

## The Analysis Of The Quality Of Teacher-Made Multiple-Choice Tests Used As Summative Assessment For English Subject

Ni Wayan Vina Krisna Yanti<sup>1</sup>, Anak Agung Gede Yudha Paramartha<sup>2</sup>, Luh Gede Eka Wahyuni<sup>3</sup>

<sup>1,2,3</sup>English Language Education, Universitas Pendidikan Ganesha,  
Singaraja, Indonesia

e-mail: e-mail:wayan.vina.krisna@undiksha.ac.id, yudha.paramartha@undiksha.ac.id,  
ekawahyuni<sup>3</sup>@undiksha.ac.id

### Abstract

*This study arose in regards to the importance of constructing high quality multiple-choice tests (MCTs) that follow the norms in making a good MCT. The norms are considered important as they can make the tests relevant to the learning objectives and ease the test-takers in taking the test. This is a qualitative document analysis study which aimed to investigate the quality of teacher-made MCTs used as summative assessment for English subject at SMP Laboratorium Universitas Pendidikan Ganesha. 3 teacher-made MCTs with 40 items for each test were taken as the samples that represent each grade. The data were collected through document study by comparing each of the teacher-made MCTs item with the norms in making a good MCT that were then clarified through interview. The comparison is then classified in order to determine the quality. The results show that the quality of the teacher-made MCTs are very good with 106 items (88%) qualified as very good and 14 items (12%) that qualified as good. However, there are some norms that need more attention as they are rarely fulfilled.*

**Keywords:** *Summative Assessment, Teacher-Made Multiple-Choice Test, Instrument Quality, Norms*

### 1. Introduction

Assessment is essentially important in teaching and learning process. It is because assessment is a process of evaluating the students' progress throughout the learning process by gathering the information needed through various methods (Gronlund, 1982; Hanna & Dettmer, 2004; Anandan, 2015). Assessment is divided into two types based on the functions; formative and summative assessments. According to Brown (2004), formative assessment is used to evaluate the students' progress in the process of learning, while summative assessment aims to measure what the students have grasped after a series of learning process. Thus, summative assessment is typically conducted at the end of a series instructions, which can be in form of middle test and final test (Chappius & Chappius, 2007); States *et al.*, 2018). Not only to merely evaluate the learning process, assessment becomes essential as it is used to measure students' improvement, rank the students, and motivate them to learn (Jones, 2005; Jabbaarifar, 2009).

Regarding the significant roles of assessment in education, Indonesia government through the Ministry of Education and Culture has established an act regulating the process of assessment in Indonesia's educational system. It is regulated in The Ministry of Education and Culture Regulation No. 23/2016 that is used as the reference for assessment standard in 2013 curriculum, the current curriculum used in Indonesia. The act regulates on article 9 paragraph 1 and item (c) that there are three aspects that must be assessed. They are attitude, knowledge, and skill. It is stated that the knowledge aspect can be assessed through written test, oral test, and assignment based on the competency that want to be achieved. Based on this regulation, the teachers surely can test the students' knowledge through written test and use it as summative assessment for middle test and final test.

One of the instruments for written test that is commonly used in Indonesia to assess the students' knowledge is multiple-choice test (MCT). Zimmaro (2016) suggests that MCT is very useful to measure the students' knowledge outcomes, which is in line with what is stated in the regulation. The use of MCT is also very effective since it requires less time in the preparation and the implementation (Adeel, 2005), which could be very beneficial for overloaded teachers. Öztürk (2007) argues that MCT seems to be more reliable compared to other types of tests that can be negatively subjective. This is supported by Tsgari (1994),

Egbule (2002), and Cheng (2004) who both find that MCTs are less discriminating than free responses tasks. Toksöz & Ertunç (2017) even argue that MCT can assess all English basic skills through appropriate response questions. Therefore, it is not strange if MCT has been used for many years by teachers in Indonesia. The advantage of MCT makes it used highly in Indonesia, especially for summative assessment as middle test, final test, and even national examination.

Since it is used as summative assessment, the MCT must be high in quality. As stated by Anderson and Morgan (in Fiktorius, 2014), the quality of any assessments in any educational settings is the result of the quality of the instrument used as the basis for decision making. It is because at the end of the day, the result of the instrument will take the higher percentage in the decision making of a summative assessment. Therefore, the construction of the MCT must follow certain standards in order to achieve the high quality.

Burton *et al.*, (1991) suggest that the quality of an MCT can be seen from the norms that are used in constructing it. Haladyna (2004) supports the statement by stating that there must be a set of guidelines or norms that are used as the basis in developing the MCT. The norms are used to make the MCT becomes relevant to the competencies that want to be achieved and easy to be read by the students as the test-takers. It is because when taking the MCT, the students need to answer the test to their best knowledge while dealing with the allotted time. Thus, making the test more readable is essential. Moreover, Hall and Marshall (2013) argue that the norms become important as writing a good MCT requires skill, experience, and attention to detail. Therefore, writing the MCT with the reference to certain guidelines or norms is essentially vital.

Several studies have been conducted previously by Kheyami *et al.*, (2017), Paramartha, (2017), Sahoo and Singh (2017), Rehman *et al.*, (2018), Mahjabeen *et al.*, (2018), Kusumawati & Hadi (2018), and Rao *et al.*, (2018) to analyze the quality of MCTs used for academic testing. However, all of these study only focus on addressing the MCTs quality from the item discrimination, item facility and distracter efficiency. None of them seen the quality from the norms that are used in constructing the MCTs. Considering the importance of norms in making a good MCT to the quality of the MCT itself, a study that investigates the quality of an MCT from the norms that are used in constructing it becomes urgent to be conducted.

Further, to support the importance of MCT's writing guidelines, Haladyna (2004), Hall and Marshall (2013), and *Puspendik Kemendikbud* (2019) provide the list of norms of making a good MCT. Haladyna (2004) suggests that there are 31 norms with 4 dimensions, including content guidelines, style and format concerns, writing stems, and writing options. Hall and Marshall (2013) suggest a total of 12 norms, while *Puspendik Kemendikbud* (2019) suggests 16 norms with 3 dimensions, including the material, the construction, and the language. These theories show that the norms are very essential in constructing a good MCT.

Even *Puspendik Kemendikbud* (2019) as the center of educational assessment in Indonesia suggests that when the MCT's items are not in line with the norms, then the tests are considered low in quality. Therefore, every MCT especially those who are made and implemented in Indonesia are expected to follow the norms in making a good MCT in order to maintain the quality.

SMP Laboratorim Universitas Pendidikan Ganesha or known as SMP Lab Undiksha is one of the private junior high schools in Buleleng regency, Bali, Indonesia that uses teacher-made MCTs as summative assessment for middle test. Lebagi *et al.*, (2017) suggests that teacher-made MCT is made by teachers as the ones who know exactly how the students are in order to measure the students' mastery on specific materials that they have been taught. The students' English achievement in SMP Lab Undiksha is considered high since it got into the top ten of junior high schools with the highest average national examination score in Buleleng regency straightly in a row for the past 2016/2017, 2017/2018, and 2018/2019 academic years (*Puspendik Kemendikbud*, 2019). Even in the last national examination, SMP Lab Undiksha got 66.87 score which brought it into the top ten.

The pre-observation data also showed that most of the students could pass the middle test by passing the minimum score standard for English subject in SMP Lab Undiksha, which

is 70. Looking at the students' national examination score and their middle test's score, it can be concluded that the students in SMP Lab Undiksha has high achievement. It is because they could pass the national examination standard that has been a final national summative assessment for the students in their last educational stage year.

Besides indicating students' high achievement, the national examination scores also indicate that SMP Lab Undiksha have conducted good assessment practice. It is because a good assessment implementation results good mastery of materials that have been taught, which can further help students for other achievement related to the materials (Black and William, 1998a). Since the students' English achievement is good, the teachers are expected to have performed good assessment practice in which the tests used in their school's assessment reflected the basic competencies that appeared in the national examination.

The teachers in SMP Lab Undiksha used one MCT for one grade, which also shows a good assessment practice since the instrument used for measuring each student in each grade is the same. However, it is still unknown whether the teachers have followed specific norms in constructing the teacher-made MCTs to ensure the quality of the test. Considering the significant roles of the MCT 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, have good format, and have paid attention to the details.

Thus, this study tries to investigate the teacher-made MCTs that are used as summative assessment for English subject at SMP Lab Undiksha. 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.

## 2. Methods

Document analysis was used as the research method of this qualitative study. It is because this study aimed at investigating the quality of teacher-made MCTs that were used as summative assessment. According to Bowen (2009), document analysis is a systematic procedure for reviewing or evaluating documents that can either be in printed or electronic. This is a design in qualitative research where the document is examined and interpreted in order to extract the meaning, gain understanding, and develop empirical knowledge (Corbin & Strauss, 2008; Rapley, 2007). Regarding these characteristics, document analysis is considered as the appropriate method for this qualitative study.

This study was conducted at SMP Lab Undiksha and took 3 teacher-made MCTs as the subjects of the study and their quality as the object of the study. Document study and interview were the method of data collection while checklist and interview guide were the instruments. The interview was conducted in order to clarify the result of the document study as detailed information can be derived from the interviewees (Berg, 2017). Checklist was chosen as Stufflebeam (2000) suggests that a checklist is very useful in planning, monitoring and guiding its operation, and assessing its outcomes. Thus, checklist is considered appropriate as the instrument of data collection for this method.

The quality of the teacher-made MCTs was analyzed based on the norms in making a good MCT. Each item of the teacher-made MCTs was compared with the 18 norms suggested by Haladyna (2004), Hall and Marshall (2013), and *Puspendik Kemendikbud* (2019) in order to see the congruity.

The MCTs' items were firstly compared with the 18 norms using the checklist. After all of the 120 items were compared, the data was displayed in an organized way. Then, the data was analyzed through descriptive statistics as suggested in Nurkencana & Sunartana's book (1992). After that, the results of data tabulation were calculated and classified to some classifications. The classifications can be seen on 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

There are five categories of classification. They are very good, good, sufficient, poor, and very poor. The teacher-made MCTs are considered very good if the percentage is more than or equal to 75%. They are good if the percentage is more than or equal to 58% but less than 75%. They are sufficient when it is more than or equal to 42% but less than 58%. They are poor if the percentage is more than or equal to 25% but less than 42% and are very poor if it is less than 25%.

### 3. Results and Discussion

There are 40 items in each grade of the teacher-made MCTs with four options for each item. The total of the items are 120 items. The MCTs were analyzed in order to see the congruity between the MCTs' items and the 18 norms. The analysis results a total of norms that are being fulfilled by the items which was then converted to a percentage.

The percentages of the items are ranged since the numbers of norms being fulfilled are different. The highest percentage is 100% in which all of the norms are fulfilled and the lowest is above 60%. There are 8 items out of the total 120 items (7%) that accomplish 100% norms fulfilled. 24 items (20%) neglect only one norm and reach above 90%. 55 items (46%) have the percentage of above 80% as a result of neglecting two to three norms. 30 items (25%) neglect four to five norms and have above 70% norms fulfilled. At the least percentage, 3 items (2%) neglect six to seven norms and are categorized as above 60% norms fulfilled.

The items with 100% norms fulfilled have followed all of the 18 norms. The items have perfectly reflected the basic competencies that wanted to be achieved. They do not depend on the previous items, give clear focus, do not ask for students' opinion, have correct grammar and spelling, do not contain double negatives, and do not give clues to the right answers. They use correct punctuation and capitalization for both the stem and the options. The options are written vertically by considering their homogeneity, one correct answer, length, placing order, repetition, independency, and distracters plausibility. Further, they also do not use *all of the above* nor using *none of the above* on the options. These make the items clear as there is no confusion in answering them.

The issues that are faced by items that are above 90% are varied from punctuation and capitalization, distracters plausibility, options homogeneity, and options writing format. The distracters are not plausible as they are out of the items' context and cannot possibly fit in as the correct answer. Thus, they can be easily marked by the students as merely distracters. Meanwhile, the options become not homogeneous since they come from different parts of speech or are in forms of phrases, clauses, and sentences. The issue of options writing format happens because some of the items' options were formatted horizontally instead of vertically. Their punctuation and capitalization issues include 5 major points, which are (1) full stops use, (2) options and stems capitalization, (3) blank space at the end of the stem, (4) colon mark use, (4) question mark use, and (5) acclamation mark use.

The punctuation and capitalization issues appear to most of the items in this percentage category. The issue is majorly about the blank space at the end of the stems. Norm 9 that governs about punctuation and capitalization discusses on point (c) that if the blank space is at the end of the stem, the sentence of the stem is ended with a space and four full stops without space. The issue happens as the stems do not follow the norm and use excessive numbers of full stops, or use only three full stops and put the last one on the options.

Neglecting 2 to 3 norms, items with above 80% norms fulfilled have different combinations of norms issues. The issues are the same with the previous percentage category. Besides that, there are also other norms that are being unfulfilled by the items in this percentage category. They are items clear focus, grammar, options placing order, overlapping options, options length, spelling, options writing format, clue to the right answers, and one correct answer. The items are not focus since the stems are not in form of questions nor incomplete sentences or dialogues. They are more likely in form of riddle. There are also some items that are in form of dialogues but there is no information of who the speakers are, which make them not clear.

The items' grammatical issues mostly include the use of (1) tenses, (2) singular and plural noun, (3) verb choice, and (4) auxiliary verb. The lengths of the options are not quite the same as one option is way longer than the others. The lengths issue can affect the options placing orders as they are not placed in a logical or numerical order. The spelling issue is relatively minor as there are only few items that misspell some words.

The issue of overlapping options happens because there are two or more options that are actually the same but are written with different adjectives to make a difference, *sorry* and *so sorry* for instance. However, the adjectives do not give much difference as their meaning is still the same. There is no one correct answer because the correct answer becomes ambiguous by the existence of highly plausible distracters. On the other hand, the items are considered giving clues to the right answer by (1) having the length of the correct answer way longer than the distracters, (2) having the correct answer the only options that are relevant to the question, and (3) having the contextual meaning of the correct answer positive while the distracters negative, and vice versa.

The options writing format also appear as the most troubled norm. This happens mostly in the teacher-made MCT grade IX. Instead of formatting the options vertically, they were formatted horizontally. There are two cases of this formatting where the options were formatted fully vertical or half vertical.

Just like the previous percentage, each of the items that have above 70% norms fulfilled has different combinations of norms issues. Their problems are all the same as the previous percentage categories. It appears that the main issue of this percentage category is also punctuation and capitalization. In addition, this percentage category also has options homogeneity as the major problem. The options are not homogeneous since they come from different parts of speech or are in forms of phrase, clause, sentence, and interrogative sentence. They are supposed to be homogeneous in terms of content being discussed and parts of speech in which they need to be in the same category of speech or be phrase only, clause only, sentence only, and interrogative sentence only.

Having the least percentage category, items with above 60% items fulfilled are neglecting 6 to 7 norms. The norms that are being neglected are the same as the previous percentage categories. The major problems are also the ones that have been explained previously. Despite having the same major issues, the issue of distracters plausibility in this category is worth discussing. The distracters implausibility happened as the results of options that are not homogeneous and/or options that are grammatically incorrect. These make the distracters cannot possibly fit in as the correct answers and are easily identified as merely distracters.

Even though most of the norms being unfulfilled are relatively similar, there is a pattern of the norms that are fully fulfilled, highly fulfilled, and rarely fulfilled. All of the teacher-made MCTs' items have fully fulfilled norms number 1, 2, 10, 15, and 18 by reflecting the basic competencies, avoiding items' dependency on each other, not containing double negatives, not repeating words or phrases with the same meaning, and not containing *none of the above* or *all of the above*.

On the contrary, the ninth norm that governs about the use of punctuation and capitalization appeared to be the norm that is rarely fulfilled. Norms number 11, 8, 17 which are about options homogeneity, options writing format, and distracters plausibility also have relatively low percentages. The detail of the norms fulfilling percentages can be seen on Table 2.

Table 2. Percentage of Norms Fulfilled by the Teacher-Made MCTs Grade VII, VIII, and IX

Norms Number	Norms' Description	Number of Items Fulfilling the Norms	Frequency
1	Reflecting basic competencies	120	100%
2	Not depending on the previous options	120	100%
3	Giving content clear focus	110	92%
4	Avoiding opinion based items	119	99%
5	Being grammatically correct	96	80%
6	Having correct spelling	119	99%
7	Not containing clues	107	89%
8	Formatting options vertically	77	64%
9	Having correct use of punctuation and capitalization	34	28%
10	Not containing double negatives	120	100%
11	Having homogeneous options in content and grammatical structure	72	60%
12	Having one correct answer	115	96%
13	Having options with relatively the same length	103	86%
14	Placing options in logical and numerical order	99	83%
15	Not repeating options with the same words or phrase	120	100%
16	Not having overlapping options	113	94%
17	Having plausible distracters	84	70%
18	Not using "none of the above" or "all of the above"	120	100%

As described previously, norm number 9 about the use of punctuation and capitalization has the lowest fulfilling frequency with only 28%. The issues in the use of punctuation and capitalization appeared in many variations. Apart from the issues that have been explained previously, there is also inconsistency in the use of full stops and underscores as the blank space. Even though this is not regulated in the norms, this inconsistency shows that the punctuation use in the tests needs more attention. It is because the issues of punctuation relating to the norms have become the major issues that appear in most of the items and now it also has inconsistency in punctuation use that is out of the norm. This issue seems to happen to many teachers as Hall and Marshall (2013) state that teachers are often inconsistent in the way they punctuate MCT's stems and options. Thus, this even further proves that more attention on the punctuation use is needed.

The attention is needed because punctuation and capitalization uses are actually important. According to Kurniawan (2016), the use of punctuation is significant as it maintains clarity and avoids ambiguity in expressions used. Further, Kurniawan (2016) and Truss (2003) suggest that accurate use of punctuation and capitalization will help the readers, which in this case are the students, to understand what the writer exact intention is. That is the reason why *Puspendik Kemendikbud* (2019) suggests punctuation and capitalization as one of the norms in making a good MCT and even provides the details on their uses.

However, despite their importance, the issues of punctuation and capitalization seem to happen to many teachers, especially the ones in Indonesia. Ariyanti & Fitriana (2017), Nasser (2019), and Toba *et al.*, (2019) state that mechanics problem that usually appear in writing is the incorrect use of punctuation.

Besides punctuation and capitalization, norms number 11, 8, and 17 that govern about options homogeneity, options writing format, and distracters plausibility are also rarely fulfilled with below 75% norms fulfilling percentage.

The teachers argued that all of the issues appeared because of three main reasons. They are the absent of peer review, inadequate time, and other school administration procedures. The teacher argued that they were informed about the submission of the middle test MCTs a week before the due date. During the week, they were also busy with other administrative procedures that were relating to the mid-term assessment and were still busy teaching the students. The teachers claimed that these made them overworked and could not maximally make the MCTs. Further, after they had finished the tests, there was no peer review where they could ask their peer teachers to check their tests so that the mistakes could be minimized. These reasons were justified to be the causes of why the tests had issues with some norms.

By looking at the number of norms each item has fulfilled, the quality of the teacher-made MCTs was then analyzed. Then, the quality of the teacher-made MCTs' items was then classified into five categories. They are very good, good, sufficient, poor, and very poor.

The analysis shows that in general, the quality of the teacher-made MCTs at SMP Lab Undiksha is very good as 106 (88%) out of the 120 items are very good by achieving more than 75% percentage judgment and 14 items (12%) are considered good. The results are in line with the teachers' expectations as they expected their tests to be good in quality. This very good quality is influenced by many factors that mainly come from the teachers as the test-makers. The interview shows that there are three major factors that contribute to the very good quality.

First, the teachers based their tests on their prior knowledge of the MCT making guidance. The prior knowledge is a result of their undergraduate study at Ganesha University of Education in which they majored in English Language Education and were taught about assessment. Through the assessment course, they have learnt that in making good MCTs, there are some norms and procedures that needed to be considered as the guidelines. They know that they need to make blueprint for the tests and validate the tests.

Studies conducted by Lipson (1982), Yeh (2012), and Wessels (2012) showed that prior knowledge contributes positively to related activities that are done in the present. When a person has prior knowledge of a certain issue, it will be easier for her or him to bring the information needed to the surface where it is ready to be applied. Lipson (1982) further suggests that prior knowledge is a powerful factor that could help people acquiring totally new information while also correcting old information that was inaccurate. Thus, having a prior knowledge of the norms in making a good MCT does seem to help the teachers in constructing their MCTs.

Their prior knowledge also guides them in making instructions for the tests. Giving complete and clear instruction of how the tests must be taken is important as Zimmaro (2016) suggests that it can help the students to allocate their time and efforts wisely. The instructions include the time allotment, how the tests are taken, and items number for each text, picture, or table in the tests.

Two of the teachers already provided instructions of how the students must answer the tests by writing *"Please choose one of the best answer a, b, c, d!"* and *"Choose the best answer a, b, c, or d."* in the very beginning of the tests. However, these instructions are not clear enough since the students do not only need to choose the correct answers but also need to cross the correct answers on their answer sheets. On the other hand, one of the teachers did not provide the instruction and argued that she knew its importance but forgot to put it due to the limited time.

The instructions for each text, picture, or table are complete in the teacher-made MCTs grades VII and VIII. However, an instruction is missing in the teacher-made MCT grade IX for a text that was supposed to be for items number 27-29. There are also three text instructions with incorrect items number in the teacher-made MCT grade IX.

To further enhance the MCTs' quality, the teachers also worked on placing the position of the correct answers. It is as suggested by Hall and Marshall (2013) that distributing the

position of the correct answers randomly will avoid students who are in doubt from guessing the pattern of the correct answers. Thus, the teachers argued that they have made sure that the positions are varied and random so that there are no patterns that can be traced. Based on the analysis, the correct answers were really placed randomly and their positions are varied.

The existence of double negatives which are *except* and *not* in an item were believed to will only cause confusion for the students as the test-takers. It is in line with the statement made by Chiavaroli (2017) who states that double negatives should be avoided in making multiple-choice questions. Thus, all of the tests are free from double negatives. Further the teachers also consider that the phrase *none of the above* and *all of the above* do not help assessing the students' higher order thinking skills since they give opportunity for the students to just choose the phrases as the correct answer. Therefore, the teachers did not write the phrases on their tests

The second factor is the teachers' experience by being involved in making MCTs for the Junior High School English final test in Buleleng regency. One of the teachers had this prosperous opportunity and learnt so much from the event. This tests making process is based Permendikbud No. 4 2018 about Assessment of Learning Outcomes by the Educational Unit and Assessment of Learning Outcomes by the Government. The act suggests that school teachers under the consolidation of Subject Teacher Meetings (MGMPs), have the responsibility to prepare 75%-80% of National Standard School Examination (USBN). The preparation is done in coordination with the educational government in the district, city, or province. Thus, through this process, the teacher could learn from other experienced teachers and then shared the knowledge to her fellow teachers at SMP Lab Undiksha. This is supported by Irvine (2018) and Podolsky *et al.*, (2019) who find that teachers' experience that were gained throughout their careers can positively affect their performance in teaching, which includes assessment.

The third factor is related seminar attended by the teachers. The interview showed that one of the teachers had the opportunity to attend an educational seminar discussing MCT items that could assess students' higher order thinking skills. The teacher learnt how to make good indicators from a basic competency. The teacher also learnt that the test is not merely testing students' memory and understanding but also their ability in analyzing and evaluating. Studies conducted by Essien *et al.*, (2016), Alestre (2016), and Al-Adawi (2017) show that there exists positive relationship between the frequency of teachers' attendance at seminars or workshops with teachers' performance that is reflected by the students' academic achievement.

Through the seminar, the teacher argued that she now knows how to make good indicators from the basic competencies and then makes MCTs' items in accordance to them. This is supported by Sunardi and Sugini (2014) who suggest that attending workshop could result significant improvement for the teachers. Even though the other two teachers could not attend it due to the limited quota, they still got the knowledge from their fellow who has attended it. This is why they have put so much attention in making sure that their tests' items reflect the basic competencies.

Besides relying on their prior knowledge and experience, the teachers argued that they also tried to find other resources. The teachers stated that they have searched over the internet for guidelines of how the MCTs should be constructed. However, the resources that they found are not comprehensive and are different from one another. Thus, they really hoped to be able to attend related seminars or have reliable sources of information about the norms.

The analysis shows that the teacher-made MCTs' quality is aligned with the students' high achievement in the national examination. The teachers argued that this happens as the students are used to take MCTs as their summative assessment and as the tests reflected the basic competencies that appear on the national examination. Thus, they believed that MCTs that are good in quality will lead the students to be successful in taking the national examination. It further proves the reason why SMP Lab Undiksha got into the top ten of



junior high school with the highest average of national examination score in Buleleng regency for the past three academic years.

The quality of the teacher-made MCTs for English subject used at SMP Lab Undiksha is very good and seems to contribute to its students' high achievement in the national examination. It further indicates that the teachers have done a good assessment practice as suggested by Black and William (1998a) that a good assessment practice results good understanding of the contents that have been taught in the classroom. However, relating to the construction of the MCTs, more attention is needed to be put to some norms that are rarely fulfilled.

#### **4. Conclusion and Suggestions**

In conclusion, the quality of the teacher-made MCTs is considered very good as their items achieved high percentage of multiple-choice test's quality criteria in the formula. The majority of the teacher-made MCTs items are qualified as very good and very few are qualified as good. The items in the teacher-made MCTs have fulfilled the norms in making a good MCT. However, there are some norms that are rarely fulfilled. They are the use of punctuation and capitalization, options homogeneity, options writing format, and distracters plausibility. The results of the study indicate that the MCTs that were constructed by the teachers have had very high quality. However, there are still some issues relating to some norms. Thus, it is suggested that more attention is needed to be put to these norms. Further, peer reviews, rechecking, and editing are highly suggested to be conducted in order to minimize errors. As the teachers claimed that there are not many seminars relating to guidelines in making good MCTs or assessment in general, they hoped to be able to attend some in the future. Thus, for the lecturers who have the responsibility to conduct public services, carrying out related seminars are suggested in order to enhance the teachers' knowledge in assessment. The results show that there might be other possible factors that contribute to very good quality of the teacher-made MCTs. There are also many other aspects that can be analyzed about the students' high achievement in the national examination. Thus, conducting further researches are recommended.

#### **References**

- Adeel. (2005). *The Criteria and Analysis of Good Multiple Choice Questions in a Health Professional Setting*. Saudi Arabia: King Saud University.
- Al-Adawi, S. (2017). Exploring the Effectiveness of Implementing Seminars as a Teaching and an Assessment Method in a Children's Literature Course. *English Language Teaching*, 10. 1. 10.5539/elt.v10n11p1.
- Alestre, L. M. (2016). The Meaning of Trainings and Seminars to Public School Teachers: Some Stories to Tell. *International Journal of Advancements in Research & Technology*, 5(8).
- Anandan. (2015). *Assessment for Learning*. Tiruchirappalli: Bharathidasan University.
- Ariyanti, A., & Fitriana, R. (2017). EFL Students' Difficulties and Needs in Essay Writing. *Advances in Social Science, Education and Humanities Research (ASSEHR)*, 158, 111–121.
- Berg, B. L. (2017). *Qualitative research methods for the social sciences*. London: Pearson.
- Black, P. & William, D. (1998a). "Assessment and Classroom Learning." *Assessment in Education: Principles, Policy, and Practice*, 5(1). 7-73.
- Bowen, Glenn. (2009). Document Analysis as a Qualitative Research Method. *Qualitative Research Journal*, 9. 27-40. 10.3316/QRJ0902027.
- Brown, H. D. (2004). *Language Assessment: Principles and Classroom Practices*. White plains, NY: Pearson Education.
- Burton, S. J., Sudweeks, R. R., Merrill, P. F., & Wood, B. (1991). *How to Prepare Better Multiple-Choice Test Items: Guidelines for University Faculty*. Brigham Young University.
- Chappius, S., & Chappius, J. (2007). *The Best Value in Formative Assessment*.

- Cheng, H. (2004). A Comparison of Multiple-Choice and Open-Ended Response Formats for the Assessment of Listening Proficiency in English. *Foreign Language Annals*, 37, 544-553. 10.1111/j.19449720.2004.tb02421.x.
- Chiavaroli, Neville. (2017). Negatively-worded multiple choice questions: An avoidable threat to validity. *Practical Assessment, Research and Evaluation*, 22, 1-14.
- Corbin, J., & Strauss, A. (2008). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (3rd ed.). Thousand Oaks, CA: Sage.
- Egbule, J.F (2002). *Fundamentals of Test, Measurement and Evaluation*. Owerri Nigeria, Whyte and Whyte Publishers.
- Essien, E. E., Akpan, O. E., & Obot, I. M. (2016). The Influence of In-Service Training, Seminars and Workshops Attendance by Social Studies Teachers on Academic Performance of Students in Junior Secondary Schools Incross River State, Nigeria. *Journal of Education and Practice*, 7 (22).
- Fiktorius, T. (2004). *A Validation Study on National English Examination of Junior High School in Indonesia*. Pontianak: Tanjung Pura University.
- Gronlund, N. E. (1982). *Measurement and evaluation in teaching* (4th ed.). Cliffs, NY: Macmillan.
- Haladyna, T. M. (2004). *Developing and Validating Multiple-Choice Test Items* Mahwah, New Jersey, London: Lawrence Erlbaum Associates.
- Hall, C. & Marshall, S. (2013). *A Guide for Developing Multiple-Choice and Other Objective Style Questions*. Centre for Academic Development, Victoria University of Wellington, New Zealand.
- Hanna, G. S., & Dettmer, P. A. (2004). *Assessment for Effective Teaching: Using context-adaptive planning*. Boston, MA: Pearson A&B.
- Irvine, J. (2018). Relationship between teaching experience and teacher effectiveness: implications for policy decisions.
- Jabbarifar, T. (2009). The Importance of Classroom Assessment and Evaluation in Educational System. *Proceedings of the 2nd International Conference of Teaching and Learning*.
- Jones, A. (2005). *Assessment for Learning*. UK: Learning and Skills Development Agency.
- Kheyami, D, et al.(2017). Item Analysis of Multiple-Choice Questions at the Department of Paediatrics, Arabian Gulf University, Manama, Bahrain. *Sultan Qaboos Univ Med J*, 18(1), 68-74. doi:0.18295/squmj.2018.18.01.011.
- Kurniawan, A. (2016). *Graduate Students' Lived Experience in Writing Academic Papers*. Sanata Darma University: Yogyakarta.
- Kusumawati & Hadi. (2018). An Analysis of Multiple Choice Questions (MCQs): Item and Test Statistics from Mathematics Assessments in Senior High School, 4(1),70-78.
- Lebagi, D., Sumardi, & Sudjoko. (2017). The Quality of Teacher-Made Test in EFL Classroom at the Elementary School and Its Washback in the Learning. *Journal of English Education*, 2(2, Dec.), 97-104.
- Lipson, M. Y. (1982). Learning New Information from Text: The Role of Prior Knowledge and Reading Ability. *Journal of Reading Behavior*, 14(3), 243-261.
- Mahjabeen, W., Alam, S., Hassan, U., Zafar, T., Butt, R., Ansari, S. K., & Rizvi, M. (2018). Difficulty Index, Discrimination Index and Distractor Efficiency in Multiple Choice Questions. *Annals of PIMS*.
- 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.
- Ministerial of Education and Culture of Indonesia. (2018). Regulation of the Ministry of Education and Culture of the Republic of Indonesia Number 4 of 2018 Regarding the Assessment of Learning Outcomes by the Educational Unit and Assessment of Learning Outcomes by the Government.
- Nasser, S. M. (2019). Iraqi EFL Students' Difficulties in Writing Composition: An Experimental Study. *International Journal of English Linguistics*, 9(1), 178-184.
- Nurkancana and Sunartana. (1992). *Evaluasi Hasil Belajar*. Surabaya: Usaha Nasional.

- Öztürk, M. (2007). Multiple-Choice Test Items of Foreign Language Vocabulary. *Eğitim Fakültesi Dergisi*, 20 (2), 399-426.
- Paramartha, A. A. G. Y. (2017). The Analysis of Multiple Choice Test Quality for Reading III Class in English Education Department, Universitas Pendidikan Ganesha Bali, Indonesia. *Journal of Education Research and Evaluation*. 1(1), pp. 45-56. Que et al, 2005.
- Podolsky, A., Kini, T., & Darling-Hammond, L. (2019). Does teaching experience increase teacher effectiveness? A review of US research. *Journal of Professional Capital and Community*. 10.1108/JPCC-12-2018-0032.
- Puspendik Kemendikbud. (2019). Average score of 2016/2017 school year grade. [online]. Available: <https://hasilun.puspendik.kemendikbud.go.id/>
- Puspendik Kemendikbud. (2019). Average score of 2017/2018 school year grade. [online]. Available: <https://hasilun.puspendik.kemendikbud.go.id/>
- Puspendik Kemendikbud. (2019). Average score of 2018/2019 school year grade. [online]. Available: <https://hasilun.puspendik.kemendikbud.go.id/>
- Puspendik Kemendikbud. (2019). *Guidelines for Implementing Test Construction Training*. Jakarta Indonesia.
- Rao, C., Kishan, P., Sajitha, K., Permi, H., Shetty, J. (2018). Item Analysis of Multiple Choice Questions: Assessing an Assessment Tool in Medical Student. *International Journal of Education and Psychology*. 201-4.
- Rapley, T. (2007). *Doing conversation, discourse and document analysis*. London: Sage.
- Rehman, et al. (2018). Item Analysis of Multiple-Choice Questions. *Pakistan Oral & Dental Journal*, 38(2), 291-93.
- Sahoo & Singh. (2017). Item and Distracter Analysis of Multiple Choice Questions (MCQs) from a Preliminary Examination of Undergraduate Medical Students. *International Journal of Research in Medical Science*, 5(12).
- States, Jack & Detrich, Ronnie & Keyworth, R. (2018). *SummativeAssessment (Wing Institute Original paper)*. 10.13140/RG.2.2.16788.19844.
- Stufflebeam, D. L. (2000). *Guidelines for Developing Evaluation Checklists: The Checklists Development Checklist (CDC)*.
- Sunardi, & Sugini, S. (2014). The Effectiveness of a Two-day Inclusion Workshop on Teachers' Attitudes, Understanding, and Competence in Inclusive Education. *The World Journal of Education*. 4(1). 77-85.
- Toba, R., Noor, W. N., & Sanu, O. (2019). The Current Issues of Indonesian EFL Students' Writing Skills: Ability, Problem, and Reason in Writing Comparison and Contrast Essay. *Dinamika Ilmu*. 19(1). 57-73. doi: <http://doi.org/10.21093/di.v19i1.1506>
- Toksöz, S., & Ertunç, A. (2017). Item Analysis of a Multiple-Choice Exam. *Advances in Language and Literary Studies*. 8. 141.
- Truss, L. (2003). *Eats shoots and leaves: The Zero Tolerance Approach to Punctuation*. New York: Gotham Books.
- Tsagari, C. (1994). *Method effects on testing reading comprehension: How far can we go?* (Unpublished MA thesis). University of Lancaster: UK.
- Wessels, S. (2012). *The Importance of Activating and Building Knowledge*. University of Nebraska: Lincoln.
- Yeh, Y. F. (2012). Teaching Language and Culture: The Importance of Prior Knowledge when Reading Chinese as a Second Language. *Journal of Multilingual Education Research*. 3(6). 91-122.
- Zimmaro, D. M. (2016). *Writing Good Multiple-Choice Exams*. University of Texas: USA.