The Analysis of the Teacher-Made Multiple-Choice Tests Quality for English Subject

M.S.D, Indrayani[1*], A.A.I.N, Marhaeini [2,], A.A.G.Y, Paramartha[3], L.G.E, Wahyuni [4]

[1,2,3,4] English Education Department Ganesha University of Education

ARTICLE INFO

ABSTRACT

This study aimed at investigating and analyze the quality of teacher-made multiple-choice tests used as summative assessment for English subject. The quality of the tests was seen from the norms in constructing a good multiple-choice test. The research design used was descriptive research. Document study and interview were used as methods of collecting the data. The data was analyzed by comparing the 18 norms in constructing a good multiple-choice test and the multiple-choice tests, then, analyzed by using formula suggested by Nurkencana. The result showed the quality of the teacher-made multiple-choice tests a is very good with 79 items (99%) qualified as very good and I item (1%) qualified good. There were still found some problems referring to some norms. Therefore, it is suggested that the teachers have to pay attention to these unfulfilled norms. To minimize the issues, it is further suggested to do peer review, rechecking, and editing process.

1. Introduction

Objectives of Curriculum 2013 (K-13) in Indonesia, regulated in Indonesian Ministry of Education and Culture (Ministerial of Education and Culture of Indonesia, 2016), includes four competencies which should be required by students, namely, spiritual attitude, social attitude, knowledge, and skills. Those competencies should be implemented in every subject including English subject that is learnt as a foreign language. (Estilden, 2017) emphasizes that English is the greatest common language which takes a big role in education field and it becomes one of the subjects taught in school. English has been introduced earlier at Elementary school as a development of the world globalization which demands the students to have English skills in order to compete with people around the world. To be able to accomplish those competencies in English teaching and learning process, one of the essential elements that should be applied by a teacher is assessment.

The existence of assessment and a teacher in English teaching and learning activities is inseparable since conducting assessment is a part of teacher's roles and responsibilities. According to (Jabbarifar, 2009), assessment is a process that includes four basic functions including measuring improvement over time, motivating students to study, evaluating teaching methods, and ranking students' capability. In line with the statement, (Dejong et al., 2002) state assessment provides teacher a report about student's improvement and gives guidance for future lesson. It is not only beneficial for the students, teachers also can improve their performance since they can figure out their strengths and weaknesses of their teaching method (Tosuncuoglu, 2018). Since assessment is important in English teaching and learning process, it is regulated in Indonesian Ministry of Education and Culture (Ministerial of Education and Culture of Indonesia, 2016).

Ministerial Regulation No. 23/3016 is used as a reference of Educational Assessment Standard in Curriculum 2013 (K-13). According to Indonesian Ministry of Education and Culture, regulated in article 3 paragraph 1 of Assessment's Scope, there are three elements in Curriculum 2013 (K-13) that should be assessed by a teacher, namely, attitude, knowledge, and skill. In terms of mechanism for evaluating learning outcomes, the student's knowledge can be assessed through written tests, oral tests, and

Corresponding author.
E-mail addresses: indrayani897@gmail.com (M.S.D, Indrayani)

assignments based on competencies that want to be achieved. One of the written tests, the most popular type of instruments that many English teachers apply in their classrooms is multiple-choice test (MCT).

(Zimmaro, 2016) states that multiple-choice test is useful for measuring knowledge outcomes and other types of learning outcomes. The statement is strengthened by (Kolte, 2015) as multiple-choice test is used as an objective and reliable instrument to evaluate knowledge of the students. Multiple-choice test is commonly utilized by classroom teacher as it contributes several advantages. First, it is categorized as fast, relatively easy, and economical to be scored (Bailey, 1998). Second, it provides practical format for assessing student's knowledge at various level of learning (Zimmaro, 2016). Third, it can be conducted in an easy way even in large number of students simultaneously as the fact that National examination (NE) in Indonesia, TOEFL, and IELTS test have been using multiple-choice for years. Fourth, there is only one correct answer so it avoids a subjectivity in scoring (Kolte, 2015). Since multiple choice test is an essential part of assessing students' knowledge, its quality needs to be tested by following certain standard.

(Burton et al., 1990) states that the quality of multiple-choice test can be determined in terms of norms which are used in the process of constructing the instrument. Norms are the starting point to develop an instrument as (Haladyna, 2004) argues that a set of guidelines or norms should be adopted in constructing items of multiple-choice test. In line with the statements, (Hall & Marshall, 2014) suggest that guidelines in making a good multiple-choice test need to be highly considered. These assertions are justified by the theories of the quality of multiple-choice test that can be seen from (Haladyna, 2004; Hall & Marshall, 2014; Puspendik Kemdikbud, 2019). (Haladyna, 2004) provides 31 norms with 4 dimensions including content, style and format, writing the stem and writing the options. (Hall & Marshall, 2014) specify 12 guidelines. (Puspendik Kemdikbud, 2019), the authority of educational assessment in Indonesia, even states that there are 16 norms with 3 dimensions including material, construction and language in which will be used as a foundation to make a good multiple-choice test in national examination (NE). Thus, the theories acknowledge that the norms in constructing a good multiple-choice test are very fundamental element to be concerned.

Within the context of Indonesia, multiple-choice test has been used for years by classroom teachers which is directly constructed and scored by the teacher. Multiple-choice test is then called as a teacher-made test. Teacher-made multiple-choice test can be applied in form of formative and summative assessment. Based on the function, formative assessment is used to monitor student's progress while summative assessment is used to evaluate student's learning. Teacher-made test is considered more applicable because the content of the test is straightforwardly related to the material taught in the class (Lebagi et al., 2017).

The congruity between the teacher-made multiple-choice test items and the content taught in the class indicates that the teachers have successfully achieved indicators that want to be achieved through national examination (NE). A good assessment implementation in terms of assessment instrument, further, will directly give an impact to the student's achievement in national examination (NE). It is suggested by (Black & Wiliam, 1998) that a good assessment practice arises good understanding of contents that have been taught in the classroom in which impacts to the student's achievement.

One of the schools that has been implementing a good assessment practice especially in the quality of the instrument is SMP Negeri 1 Singaraja. SMP Negeri 1 Singaraja, a public school located in Buleleng regency, uses teacher-made multiple-choice test for English subject as an instrument to assess the student's knowledge for middle test. A good assessment practice in terms of the quality of the instrument that is implemented at SMP Negeri 1 Singaraja expects the teachers to be able to perform a good multiple-choice test in which the items should reflect the basic competencies that want to be achieved through the national examination (NE).

In the pre-observation, the good assessment practice at SMP Negeri 1 Singaraja has been identified as according to (Puspendik Kemdikbud, 2019), SMP Negeri 1 Singaraja occupied the highest score for English subject of the average of national examination (NE) for the past five years in a row. In 2018/2019 academic year, the average of the students' achievement of national examination (NE) for English subject reached 80.74. The high score indicates that the English teachers had constructed a good multiple-choice test as there was a congruence between the items and the basic competencies in which automatically assisted the students to be able to answer the items correctly in the national examination (NE). Furthermore, the English teachers have performed a blueprint to guide them to make a good multiple-choice test as well as conducted item analysis of the multiple-choice test to estimate the reliability and validity of the instrument.

The high average of national examination (NE) 2019 for English subject, the good performance of the teachers who have accomplished blueprints as well as item analysis can be considered as a reflection of a good assessment practice at SMP Negeri 1 Singaraja. Unfortunately, the norms, one of the fundamental requirements to ensure the quality of the instrument which eventually influences the assessment practice

(Puspendik Kemdikbud, 2019) are yet to be identified. There is no further report regarding the construction of the teacher-made multiple-choice test whether or not the items have followed certain norms. These statements are emphasized by   Crockett and Churches (2016) as teachers have to concern in constructing an instrument by following certain norms. In consideration of the substantial roles of teacher-made multiple-choice test as the instrument for summative assessment, the norms in constructing the test need to be crucially investigated.

Hence, this study is arisen to investigate the quality of multiple-choice tests as summative assessment for English subject used by English teachers at SMP Negeri 1 Singaraja. To ensure the quality of the instrument, this study concerns on the analyzing the congruity and discrepancy between the teacher-made multiple-choice test items taken by the seventh, eighth, and ninth grades at SMP Negeri 1 Singaraja and the norms in constructing a good multiple-choice test, one of the fundamental requirements in implementing a good assessment practice, suggested by (Haladyna, 2004; Hall & Marshall, 2014; Puspendik Kemdikbud, 2019; Zimmaro, 2016).

2. Methods

The research design used in this study was descriptive research. The purpose of a descriptive research is to describe a picture of a phenomena or demonstrate how things are related to each other as accurate as possible (Blumberg et al., 2005). In this study, the descriptive approach was intended to investigate the quality of multiple-choice test as summative assessment for English subject taken by the students at SMP Negeri 1 Singaraja in terms of the norms in making a good multiple-choice test.

This study was conducted at SMP Negeri 1 Singaraja as it had high average score of national examination (NE) in 2018/2019 academic year for English subject among junior high schools in Buleleng regency which the score reached 80.74 (Puspendik Kemdikbud, 2019). It can be considered as a reflection of one of good assessment practices

The quality of the teacher-made MCTs was analyzed by comparing the 18 norms in constructing a good multiple-choice test suggested by (Haladyna, 2004; Hall & Marshall, 2014; Puspendik Kemdikbud, 2019; Zimmaro, 2016) and the MCTs to see the congruity. After the total of the 80 items were compared, then, the data was analyzed by using (Nurkancana dan Sunarta, 1992) formula. After that, the results of data tabulation were calculated and classified to some classifications.

There are five level qualifications of the quality of the teacher-made multiple-choice test which are very good, good, sufficient, poor, and very poor. The teacher-made multiple-choice test is considered very good if the percentage is more than or equal to 75%, good if the percentage is less than 75% and more than or equal to 58%, sufficient if the percentage is less than 58% and more than or equal to 42%, poor if the percentage is less than 42% and more than or equal to 25%, and very poor if the percentage is less than 25%.

3. Result and Discussion

Findings

Teacher-made multiple-choice test as summative assessment for English subject grade VII consists of 20 items, grade VIII consists of 30 items, and grade IX consists of 30 items with four options of each item which are a, b, c, and d. It was analyzed by identifying the congruity between the 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 result of the analysis demonstrates the percentages of the items are ranged. Out of 80 items, 33 items (41%) perfectly fulfill all the norms in making a good multiple-choice test. There are 19 items (24%) with 1 norm unfulfilled that are considered to have above 90%. The items that have above 80% norms fulfilled indicating 2-3 norms unfulfilled consist of 23 items (29%). 5 items (6%) with 4 norms unfulfilled are considered to have above 70% norms fulfilled.

The quality of teacher-made multiple-choice tests as summative assessment for English subject at SMP Negeri 1 Singaraja is very good. The result shows that 79 items (99%) that are very good with the percentage above 75% and 1 item (1%) is good with the percentage 72%. The results are in line with the teachers' expectation.

Discussion

The 33 items that have 100% norms fulfilled completely follow all the 18 norms. In the content guideline, the items already reflect to the basic competency, there is no dependency on the answer of the previous items, the items give clear focus of what is being asked, there is no subjectivity that makes the

student give their opinion, and the spelling and grammar are correct. In the style and format, the options are already formatted vertically, and the punctuation and capitalization are correct. In the writing stem, there is no stem that contains double negatives. In the writing opinion, the options have the correct answer, do not contain clues, do not repeat words, and do not use "all of the above" or "none of the above". Moreover, the options are homogenous, about the same length placing orderly, not overlapping, and plausible.

The 19 items that have above 90% norms fulfilled have five different issues which are the use the punctuation and capitalization, the reflection of the basic competency, the homogeneity of the options, the correct answer, and the placing order of the options. The use of punctuation and capitalization are frequently caused by the errors in the blank space at the end of the stem and the writing clause on the options. The basic competency is not reflected because the item does not provide topic that is being asked resulting no correct answer. For the options, there is inconsistency in terms of the part of speech that causes the options are heterogeneous in grammatical structure. The placement of the options is not in logical order since there is an option that is much longer or much shorter than the others. The problems that have above 90% fulfilled norms are mostly about the use of the punctuation and capitalization. The problem can be seen in the ellipsis. The options are in form of the clause ending up with full stop. The stem of the items is in form of uncompleted sentence that is finished by three full stops. Since the stem is in form of a sentence, there must be ended by four full stops, and for the options, they must not be punctuated with period.

The 23 items that have above 80% norms fulfilled have variety of problems. In this percentage, the unfulfilled norms are the same with the previous percentage; however, there is further information of norms that are being unfulfilled by the items. They are the use of grammar on the items, the options plausibility, the item clue, the options length, and the item clear focus. The problems of the use of grammar on the item are caused by the inaccurate structure, parallelism, mistake of the pronoun, and verb. It indirectly leads to the implausibility of the options as the distractors are contextually deviated in which they are not relevant with topic that is being asked. There are some items that one option is much longer than the others indicating a clue or information directing the test-taker to find the correct answer easily. Since the one option is contrast in terms of the length, the placement of the options is automatically not in logical order. The unclear focus on item is resulted by the inconsistency in using the pronoun and the incorrect use of the grammar. The implausibility of the distractors as the consequence of the inaccurate use of the grammar and the homogeneity of the option is the most common mistake of the items that have above 80% fulfilled norms. The stem is grammatically correct, but the options are not. Besides, the options do not indicate a homogeneity in grammatical structure because three options are in form of verb + object + adverb of time and one stem only states the verb. The redundancy on the item makes the distractors implausible.

The 5 items, considered to have above 70% norms fulfilled, have the same mistake with the previous percentage. However, the major problems are about the length and the placing order of the options. The options show different length. If the options are not about the same length, they will automatically result an illogical order placement. One of the options is much longer than the others making the options contrast in terms of the length. Thus, the options are not placed orderly.

There are found two minor problems in the MCTs (98%) which are about the item clear focus and the correct answer. The number of items fulfilling the norm is only 78 out of 80 items. The ambiguous focus on the item is caused by the problem of the use of the pronoun. On the stem, it is stated that Rafasya is a boy that is marked by the term his (possessive adjective). Based on the interview, the teacher stated the same. However, on the options, the subject pronoun is changed becoming She. There is no clear focus that is being asked that might lead confusion. Based on the interview, it is caused by a typographical error. The teacher realizes the problem when the test is already printed out. Limited time and expense make the teacher not to revise it, but the instruction is directly given in the classroom.

Another minor problem is about the correct answer (98%). On the item, there must be one correct answer, however, item 16 provides two correct answers. Option C and D are contextually the same. Based on the interview, the teacher has realized it as there must be another expression used to avoid students' confusion.

Based on the result of the analysis above, it demonstrates that the most unfulfilled norm is about the use of punctuation and capitalization causing this norm covers only 71% items following it. The number of items fulfilling the norm is only 57 out of 80 items. The rest of the norms have varieties of percentage ranging from 84-100%.

Table 1. Percentage of Norms Fulfilled by the Teacher-Made MCTs Grade VII, VIII, and IX

Norms Number	Norms' Description	Number of Items Fulfilling the Norms	Norms Fulfilled (%)
1	Reflecting basic competencies	74	93%
2	Not depending on the previous options	80	100%
3	Giving content clear focus	78	98%
4	Avoiding opinion-based items	80	100%
5	Being grammatically correct	68	85%
6	Having correct spelling	80	100%
7	Not containing clues	75	94%
8	Formatting options vertically	80	100%
9	Having correct use of punctuation and capitalization	57	71%
10	Not containing double negatives	80	100%
11	Having homogeneous options in content and grammatical structure	68	85%
12	Having one correct answer	78	98%
13	Having options with relatively the same length	69	86%
14	Placing options in logical and numerical order	68	85%
15	Not repeating options with the same words or phrase	80	100%
16	Not having overlapping options	80	100%
17	Having plausible distracters	67	84%
18	Not using *"none of the above"* or *"all of the above"*	80	100%

The most unfulfilled norm, 23 unfulfilled norms (71%) is about the use of the punctuation and capitalization. Punctuation and capitalization are essential to convey meaning of items in making a good-multiple choice test. According to (Connelly, 2009), incoherent use of punctuation can cause ambiguity. In line with the statement, proper punctuation leads appropriate meaning and understanding (Truss, 2003). Another element which affects readability and intelligibility is capitalization (Salamin et al., 2016). In addition, (Solomon, 1990; Zolfaghari, 1999) believe that the function of punctuation is to split sentences into clauses and make clear connection between the parts of a sentence.

(Mann, 2003) claims that applying punctuation marks serves a difficulty. The statement is strengthened by (Kurniawan et al., 2014) which demonstrates that the ability of teachers in Indonesia in comprehending the use of punctuation is considered low. They claim that out of 14 types of punctuation, ellipsis that is a part of a sentence that has been omitted, got the lowest number of teachers understood it. It is related with the result of this study in which the problem of the use of punctuation is mostly in form of ellipsis whether the items use three or four full stops. Based on the interview with the test makers, it is caused by an inadequate time and misconception in applying the ellipsis. Two of them claimed that the cramped schedule, for instance doing administration business and teaching, lead them to not recheck the MCTs especially for the punctuation and capitalization. Another test maker has misconception towards the use of ellipsis which results an error.

Besides, it also happens due to the decrease of MGMP (Teachers Professional Development Forum) program that is usually conducted every Saturday. MGMP program evaluates the teachers' performance or discusses any problems that ruin the essence of teaching and learning in the classroom. Based on the interview, the teachers give an illustration about the existence of MGMP saying that if there is problem in terms of the quality of test, for instance, there will be a discussion so that it can be resolved immediately. Unfortunately, this program is not currently running well. However, the teachers claim that they will put some effort to revive the MGMP program so that all idea will be constantly integrated.

Even though eight norms are perfectly followed by all the items, the other norms also contribute high number of items following it ranging 84-98%. Except the norm about the use of punctuation and capitalization that has 57 out of 80 items following it.

4. Conclusion

The quality of teacher-made multiple-choice tests used as summative assessment for English subject at SMP Negeri 1 Singaraja is very good with 79 items (99%) qualified very good and 1 item (1%)

qualified good. However, there is a norm that is highly unfulfilled by the items that need more attention which is the use punctuation and capitalization.

References

Bailey, K. M. (1998). *Learning about language assessment: dilemmas, decisionjs, and directions*. Heinle & Heinle.

Black, P., & Wiliam, D. (1998). "Assessment and Classroom Learning,. *Assessment in Education: Principles, Policy and Practice*, *5*(1), 7–73. https://doi.org/https://www.gla.ac.uk/t4/learningandteaching/files/PGCTHE/BlackandWiliam1998.pdf

Burton, S. J., Sudweeks, R. R., Marrill, P. F., & Wood, B. (1990). How to prepare better multiple-choice test items: Guidelines for university faculty. *Doctoral Dissertation, Brigham Young University. Department of Instructional Science*. https://doi.org/http://sites.educ.ualberta.ca/staff/ldelia/How-to-prp-MC-Tests=UK'University.pdf.

Connelly, M. (2009). *Get writing: Sentences and paragraphs*. Cengage Learning.

Dejong, G., Kokinakis, C. L., & Kuntzleman, C. (2002). The role of assessment in meeting the NASPE physical education content standards. *Journal of Physical Education, Recreation & Dance*, *73*(7), 22–25. https://doi.org/https://doi.org/10.1080/07303084.2002.10607842

Estilden, K. P. (2017). Why is it important to learn English: A study of attitudes and motivation towards English and English language learning in Swedish upper secondary school. *Retrieved September 2nd, 2019*. https://doi.org/https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1078032&dswid=-1149

Haladyna, T. M. (2004). *Developing and Validating Multiple-Choice Test Items*. Lawrence Erlbaum Associates.

Hall, & Marshall. (2014). *A Guide for Developing Multiple-Choice and Other Objective Style Questions. Centre for Academic Development*. Victoria University of Wellington, New Zealand.

Jabbarifar, T. (2009). The Importance of Classroom Assessment and Evaluation in Educational System. *Proceedings of the 2nd International Conference of Teaching and Learning*. https://doi.org/https://www.fche.cl/wp-content/uploads/2019/02/Jabbarifar-Taghi_2009_The-Importance-of-Classroom-Assessment-and-Evaluation-in-Educational-System.pdf.

Kolte, V. (2015). Item analysis of Multiple-Choice Questions in Physiology Examination. *Indian Journal of Basic and Applied Medical Research*, *4*(4), 320–326. https://doi.org/https://ijbamr.com/pdf/september%202015%20320-326.pdf.pdf

Kurniawan, O., Noviana, E., & Muhammad, N. (2014). Analisis Kemampuan Guru Sekolahdasar Dalam Memahami Konsep Penggunaan Tanda Baca Se-Kecamatantampan Pekanbaru. *Jurnal Primary Program Studi Pendidikan Guru Sekolah Dasar Fakultas Keguruan Dan Ilmu Pendidikan Universitas Riau*, *3*(1). https://doi.org/http://dx.doi.org/10.33578/jpfkip.v3i1.2107

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). https://doi.org/https://doi.org/10.31327/jee.v2i2.289

Mann, C. (2003). Point counterpoint: Teaching punctuation as information management. *College Composition and Communication*, *54*(3), 359–393. https://doi.org/DOI: 10.2307/3594170

Ministerial of Education and Culture of Indonesia. (2016). Regulation of the Ministry of Education and Culture of the Republic of Indonesia Number 4 of 2018 Regarding Assessment of Learning Outcomes. *Jakarta, Indonesia*.

Nurkancana dan Sunarta. (1992). Evaluasi Hasil Belajar. *Surabaya: Usaha Nasional*.

Puspendik Kemdikbud. (2019). Guidelines for Implementing Test Construction Training. *Jakarta Indonesia*.

Salamin, A., Farrah, M., Zahida, R., & Zaru, N. (2016). An Investigation into Punctuation and Capitalization Errors Made by Hebron University EFL Students. *Al-Azhar University Research Journal*.

https://doi.org/http://dspace.hebron.edu:80/xmlui/handle/123456789/166

Solomon, M. (1990). The power of punctuation. *Design Issues*, *6*(2), 28–32. https://doi.org/DOI: 10.2307/1511434

Tosuncuoglu, I. (2018). Importance of Assessment in ELT. *Journal of Education and Training Studies*, *6*(9), 163–167. https://doi.org/https://doi.org/10.11114/jets.v6i9.3443

Truss, L. (2003). Eats shoots and leaves: The Zero Tolerance Approach to Punctuation. *New York: Gotham Books.*

Zimmaro, D. M. (2016). Writing Good Multiple-Choice Exams. *Texas: University of Texas – Austin.*

Zolfaghari, H. (1999). How to write a composition. *Tehran: Asatir Press.*