniawan et al., (2014) which shows that the ability of teachers in Indonesia in understanding the use of

punctuation is considered low. Different with norm about punctuation, the norm about reflecting basic competency, independency, option style and format, double negative, correct answer, word repetition, and phrase all of the above or none of the above are perfectly followed by all of the items resulting these norms covers 100% items following it.

Even though there are only 6 out of 18 norms that have 100% items following it, the other norms also have high number of item following it. Except norm about punctuation, the rest of the norms have more than 100 items following it. Each item also has many norms being unfulfilled. To analyze the quality of the teacher-made multiple-choice tests, formula suggested by Nurkancana and Sunartana (1992) is being applied. The criteria of the test quality can be seen in Table 4.

Table 4. The Criteria of Test Quality

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

There are five criteria of the quality which are very good, good, sufficient, poor, and very poor. The multiple-choice test which has percentage more than or equal to 75% is considered as very good, percentage less than 75% and more than or equal to 58% is considered as good, less that 58% and more than or equal to 42% is considered as sufficient, less that 42% and more than or

equal to 25% is considered as poor, and percentage less than 25% is considered as very poor.

In general, the quality of the teacher-made multiple-choice tests is considered very good since the majority of the items achieved more than 75% of multiple-choice test's quality criteria in the formula. The result of the judgment showed that only 1 item has

percentage less than 75% which is 72% so that this item is considered as good item.

Besides, the quality of the multiple-choice test in general can be seen from the location of the correct answer and the clarity of the instruction. The location of the correct answers is varied and is not mainly put in 1 option. The position of the right answers is assigned randomly so it does not form a pattern which gives clue to the students. Therefore, it supports the quality of the multiple-choice tests to be very good.

Meanwhile, the instructions are not clear enough. Some of the instructions do not clearly give the information about how to do the test. There are 2 tests which instruct the students to cross the options of the correct answer when the students actually have to rewrite the answer on their answer sheet. Besides, there is also a test which does not provide instruction about information of which items are a text for. Other than that, the instructions about information of which items are a text for in the rest of the multiple-choice tests are clearly stated.

This result of quality analysis is in line with the satisfaction of the teacher. Based on the interview, the teachers were confirmed to feel satisfied with their own works. Their knowledge about making a good multiple-choice test is derived from three factors. First, the teachers are graduated from English Education Department which teaches specifically about making a multiple-choice test in Assessment course. Second, the teachers ask the other teacher who had opportunity to join a workshop about making a good multiple-choice test. Third, the teachers learn how to make a multiple choice test by analyzing each item in National Standard School Examination (USBN).

First, the knowledge about making a good multiple-choice test is derived from college years. When the teachers took their bachelor degree at English Language Education at Ganesha University of Education, they had the opportunity to learn about how to make a multiple-choice test. Based on interview, the teachers remember some important aspects in making a good multiple-choice test such as using basic competency as the basis in making the items,

the format of the multiple-choice test, and some aspects that must be avoided such as the use of double negative in stem and the use of phrase all of the above or none of the above.

Second, the teachers ask the other teacher who attended a workshop about making a multiple-choice test. Workshop about making a good multiple choice test is held every year in Buleleng. The government conducted a Middle School Test Analysis Workshop, in order to train and prepare subject teachers in preparing items for the National Standard School Examination (USBN). However, the teachers in SMPN 4 Singaraja, who make multiple-choice test as a middle test, do not have the opportunity to attend the workshop yet.

The third is by analyzing each item in National Standard School Examination (USBN). Based on interview, the teachers believed that the quality of each item in National Standard School Examination (USBN) is guaranteed. Therefore, they make it as a basis in making their midterm and final examination test which are also in form of a multiple-choice test.

Those are the factors which influence the quality of teacher-made multiple-choice test in SMPN 4 Singaraja to be very good. The quality of the multiple-choice test is not aligned with the national examination result of SMP N 4 Singaraja for English subject. The average score of national examination of English subject of SMPN 4 Singajara in 2018/2019 academic year is 52.15 which is below the minimum standard score (55.00). It means that the quality of teacher-made multiple-choice test is not the factor that influences students' national examination result. There are other factors which seem to influence the performance of the students in National Standard School Examination (USBN) which are the competency of the students itself and the teacher performance.

Undeniably, the average score of national examination is the reflection of the knowledge and ability of the students. Based on interview, to verify that the quality of teacher-made multiple-choice test is not the influential factor which affects the average score national examination, the teachers stated some significant factors which might be the

reason of the low average score of national examination of English subject of SMPN 4 Singajara which are 1) the lack of vocabulary mastery; 2) the lack of motivation to learn; 3) family background; 4) the anxiety while taking the national examination; and 5) the application of computer-based test in national examination.

Even though the quality of the multiple-choice tests is considered to be very good, it is presumably that the lack of vocabulary mastered by students affects to their performance in national examination resulting to the low average score. A learner will not perform well in every aspect of language if a person does not have sufficient vocabulary size (Susanto, 2017). Vocabulary plays an important role on student success (Baker et al., 1997). This statement is strengthened by Marzano & Pickering (2005) who agrees that vocabulary is one of the key indicators of students' success in school especially on standardized tests.

Based on interview, the vocabulary mastery of students in SMP N 4 Singaraja needs to be improved. The lack of vocabulary is caused by the low exposure of the students in using the vocabulary itself. While teaching in class, the teachers use the students' native language to make them easier to understand the material. Besides, the type of vocabulary used in making multiple-choice test as middle test is made to be in accordance to the students' level instead of constructing it to meet the standard of national examination test.

The vocabulary used in national examination is varied and definitely different with what they got in middle test. The texts provided in national examination tend to be longer and the vocabulary is more complex. In this situation, the students hardly understand the question and find the correct answer causing them to get low score. It happens because the students are accustomed to get vocabulary that is not in accordance with the vocabulary in national examination provided in the multiple-choice test that is used for their middle test.

Besides the students' lack of vocabulary, to magnify that the quality of teacher-made multiple-choice test is not the

influential factor of the average score of national examination, student learning motivation is argued to be the other possible factor. According to Vansteenkiste et al., (2005), motivation has been shown to positively influence students' academic performance. Based on interview, the teachers stated that most of the students in SMP N 4 Singaraja do not have motivation and self-belief that they will be able to master English especially the students in parallel class. A study conducted by Kusukar et al., (2012) shows that students with autonomous motivation (motivation that originates within an individual) perform better than students with controlled motivation (motivation that originates from external source). Therefore, no matter how good the quality of the teacher-made multiple-choice test is, this low autonomous motivation must affects to their performance in national examination resulting to low average score.

Since the quality of the teacher-made multiple-choice test is not aligned with the national examination result of SMP N 4 Singaraja, teachers argued that family background is one of the significant factors which influence the average score of national examination. According to Li et al., (2018), the higher the social-economic status of the family, the more the participation of parent in their children's education is. Based on interview, the majority of student in SMP N 4 Singaraja comes from middle to low socio-economic status whose parents work as a farmer. Only a few parents are aware of the development of education that must be faced by their children as a student. Besides, parental participation also affects to the vocabulary exposure of their children during childhood. Hart and Risley (1995) found that children in lower socioeconomic classes experience less vocabulary than children in higher socioeconomic classes.

Not only through the differences in parental education participation, socio-economic status also become one influential factor through the differences in learning opportunities. Low socio-economic status makes them do not have the opportunity to just do their homework or even take private lessons. Based on interview, rather than do

their school homework, the teachers stated that some of their students must free their time to help their parents to work. A study conducted by Jez et al., (2013) found that increasing students' learning time resulting them perform better in standardized test which means that learning time is positively affects to the average score of national examination.

The other factor which affirms that the quality of teacher-made multiple choice test is not the factor which influences the students' score in national examination is the anxiety of the students while taking the national examination. The atmosphere while taking middle test must be different with the one in national examination. Based on interview, this anxiety leads to nervous students causing them to be unfocused to answer the test. This statement is strengthen by Jin, et al., (2014); Karatas, Wiryani et al., Alci & Aydin (2013) who agree that student anxiety in taking examination is a common issue happened in every country and every level of education. A research conducted by Ratih et al., (2012) found that out of 153 high school students in Jakarta, 1Therefore, the quality of the multiple-choice test used for middle test is not an influential factor since it concerns about the atmosphere in national examination which is different with the middle test.

Beside those factors, the other factor which shows that the quality of the multiple-choice test is not the factor which influences the average score of national examination is the ability of students to operate a computer. Students in SMP N 4 Singaraja no longer take examination with paper and pencils but use computers instead. This requires students to be able to login, view the texts in each item, and submit their answers in computer. However, based on interview, most of the students do not even know the basics in operating a computer. This problem relates to the previous factor which is student anxiety. Non-fluency in using computers increases the level of student nervousness because the situation during the examination is very different from what they get used to.

Besides the factor from the student itself, to affirm that the quality of teacher-made multiple-choice test is not the influential factor of the average score of national

examination, teacher performance is argued to be the other possible factor. Rockof (2004) states that teacher performance also significantly affects student performance. The result of the interview shows that the implementation of assessment practice in SMP N 4 Singaraja needs to be improved.

Instead of constructing items that meet the standard of national examination test, the teacher constructed the items in accordance to the students' level. The teachers argued that when their students successful in taking the quizzes and middle test even though the item difficulty is lower than the national examination, it will motivate the students to learn English more as they get high score in the test. In fact, this strategy does not seem to help the students to perform better in national examination since the average score of national examination is low. It might happen because the comfort provided by the teachers causes the absence of students' learning anxiety in achieving their goal which is to be success in national examination.

A study conducted by Strack et al., (2017) found that student who feels anxious in learning keeps them focus in achieving their learning goal. Instead of demotivating the students, constructing multiple-choice test which has the same level of difficulty with national examination will raise the students' learning anxiety causing them have the eagerness to learn more since they want to be success in national examination. A study conducted by Elmelid et al., (2015) showed that anxiety symptoms were positively correlated with higher academic motivation which was measured by students' positive attitudes toward learning and school. However, the multiple-choice test made by the teacher that is used for middle test does not promote the anxiety of the students which leads to unconcerned and unwillingness to learn.

The teachers argued that by not placing the students in anxious situation makes them willing to learn in class especially for English subject. Therefore, the teachers completely try to use teaching strategy which makes them easier to understand the material such as by using the students' native language while teaching in class. This teaching strategy which

provides low English exposure leads to the lack of vocabulary of students influencing to their performance in national examination. According to McGregor et al., (2007), when students had greater frequency of vocabulary exposure, greater number of vocabulary will be possessed which leads them to more likely to understand and remember the targeted words.

The result of the analysis shows that the high quality of teacher-made multiple-choice test is not aligned with the students' achievement in national examination. There are other possible factors that seems to contribute to the average score of national examination for English subject in SMP N 4 Singaraja which includes students' competency and teacher performance.

The teacher-made multiple-choice tests used for English middle test for at SMP N 4 Singaraja have a very good quality. However, there are some norms that are unfulfilled which range from 0-5. Therefore, relating to the construction of the multiple-choice test, some improvement is needed by paying more attention specifically for the unfulfilled norms.

CONCLUSION

As a conclusion, the teacher-made multiple-choice tests in SMP N 4 Singaraja have followed the norms in making a good multiple-choice test. However, some norms were unfulfilled by each item which ranges from 0-5 norms. In general, the quality of the teacher-made multiple-choice tests is considered very good since the majority of the items achieved more than 75% of multiple-choice test's quality criteria in the Nurkencana and Sunartana (1992) formula. The result of the judgment showed that only 1 item has percentage less than 75% which is 72% so that this item is considered as good item. Since there are some norms that are still neglected, more attention is needed to be applied for these norms. It is expected that by considering the norms in constructing a multiple-choice test, the quality of a multiple-choice test can be well maintained.

Based on the results of this research, there are some recommendations that can be

given to the teachers as the teachers, the lectures, and the other researchers. For teacher, it is suggested that more attention is needed to be put to these norms. Further, it is suggested to the teachers to attend related workshop. Besides, it is suggested that the level of difficulty of the multiple-choice tests is made to be met the standard of national examination and higher vocabulary exposure is provided to prepare the students in taking the national examination test. For lecturers who have the responsibility to conduct public services is suggested to conduct a related workshop in order to enhance the teachers' knowledge in assessment. For other researchers, it is suggested to conduct further research related to the quality of multiple-choice test that is seen not only from the norms but also other certain standard in making a good multiple-choice test. Moreover, it is recommended to investigate other possible factors that contribute to the students' performance in national examination.

REFERENCES

- Atmowardoyo, H. (2018). Research Methods in TEFL Studies: Descriptive Research, Case Study, Error Analysis, and R&D. *Journal of Language Teaching and Research*. 9(1). 197-204. DOI: <http://dx.doi.org/10.17507/jltr.0901.25>
- Baker, S., Simmons, D., & Kame'enui, E. (1995). *Vocabulary instruction: Synthesis of the research* (Technical Report No. 13). Eugene, OR: National Center to Improve the Tools of Education.
- Berg, B. L. (2007). *Qualitative research methods for the social sciences*. London: Pearson.
- Black, P. & D. William (1998a). "Assessment and Classroom Learning," *Assessment in Education: Principle, Policy, and Practice* 5(1): 7-73
- 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 paragraphs*. Cengage Learning
- Elmelid, Andrea & Stickley, Andrew & Lindblad, Frank & Schwab-Stone, Mary & Henrich, Christopher & Ruchkin, Vladislav. (2015). Depressive symptoms, anxiety and academic motivation in youth: Do schools and families make a difference?. *Journal of adolescence*. 45. 174-182. 10.1016/j.adolescence.2015.08.003
- Haladyna, T. M. (2004). *Developing and Validating Multiple-Choice Test Items*. Mahwah, New Jersey, London: Lawrence Erlbaum Associates.
- Hall, and Marshall. 2013. *A Guide for Developing Multiple Choice and Other Objective Style Questions*. Centre for Academic Development, Victoria University of Wellington, New Zealand.
- Hameed AA, Al-Faris EA, Alorainy IA. (2005). *The criteria and analysis of good multiple choice questions in a health professional setting*. *Saudi Med J*.26(10):1505–1510.
- Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Paul H. Brookes Publishing Company
- Jez, Su & Wassmer, Robert. (2013). The Impact of Learning Time on Academic Achievement. *Education and Urban Society*. 47. 284-306. 10.1177/0013124513495275.
- Jin, Y., He, L., Kang, Y., Chen, Y., Lu, W., Ren, X., . . . Yao, Y. (2014). Prevalence and risk factors of anxiety status among students aged 13-26 years. *Int J Clin Exp Med*, 7 (11), 4420- 4426. Diperoleh dari <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4276221/>
- Karatas, H., Alci, B., & Aydin, H. (2013). Correlation among high school senior students' test anxiety, academic performance and points of university entrance exam. *Educational Research and Reviews*, 8 (13), 919–926. doi: <http://dx.doi.org/10.5897/ERR2013.1462>
- Kemendikbud, 2014. *Implementasi Kurikulum 2013*. Jakarta: Kementrian Pendidikan dan Kebudayaan RI.
- Kurniawan, O., Noviana, E., Muhammad, N., (2014). ANALISIS KEMAMPUAN GURU SEKOLAH DASAR DALAM MEMAHAMI KONSEP PENGGUNAAN TANDA BACA SE-KECAMATAN TAMPAN PEKANBARU. *Jurnal Primary Program Studi Pendidikan Guru Sekolah Dasar Fakultas Keguruan dan Ilmu Pendidikan Universitas Riau* | Volume 3 Nomor 1, April 2014 | ISSN: 2303-1514
- Kusurkar RA, Ten Cate TJ, Vos CMP, Westers P, Croiset G. How motivation affects academic performance: a structural equation modelling analysis. *Adv Heal Sci Educ*. 2013;18(1):57–69. doi: 10.1007/s10459-012-9354-3.
- Li, Z., Qiu, Z. How does family background affect children's educational achievement? Evidence from Contemporary China. *J. Chin. Sociol*. 5, 13 (2018). <https://doi.org/10.1186/s40711-018-0083-8>
- Mann, C. (2003). Point counterpoint: Teaching punctuation as information management. *College Composition and Communication*, 54(3), 359-393.

- Marzano, R., & Pickering, D. (2005). *Building academic vocabulary: Teacher's manual*. Alexandria, VA: ASCD.
- McFarland, Laura. (2008). Anecdotal records: Valuable tools for assessing young children's development. *Dimensions; Journal of the Southern Early Childhood Association*. 36. 31-36.
- McGregor, K., Sheng, L., & Ball, T. (2007, October 1). Complexities of expressive word learning over time. *Language, Speech, and Hearing Services in Schools*, 38(4), 353–364. (ERIC Document Reproduction Service No.EJ776268). Retrieved August 18, 2009, from ERIC database.
- Ministerial of Education and Culture of Indonesia. (2016). Regulation of the Ministry of Education and Culture of the Republic of Indonesia Number 23 of 2016 Regarding the Standard of Education Assessment. Jakarta, Indonesia.
- Nassaji, Hossein. (2015). Qualitative and descriptive research: Data type versus data analysis. *Language Teaching Research*. 19. 129-132. 10.1177/1362168815572747.
- Nurkencana dan Sunartana. (1992). *Evaluasi Hasil Belajar*. Surabaya: Usaha Nasional.
- Puspendik Kemdikbud. (2019). Average graph of 2018/2019 school year grades. [Online]. Available: <https://hasilun.puspendik.kemdikbud.go.id/>
- Ratih, NK., Fitriyani, P., & Nurviyandari, D. (2012). Hubungan tingkat kecemasan terhadap koping siswa SMUN 16 dalam menghadapi ujian nasional. *FIK-Universitas Indonesia*, Depok, Jawa Barat.
- Roberts T S (2006), *The Use of Multiple Choice Tests for Formative and Summative Assessment*, Australasian Computer Education Conference (ACE 2006), Hobart, Tasmania, 16-19 January 2006.
- Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *The American Economic Review*, 94(2), 247-252.
- Strack, Juliane & Lopes, Paulo & Esteves, Francisco & Fernández-Berrocal, Pablo. (2017). Must We Suffer to Succeed?: When Anxiety Boosts Motivation and Performance. *Journal of Individual Differences*. 38. 113-124. 10.1027/1614-0001/a000228.
- Stufflebeam, D. L. (2000). *GUIDELINES FOR DEVELOPING EVALUATION CHECKLISTS: THE CHECKLISTS DEVELOPMENT CHECKLIST (CDC)* .
- Susanto, Alpino. (2017). THE TEACHING OF VOCABULARY: A PERSPECTIVE. *Jurnal KATA*. 1. 182. 10.22216/jk.v1i2.2136.
- Taras, M. (2005) Assessment Summative and Formative Some Theoretical Reflections. *British Journal of Educational Studies* , 53 (4), 466 478. <https://doi.org/10.1111/j.1467-8527.2005.00307.x>
- Toksöz, S., & Ertunç, A. (2017). Item Analysis of a Multiple-Choice Exam. *Advances in Language and Literary Studies*. 8. 141. 10.7575/aiac.all.v.8n.6p.141.
- Tosuncuoglu, Irfan. 2018. Importance of Assessment in ELT. *Journal of Education and Training Studies*. 6. 163. 10.11114/jets.v6i9.3443.
- Truss, L. (2003). *Eats shoots and leaves: The Zero Tolerance Approach to Punctuation*. New York: Gotham Books.

Vansteenkiste M, Zhou M, Lens W, Soenens
B. Experiences of autonomy and control
among Chinese learners: Vitalizing or

immobilizing? *Journal of Educational
Psychology*. 2005;97(3):468–483. doi:
10.1037/0022-0663.97.3.468.
[CrossRef] [Google Scholar]