THE OPTIMAL NUMBER OF OPTIONS USED IN MULTIPLE-CHOICE TEST FORMAT FOR NATIONAL EXAMINATIONS IN INDONESIA

Herland Franley Manalu[1*], Diana Anggraeni[2]

[1]Lecturer, Department of English Literature, Faculty of Social Science and Political Science, Universitas Bangka Belitung, Bangka Belitung, Indonesia; [2]Lecturer, Department of English Literature, Faculty of Social Science and Political Science, Universitas Bangka Belitung, Bangka Belitung, Indonesia.
Emails:[1*]herland-fm@ubb.ac.id, [2]diana-anggraeni@ubb.ac.id

Abstract

Purpose of the study: The present research purports to find out the most favourable number of options used in the Multiple Choice (MC) format for Senior High Schools in Indonesian National Examinations (UN) and suggest the testing division within the Indonesia ministry of education to consider the result of the study

Methodology: Two English tests using MC questions with five options, consisting of 50 questions per test format, were used as the elicitation devices for this research. One of these English tests was rewritten to create four options by deleting the non-functioning distractors in each question. Both tests with different MC test formats were administered to 2 groups within two state Senior High Schools (SMA). The Classical Method and Rasch Analysis were utilized to compare item facility, item discrimination, distractor measure correlation, and reliabilities across the two MC test formats. To corroborate the findings of the study, questionnaires were randomly distributed to 120 SMA students and 15 SMA English teachers.

Main Findings: The findings suggested that four options were more difficult than five options. There was significant change observed in Item Facility (p<0.05) and Item Discrimination across the two MC test formats. Based on the questionnaire data analysis, the four-option is the more optimal and preferable format to be used in the National Examinations.

Applications of this study: Regarding the practicality issues like saving time and money for implementing the tests, minimizing the amount of time and effort needed for test-makers to create the tests, and also reducing the risk of providing implausible distractors for developing MC tests, this study concluded that the four-option MC format is more optimal to be used for Indonesian National Examinations in Senior High School.

Novelty/Originality of this study: The testing division within the Indonesia Ministry of Education uses different numbers of MC options in Elementary, Junior High, and Senior High Schools examinations. The Ministry has predetermined three options for elementary schools, four options for junior high schools, and five options for senior high school students. The decision made the researchers of the present study eager to find out whether the use of MC format with reduced options in the UN SMA will produce the same results or maybe increase or reduce the efficiency, effectiveness, reliability, and practicality of the test administration. This is the first study to compare the number of choices in the MC test format used in Indonesian National Examinations.

Keywords: MC Test Formats, Distractors, National Examinations, English Tests, Item Facility, Practicality.

INTRODUCTION

Language testers all around the world have frequently used Multiple Choice (MC) in any English tests and this test format has also been extensively applied in all subjects to measure test-takers' language ability based on what students have learned, what teachers have taught or what criteria have been described in the syllabus, textbook, materials, etc. MC is used for National Examinations in Indonesia because test-makers believe that this format contributes to the effective and efficient scoring system in administering the National Examinations (UN) for senior high school level. The testing division within the Indonesian Ministry of Education and Culture (or *Kementerian Pendidikan dan Kebudayaan Republik Indonesia,* abbreviated as *Kemdikbud*) creates MC items which consist of five options per item. However, it raises questions whether the five-option MC format is the most optimal format to be used in the UN for SMA, while the reduced-option MC format is still able to distinguish strong students from weak students in Senior High School. These questions have led scholars who are interested in language-testing to the investigation of this area (Redmond et al., 2012).

Haladyna, et al. (2002) and Hughes (2003) stated that MC questions occur in two, three, four, five, and infrequently with more options. However, it raises the question of whether there is the most optimal number of options in MC items to be used at a higher level of education, such as senior high school and university. From a purely psychometric perspective, more options in MC items are preferable because with more options, the effect of guessing the answer can be reduced, and the reliability will be higher. By using two options (A, B), the guessing chance is 50% (Dirgeyasa, 2017). However, the five-option format reduces the student's chance of getting the item correct to 20%.

There are other matters to regard as the number of options rises. The test takers must do more reading and consider more possible answers carefully to a question simultaneously (Sujinah et al., 2019; Susanto, 2017; Vyas & Supe, 2008). So, it

creates questions about whether many options used in the MC format will challenge the students' cognition. In the very early level of education, such as kindergartens and primary schools, two to three options are more appropriate. Meanwhile, in the higher levels, students are ordinarily able to manage four or more options. It echoes Garners' argument about human thinking capacity. He concluded that human beings of different ages have different capacities to process multi-dimensional information.

Due to the cognitive challenge, the testing division in Kemdikbud uses different numbers of MC options in Elementary, Junior High, and Senior High School examinations. The Ministry has predetermined the three-option format for primary schools, the four-option format for junior high schools, and the five-option format for senior high school students. This decision made the researchers of the present study eager to find out whether the use of MC format with reduced options in the UN SMA will produce the same results or maybe increase or reduce the efficiency, effectiveness, reliability, and practicality of the test administration. This question is the focus of this study.

LITERATURE REVIEW

Since the last decade, many researchers from various fields of study (Abad, et al., 2001; Baghaei & Amrahi, 2011; Dehnad et al., 2014; Epstein, 2007; Lee & Winke, 2013; Shizuka, et al., 2006; Tarrant &Ware, 2010) have focused their research on investigating the most optimal number of MC options. Even though most research in this area has shown that the three-option MC test format seems to be favoured in many fields, we see that many test-makers still rely on the five- or the four-option MC formats because they consider that the three-option MC format which has 33.3% guessing chance is not appropriate to test students with a higher level of understanding in senior high school or university. Based on our survey, mostly Senior High School teachers and students in Indonesia consider that the three-option format does not really cognitively challenge the students' understanding and does not clearly differentiate the level of education among TK, SD, SMP, and SMA. It echoes Allison and Palmer's (2010) arguments over the best number of options for MC test items that remain happening between the five-option and the four-option MC formats for a higher level of education.

Another reason is that some researchers in this area have different results for the best number of MC options. Baghaei and Amrahi(2011)divided the researchers' results of MC studies into three groups: researchers who found that the 5-option MC format is the optimal number of options if the distractors are plausible constructed (Rodriguez, 2002; Sidick et al.,1994), researchers who found out that the 4-option format is the best number of options (Owen &Froman, 1987; Costin, 1972), and researchers who found out that the 3-option format is the best number of options (Landrum, et al., 1993; Rogers & Harley, 1999).

The subsequent reason why the debate over the optimal number of options for MC items is still going on is that the test-makers' belief concerning human capacity to process multi-dimensional information, as I have mentioned above. Test-makers believe that students from a different level of education such as kindergarten (TK), elementary (SD), junior high (SMP) and senior high schools (SMA) have different limits on their capacity or mental capability to process the information (Kubiszyn & Borich, 2004; Rodriguez, 2002). The belief echoes with the government's decision to differentiate the number of options used in TK, SD, SMP, and SMA. It makes teachers and test writers neglect the result of MC studies which mostly favoured of the three-option format even though test-makers still favour the use of the two-option MC format for TK, the three-option MC format for SD, the four-option MC format for SMP and the five-option MC test format for SMA (Tarrant, et al., 2009).

From the explanations of the strengths and the weaknesses of the different number of options in MC test formats stated above, this study seeks to find out which format gives the optimal result to the teachers, as the examiners, and to the students, as the examinees. This study compared the five-option and the four-option MC test formats. It is hypothesized that reducing one option from the five-option MC format to create the four-option format in the English test for UN SMA is a good way to find out the most optimal number of options in MC questions. This study directed the achievability of that goal by deleting the non-functioning distractor in each question (Schneid, et al., 2014). Non-functioning distractors are options that are selected infrequently (<5%) by examinees or otherwise do not perform as expected (Tarrant, et al., 2009). The researchers of this study used two sets of English UN 2017 test books. Therefore, in this study, the researchers decided to compare the five-option and the four-option MC formats to find the most optimal number of MC options to be used in the UN SMA by measuring the results of the UN-like tests and corroborating the scores with the questionnaire analysis.

Each type of MC format has its strengths and weaknesses. The primary strength of the five-option format is a relatively low guessing factor. It has a 20% guessing factor. The five-option format also has the advantage of testing students' comprehension more deeply than the four- and the three-option formats, if the distractors are plausibly constructed. However, this format gives disadvantages to teachers and students. The teachers have to spend a lot of time writing the distractors because it is not easy to create a plausible distractor to differentiate strong students from weak students. A distractor is easily eliminated by both strong and weak students if it is not plausibly constructed. From the students' point of view, the five-option format requires them to spend a lot of time reading the options and to think out of the correct option.

The five-option MC format remains as the standard format to be used in the Indonesian senior high school level. This format is used by the teachers and test-makers to create the monthly test, to administer the UN SMA, and to develop test banks (Masters, et al., 2001). Test-makers who are in favour of writing the MC questions with five options consider that the five-option format is the best because it is difficult for students to guess the correct option from the five available options simply by guessing. Only the students who understand the material can get the correct option. That is why many test-makers think that the five-option MC format is the best format to test students from a higher level of education if the distractors are plausibly constructed (Rodriguez, 2002; Sidicket al.,1994).

The MC test with four options and five options have the advantage of a relatively low guessing chance. It has a 25% guessing chance (Kubinger et.al, 2010). The four-option format also has the advantage of being less strenuous for the students to find the correct option because the number of options in this format is fewer than the five-option format. Thus, the four-option format takes less time to be conducted than the five-option format. However, the four-option format also has a problem that requires the students to spend a lot of time reading the options and think out of the correct option (Owen & Froman, 1987; Costin, 1972).

Test-makers who are in favour writing the MC questions with four options think that this is the best format because the two-option and the three-option formats are indicated to be easier to do than the four-option format and these formats are not appropriate to be used in Indonesian senior high schools and universities (McCoubrie, 2004). The five-option format is considered to be slightly more difficult than the four-option format. The test makers are encouraged to use the four-option to minimize the guessing effect. As a result, the majority of the international English proficiency tests (e.g., TOEFL, TOEIC, SAT, and IELTS) usually use the MC test format with four options.

METHODOLOGY

This research is qualitative and quantitative by nature because to do a thorough job of test analysis, a mixed-method on item analysis was employed (Bachman, 2004; Frankfort-Nachmias & Nachmias, 2007). Moreover, the samples were chosen through a random sampling technique. Four classes from two different schools were available to be involved in the study. On the whole, this study initially involved 160 students from two State Senior High Schools (SMAN) in Jakarta in doing the UN-like tests. However, only 120 students were able to take the two different MC test formats of the UN-like tests. These students were asked to do the English UN-like tests and fill out the questionnaires.

The participants were from SMAN 8 and SMAN 105 Jakarta. We chose these schools due to their ranks in the result of the national examinations every year. SMAN 8 is reputed to get the best scores in the English exams every year and hence their students are academically assumed to be the strong students in this study. SMAN 105 annually reaches the lowest rank, and hence the students are considered to be the weak students. This specific choice of schools should add to the reliability of this study. The UN-like tests were administered during their English classes. The English teachers told the students that they were all going to have the UN-like tests two times and fill out questionnaires after taking the tests.

The SMAN students took the two different types of English UN-like tests which were divided into two groups. The first group was the English UN-like test with the five-option format and the second group was the English UN-like test with the four-option format. Only the participants, who took the two types of English UN-like tests, were taken as the samples in this study. This study also involved 15 SMA English teachers in Jakarta to gain their opinion about the best number of MC options in the UN SMA. All of them were asked to fill out the questionnaires.

To be able to find out the best number of options in the MC test format, the data was collected using various instruments. Two UN Bahasa Inggris SMA 2017 test books were used as the elicitation devices to gather students' scores by administering the tests in their class. Questionnaires were used to garner SMA students' and English teachers' views about the simplification of options (Bailey &Nunan, 2009). Statistical software programs were used to analyse the data.

The 50 items of the five-option MC test format from the past UN test book, which was coded as B-12, were reconstructed to create a reduced-options format comprising 50 four-option MC questions. The undesired option was deleted. The deletion of the undesired option was conducted by finding the least frequently chosen distractor or the non-functioning ($< 5\%$) distractor from the statistical data that we obtained from Litbang Kemdikbud.

To confirm the result of the study, the questionnaires as a useful elicitation tool to get more information about the best number of MC options were developed. The respondents were students from two state senior high schools (SMAN) and English teachers from five senior high schools in Jakarta. The questionnaire for English teachers was written in English and was disseminated to fifteen English teachers from five senior high schools in Jakarta, but the questionnaire for SMAN students was written in English with its translation in Bahasa Indonesia because there were many weak students taken as respondents in this study. The type of questions used in the questionnaires was closed-ended items. Respondents answered by putting a thick on the available columns and filling out some blank spaces provided in the questionnaires. In the questionnaires, the teachers and students were asked to choose the best number of options in MC format and write their opinion about the reduced options. Questionnaire dissemination was conducted as soon as the students finished doing the second UN-like test and when the English teachers had no class to teach so that they would be able to give reliable answers.

To data analysis, three statistical software programs were used to assist the process of data analysis in this study. This study used *Lertap 5.8 for MS Excel 2007* to measure Descriptive Statistics, including Median, Mean, Standard Deviation, and Variance. The software was also used to measure item analysis, such as Distractors Performance, Item Facility, and Item Discrimination. To analyze the questionnaire data and count the *dependent t-test*, then *IBM SPSS v.20* software was used. *Jmetrik v.03* software was used to measure the Rasch model to obtain Person Separation Indices and Cronbach's Alpha Reliability. This study used different statistical software programs because each statistical software program has different statistical features.

RESULTS

Qualitative Descriptive Analysis

Qualitative descriptive analysis was conducted to check the content validity of the two English UN-like tests used in this study by examining all the items according to the aspects of materials, construction, and language in the MC item analysis card. Based on the result of the qualitative descriptive analysis, the English UN 2017 tests which were coded as B-12 and C-24 had very high content validity 95% and 96% respectively. The result showed that the content validities of the two English UN tests were good because each item reached more than 95% aspects of good content validity. Therefore, we can conclude that the two tests are valid to measure what they were intended to measure and can give the accurate result of what was meant to be measured.

42 out of 50 items in the English UN-like test coded as B-12 qualified 95% ≥ the aspects of content validity which were written in the MC items analysis card. It means that these items met the criteria of good MC items. The other 8 items did not fully qualify all criteria. Whereas 44 out of 50 items in the English UN-like test coded as C-24 qualified 95% ≥ the aspects written in the MC items analysis card. It means that these items also met all the criteria for writing good MC items (Boland, et al., 2010).

Quantitative Descriptive Analysis

Descriptive Statistics for the two UN-like tests

Firstly, to check the comparison between the two groups, this study examined the Descriptive Statistics of the two different English UN-like tests which consist of 50 MC questions. The mean for the five-option group (n=120) was 34.93 (SD=9.01), the mean for the four-option group (n=120) was 32.86 (SD=8.54). The variances for the five-option format (F= 81.15) and the four-option format (F= 72.94) were significantly different from each other.

Table 1: Descriptive statistics for the two UN-like tests

	5-option format	**4-option format**	**TOTAL**
N	120	120	120
Min	16.00	14.00	30.00
Median	36.50	34.50	70.50
Mean	34.93	32.86	67.79
Max	49.00	46.00	95.00
S.D.	9.01	8.54	17.48
var.	81.15	72.94	305.48

Based on the line-chart of students' scores shown in figure 1 below, we can see that the line of the four-option format has the same pattern as the five-option format. We can conclude that strong students and weak students tend to get the same result after doing the two different tests. It means that the reduced-options MC test is still able to differentiate the strong students from the weak ones.
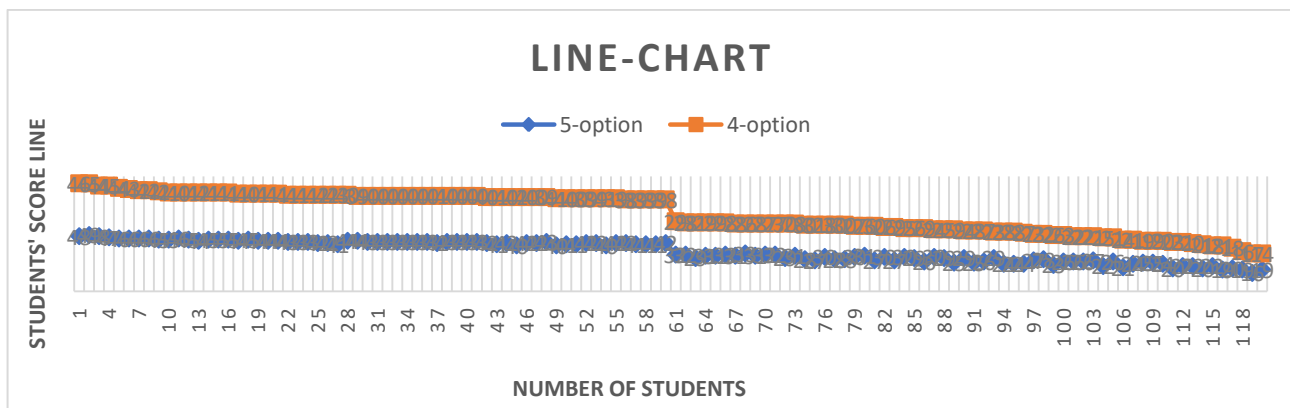
Figure 1: Line-chart of the students' score

Item Facility and Item Discrimination of the two UN-like tests

Difficulty parameters are summarized in table 2. The average value of IF for the five-option format is .70 and the average value of IF for the four-option format is .66. Hence, the reduction of MC options did not make the questions easier; as a matter of fact, the mean difficulty parameter remains the same, and even the four-option IF value is 0.04 bigger than the five-option IF value.

The Rasch analysis set discrimination to be consistent across all the dichotomous items in a given dataset, as a measurement ideal (Rasch, 1980; Wolfe & Smith, 2007). Person separation indices – the Rasch equivalents of Cronbach's alpha coefficients – of the two different English UN-like tests were 0.39 for the five-option set and 0.36 for the four-option set.

The Correlation between Item Facility (IF) and Item Discrimination

The results of the two different MC format UN-like tests showed that there was no significant difference between Item Facility and Item Discrimination in the result of the five-option and the four-option UN-like tests. The spread of items below shows the correlation between Item Facility and Item Discrimination in the result of the two different format UN-like tests:

Table 2: The correlation between Item Facility (IF) and Item Discrimination (ID)

IF	5-option MC Format	4-option MC Format
.00:		
.10:		
.20:	Q28 Q45	Q22 Q30 Q35 Q45
.30:	Q22 Q41 Q43	Q3 Q16 Q17 Q20 Q27 Q43
.40:	Q38	Q5 Q26 Q38 Q39 Q40 Q41 Q44 Q47
.50:	Q17 Q23 Q29 Q35 Q40	Q4 Q7 Q9 Q19 Q23 Q29 Q31 Q36 Q37
.60:	Q3 Q11 Q13 Q16 Q18 Q20 Q25 Q26 Q33 Q42 Q44 Q47	Q8 Q10 Q12 Q13 Q21 Q28 Q33 Q48 Q50
.70:	Q4 Q5 Q12 Q30 Q32 Q34 Q36	Q1 Q6 Q11 Q15 Q18 Q25 Q34 Q42 Q46 Q49
.80:	Q1 Q6 Q7 Q8 Q9 Q19 Q21 Q27 Q31 Q37 Q39 Q48 Q49	Q2 Q14 Q24 Q32
.90:	Q2 Q10 Q14 Q15 Q24 Q46 Q50	
ID	**5-option MC Format**	**4-option MC Format**
.00:		
.10:	Q3 Q13 Q23 Q38	Q7 Q14 Q45
.20:	Q14 Q33 Q45	Q2 Q5 Q13 Q15 Q26 Q32 Q39 Q42 Q46
.30:	Q15 Q28 Q46	Q3 Q18 Q19 Q21 Q24 Q41 Q49 Q50
.40:	Q1 Q5 Q6 Q7 Q8 Q9 Q10 Q20 Q22 Q24 Q39 Q43 Q49 Q50	Q22 Q25 Q28 Q30 Q34 Q37
.50:	Q2 Q4 Q11 Q16 Q18 Q19 Q21 Q30 Q34 Q35 Q37 Q44 Q48	Q1 Q6 Q8 Q10 Q11 Q17 Q23 Q36 Q38 Q40 Q43 Q44 Q48
.60:	Q12 Q25 Q27 Q31 Q36 Q41	Q4 Q12 Q29 Q33 Q35 Q47
.70:	Q32 Q42 Q47	Q9 Q16 Q20 Q27 Q31
.80:	Q17 Q26 Q29 Q40	
.90:		

Distractors Performance of the Two UN-like Tests

Distractor performance, as highlighted in Table 3, shows the number of functioning distractors and non-functioning distractors, which occur in the result of the two different MC English UN-like tests. The table shows that the 68% functioning-distractors in the four-option format UN-like test items were classified as the higher functioning-distractors when compared to the five-option format English UN-like tests with 36.5%. Similarly, 25% of items in the four-option test format had optimal functioning-distractors compared with only 2% of items on the five-option test format. Despite having fewer distractors, the four-option format had more functioning distractors per item than the five-option test format (2.372 times vs. 1.81 times).

Since it is expected that weak students choose the distractors instead of the right choice, negative coefficients of correlation for distractors were intensely wanted by the weak students. It means the greater the absolute value of negative correlations, the more discriminating the distractor is. Distractor performance revealed that 36.5 % of distractors in the five-option MC format and 68 % of distractors in the four-option format were of quite high discrimination power. Thus, it is concluded that as the number of options decreases, the discrimination power of distractors increases. Therefore, the investigation of distractor performance corroborated the results found from the other

similar studies (Tarrant, et al., 2009; Shizuka, et al., 2006) that revealed the number of options for the English UN-like test questions has an inverse effect on the discrimination power of distractors.

Table 3: Distractors performance

	4-option format	5-option format
No. of distractors with: n(%)	150	200
Non-Function Distractor (Frequency < 5%)	48 (32)	127 (63.5)
Functioning Distractors n (%)	102 (.68)	73 (.365)
Functioning Distractors per item n (%)		
None	7	9
One	8	15
Two	10	11
Three	25	13
Four	-	2
The Average of Functioning Distractors per item(*M)*	2.372	1.871

Test reliability of the two UN-like tests

Person separation Indices and Cronbach's Alpha reliability were applied to compare the two UN-like tests MC formats as indicated in Table 4. Person separation is used to indicate the number of ability strata that a test can distinguish. As we can see in the table, the person separation index in the items of the five-option format is only slightly higher than the items in the four-option format. Thus, the researcher concluded that the number of options for MC English UN-like test items has no significant impact on test reliabilities (Rao & Haque, 2019).

Table 4: Reliability and Person separation indices for the two UN-like tests MC formats

	N	Separation Index		Cronbach's Alpha Reliability	
		Items	Persons	Items	Persons
The five-option format	120	4.6764	2.8514	0.9563	0.8905
The four-option format	120	4.2242	2.5292	0.9469	0.8648

n= number of persons

Paired sample t-test Analysis results

A paired *t-test* analysis was used to compare two means in factorial studies (Bailey &Nunan, 2009). The Interpretation of the Paired *t-test* is shown below:

Table 5: Interpretation of SPSS statistical software for the paired sample correlation between the five- and the four-option MC formats (The t-test statistical result)

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	5-option MC format	34.93	120	9.046	.826
	4-option MC format	32.86	120	8.576	.783

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	5-option MC format & 4-option MC format	120	.954	.000

Paired Samples Test

		Paired Differences					t	Df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		t	Df	
					Lower	Upper			
Pair 1	5-option MC format - 4-option MC format	2.075	2.714	.248	1.584	2.566	8.376	119	.000

Interpretation:

From the paired sample test table above, this study achieved:

$$\underline{d} = \frac{\sum_{i=1}^{n} {}^{\wedge} d_i}{n} = \underline{X} - \underline{y} = 34.93 - 32.86 = 2.075$$

$$S_d = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n} {}^{\wedge}(d_i - \underline{d})^2} = 2.714 \text{ where } d_i = x_i - y_i, \text{ i} = 1, 2, 3\dots\text{n}$$

The SPSS statistical software counted Pearson correlation between the two variables: $r = 0.954$ and the hypothesis testing was conducted to find out whether the correlation is significant with *p-value*= 0.00. In this case, *p-value*= 0.00 is less than α= 0.05. Thus, the Pearson correlation is significant.

The mean difference between the five-option and the four-option formats is 34.93 – 32.86 = 2.075. The paired sample t-test= H$_o$: $\mu_{5-option} = \mu_{4-option}$, resulting t value = 8.376 with the degree of freedom= n-1= 120 – 1 = 119. The SPSS produced *p-value* for 2-tailed t-test = 0.000. The result of the *p-value* is less than α= 0.05. Thus, it gives strong evidence to reject Ho: $\mu_{5-option} = \mu_{4-option}$

It interprets that the mean of the five-option and the four-option formats are different. Therefore, the conclusion of this study is there is a statistically significant difference in the results of English UN-like tests, which were examined by using MC questions by 5-option and 4-option MC formats. The paired sample t-test results show that the 4-option format is more difficult than the 5-option format.

Questionnaires Data Analysis

The questionnaire's data were elicited from 15 SMA English teachers of whom 80% regularly wrote monthly tests and 20% rarely wrote monthly tests. The questionnaire data were also taken from 120 SMA students who took the UN-like tests in this study. The result of the students' questionnaire data analysis is firstly described as follows:

Table 6: Students' questionnaire data analysis (weaknesses)

No	Weaknesses	5- option n (%)	4- option n (%)	5- & 4- options n (%)
1	It takes long time to finish	93 (72.7)	7 (5.5)	2 (1.6)
2	There are many implausible *distractors*	147 (78.2)	11 (8.6)	2 (1.6)
3	It is easy to guess the correct option	28 (21.9)	7 (5.5)	-
4	It is NOT appropriate to be used in Senior High School	36 (28.1)	7 (5.5)	1 (0.8)

Table 7: Students' questionnaire data analysis (strengths)

No	Strengths	5- option n (%)	4- option n (%)	5- & 4- options n (%)
1	It is easy to answer	2 (1.6)	30 (23.4)	2 (1.6)
2	It gives more time to think	8 (6.3)	35 (27.3)	2 (1.1)
3	The quality of the options is good because the *distractors* are similar to the correct option	16 (12.5)	44 (34.4)	2 (1.6)
4	It is appropriate to be used in Senior High School	20 (15.6)	34 (26.6)	-
5	Others:			
	Used in International test	-	1 (0.5)	-

According to the students' responses from the questionnaire data analysis, they mostly preferred the 4-option format to be tested in the UN SMA. The students' maximum response for the strength of the four-option MC test format was 34.4% (n=120). Most students considered that the four-option MC format has good quality because this format consists of many plausible distractors. From the table, we can see that most of the students preferred fewer options to be used in the UN SMA and disliked the five-option format. Most students were against the five-option because this format contains many implausible distractors. Furthermore, it is also surprising that 28.1% of the students (n=120) thought that the 5-option is not appropriate to be used in Senior High School because this format wastes their time to read the illogical option during the test. Therefore, they considered the 5-option is not suitable to be implemented in the UN SMA tests.

The result of the teachers' questionnaire data analysis is described in Tables 8 and 9.

Table 8: Teachers' questionnaire data analysis (strengths)

No	Strengths	5- option n (%)	4- option n (%)	5- & 4- options n (%)
1	It contributes to efficiency, i.e. you can finish writing the Multiple-Choice items faster	2 (13.3)	5 (33.3)	-
2	It makes students finish the test faster	2 (13.3)	4 (26.7)	-
3	It gives students more time to think	11 (73.3)	1 (6.7)	3 (20)
4	The quality of the options is good because the *distractors* I made are similar to the correct option	3 (20)	11 (73.3)	-
5	It is appropriate to be used in Senior High School	4 (26.7)	9 (60)	-

Table 9: Teachers' questionnaire data analysis (weaknesses)

No	Weaknesses	5- option n (%)	4- option n (%)	5- & 4- options n (%)
1	It takes a long time to write the items (MC Questions) because I have to think of plausible distractors	15 (100)	-	-
2	It takes a long time for the students to finish the test	15 (100)	-	-
3	Most *distractors* I made are implausible	15 (100)	-	-
4	It is easy for the students to guess the correct option	2 (13.3)	-	-
5	It is NOT appropriate to be used in Senior High School	2 (13.3)	-	-

According to the teachers' maximum percentile responses for the MC test format which contributes to efficiency, 33.3% (5) teachers (n=15) chose the four-option format because this format can make them finish writing the MC items faster, 26.7% teachers considered that the four-option MC format makes their students finish the test faster. The teachers' percentile responses for the MC test format which gives students more time to think, 73.3% of the teachers (n=15) chose the five-option MC items. From the table, we can see that most of the teachers preferred fewer options to be used in the UN SMA and disliked the five-option format. Moreover, 13.3% of the teachers (n=15) stated that the 5-option is not appropriate to be used in Senior High School because this format wastes their time to write more plausible distractors to trap weak students and this MC format consumes much time to finish the test (Shizuka, et al., 2006). Therefore, they considered the 5-option is not suitable to be implemented in the UN SMA tests.

Teachers' and students' opinions on the most optimal MC test format

Based on the result of questionnaire data analysis, 53.3% (8) teachers (n=15) chose the four-option MC format as the most optimal number of options to be used in the UN SMA and 39.8% (51) of the students (n=120) chose the four-option MC format as the best number of options. We can conclude that the fewer MC options, especially the four-option MC test items, were mostly preferred by SMA students and English teachers.

DISCUSSION

Based on the data shown in table 10 below, a statistically significant difference was noted among the results of the UN-like test and questionnaire data analysis. The 5-option format is easier than the four-option format. The result is in line with the previous findings in this same area (Abad, et al., 2001; Baghaei&Amrahi, 2011; Dehnad, et al., 2014; Lee &Winke, 2013; Shizuka, et al., 2006; Tarrant & Ware, 2010). However, no significant difference was found from the two different MC test formats considering the Index of Fit and reliability of the questions. The findings showed that the number of options for English UN-like test questions has inverse consequences on the discriminating power of items that measure the distractors while the Index of Fit was unaffected. Surprisingly, person response behaviours were shown to be similar throughout the two MC test formats with a different number of options per test item. The result of the study negates the general belief which states that fewer options increase the students' chance of guessing the correct option which contributes to affect the reliability of the test (Shizuka, et al., 2006; Tarrant & Ware, 2010). It means students often guess the answer when the questions are too hard or they run out of time to finish the test. Moreover, the results showed no significant change in the reliability of two test formats with different numbers of MC options per item, but they revealed the significant difference in the difficulty of the two different tests (Baghaei &Amrahi, 2011; Dehnad, et al., 2014; Lee & Winke, 2013; Tarrant & Ware, 2010).

Table 10: The tabulation of the five-option and the four-option MC test results

	IF	ID	Coefficient Alpha Reliability	Functioning Distractors	Questionnaires data analysis	Mean	*Point Biserial*	*t-test analysis*
5-option MC format	0.70	0.39	0.91	0.36	14.1 (students) 40 (teachers)	34.93	0.43	easier
4-option MC format	0.66	0.36	0.88	0.68	39.8 (students) 53.3 (teachers)	32.86	0.39	more difficult

From the tabulation of the five- and the four-option data analysis results in table 8, the four-option format exceeds the five-option format in the number of functioning distractors (0.68 VS 0.36), mostly preferred MC test format based on the result of teachers' questionnaires data analysis (53.3 VS 40) and students' questionnaires data analysis (39.8 VS 14.1). Even, based on the value of the *mean* of the UN-like test scores (32.86 VS 34.93) and the value of IF (0.66 VS 0.70), the four-option MC format is harder than the five-option MC format because the less the value of the *mean* and the IF, the more difficult the test is.

The finding of this study which stated that the four-option MC items were harder than the 5-option MC format is interesting. Despite the fact that the chances of getting the correct answer without being familiar with the content evaluated in tests with the 4- and the 5-option formats are 25 % and 20 % respectively, the researchers noted that the guessing factor did not affect the Item Facility throughout the tests.

From the analysis, we can see the findings of the research regarding the optimal number of MC options for the English test items in the Indonesian National Examinations affirm the very previous findings (Bruno &Dirkzwager, 1995; Delgado & Prieto, 1998; Tversky, 1964) and subsequent empirical studies (Baghaei&Amrahi, 2011; Dehnad, et al., 2014; Lee &Winke, 2013; Haladyna& Rodriguez, 2013; Shizuka, et al., 2006; Tarrant & Ware, 2010) which recommend the use of fewer choices per item.

CONCLUSION

Having analysed the data from the results of the UN-like tests and questionnaires data analysis, we could conclude the research findings. The *t-test* result in this study indicated the five-option format is surprisingly easier to do than the four-option MC formats. This study revealed that the four-option is the most optimal MC format to be used in the UN SMA. Since the distractors were all plausibly constructed, the number of options is fewer than the five-option format, and students had more time to think out of the correct option. Thus, the four-option is the most optimal MC format to be used in the UN SMA.

Subsequently, based on the result of the UN-like tests statistical analysis in this study, the reduction of the MC options does not significantly affect the Item Facility in each item. The difficulty of the items remains the same between the five- and the four-option MC formats (Hohensinn&Kubinger, 2009). The reduction of the MC options is still able to differentiate strong students from weak students (Abad, et al., 2001; Baghaei&Amrahi, 2011; Dehnad, et al., 2014; Lee &Winke, 2013; Shizuka, et al., 2006; Tarrant & Ware, 2010). Both MC formats have the same discrimination power.

The result of distractors analysis in this study showed that the four-option format has more functioning distractors than the five-option format. Besides, there was a slight increment in the discriminating power of distractors as a consequence of decreasing the number of MC options in the English UN-like tests. Consequently, fewer options formats (the four-option MC format) seem to be optimal to be used in the UN SMA regarding the duration and attempt needed to create MC English tests with more options.

Based on the result of questionnaires data analysis, the four-option is the most preferable option to being used in the senior high school examinations for SMA students and SMA English teachers. Since the psychometric properties of MC tests with the four-option and the five-option MC formats were shown to be almost similar, this study strongly recommends the use of the four choices for MC English test items as a result of practicality-related issues. Regarding the matters like reducing test duration, fund, and energy, minimizing the risk of writing implausible distractors, and lowering the probability of providing clues which would test-wise the test-takers, the results of the present research have found to concord on the use of four options for MC English test items in the Indonesian Senior High School Examinations. Therefore, we suggest the test makers of English UN test within the Indonesia Ministry of Education and Culture use the four-option MC format for the Senior High School level in order to improve the quality of the national examinations and satisfy not only the students but also the test makers in implementing the tests (Paharia, 2019).

LIMITATION AND STUDY FORWARD

Acknowledging some areas that have not been touched in this study, we suggest some areas for future research. To the researchers' knowledge, this kind of study about the most optimal number of MC options is the first study conducted in Indonesia. The researchers only took 120 samples in this study. Therefore, other researchers could conduct a similar study by taking more samples. Further study should focus on the impact of reducing the number of options to find the

most favourable number of options for the primary school students or the junior high school students in doing English for the UN-test. It is also possible to inquire about the impact of reducing the number of MC options to find the most favourable number of options at the university level.

ACKNOWLEDGMENT

AUTHORS CONTRIBUTION

The authors of this article are Herland Franley Manalu and Diana Anggraeni. Herland's contributions are mostly on writing the article, analysing the data, and revising the manuscript. Diana's contribution is collecting the data and Herland supervised her with valuable suggestions for the completion of the manuscript.

REFERENCES

1. Abad, F. J., Olea, J., & Ponsoda, V. (2001). Analysis of the optimum number of alternatives from the Item Response Theory. *Psicothema*, *13*(1), 152-158.
2. Allison, D., & Palmer, A. (2010). *Language assessment in practice*. Oxford University Press.
3. Bachman, L.F. (2004). *Statistical analyses for language assessment*. Cambridge University Press. https://doi.org/10.1017/CBO9780511667350
4. Baghaei, P., & Amrahi, N. (2011). The effects of the number of options on the psychometric characteristics of multiple-choice items. *Psychological Test and Assessment Modelling*, 53(2), 192-211. https://doi.org/10.4304/jltr.2.5.1052-1060
5. Bailey, K.M., & Nunan, D. (2009). *Exploring second language classroom research: A comprehensive guide.* Heinle.
6. Boland, R.J., Lester, N.A., & Williams, E. (2010). Writing multiple-choice questions. *Acad Psychiatry*, *34*(4), 310-316. https://doi.org/10.1176/appi.ap.34.4.310
7. Bruno, J. E., & Dirkzwager, A. (1995). Determining the optimal number of alternatives to a multiple-choice test item: An information theoretic perspective. *Educational and Psycho-logical Measurement*, *55,*959-966. https://doi.org/10.1177/0013164495055006004
8. Costin, F. (1972). Three-choice versus four-choice items: Implications for Reliability and Validity of objective achievement test. *Educational and Psychological Measurement*, *32*, 1035-1038. https://doi.org/10.1177/001316447203200419
9. Dehnad, A., Nasser, H., & Hosseini, A. F. (2014). A comparison between three-and four-option multiple choice questions. *Procedia-Social and Behavioural Sciences*, *98*, 398–403. https://doi.org/10.1016/j.sbspro.2014.03.432
10. Delgado, A. R., & Prieto, G. (1998). Further evidence favoring three-option items in multiple-choice tests. *European Journal of Psychological Assessment*, *14,* 197-201. https://doi.org/10.1027/1015-5759.14.3.197
11. Dirgeyasa, I. W. (2017). Model answers for yes/no questions from EFL students in public senior high schools. *Studies in English Language and Education*, *4*(2), 235-246. https://doi.org/10.24815/siele.v4i2.7717
12. Epstein, R.M. (2007). Medical education: Assessment in medical education. *The New England Journal of Medicine, 356*(4), 387-396. https://doi.org/10.1056/NEJMra054784
13. Frankfort-Nachmias, C. &Nachmias, D. (2007). *Study guide for research methods in the social sciences,* 7th ed. Macmillan Higher Education.
14. Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education, 15*(3), 309–334. https://doi.org/10.1207/S15324818AME1503_5
15. Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating test items.* Routledge. https://doi.org/10.4324/9780203850381
16. Hohensinn, C. H., &Kubinger, K. D. (2009). On varying item difficulty by changing the response format for a mathematical competence test. *Austrian Journal of Statistics*, *38*(4), 231-239. https://doi.org/10.17713/ajs.v38i4.276
17. Hughes, A. (2003). *Testing for language teachers* (2nded.). Cambridge University Press.
18. Kubinger, K. D., Holocher-Ertl, S., Reif, M., Hohensinn, C., &Frebort, M. (2010). On minimizing guessing effects on multiple-choice items: Superiority of a two solutions and three distractors item format to one solution and five distractors item format. *International Journal of Selection and Assessment*, *18*(1), 111-115. https://doi.org/10.1111/j.1468-2389.2010.00493.x
19. Kubiszyn, T., & Borich, G. (2004). *Educational testing and measurement: Classroom application and practice* (7thed.).John Wiley & Sons, Inc.

20. Landrum, R. E., Cashin, J. R., & Thesis, K. S. (1993). More evidence in favor of three-option multiple-choice tests. *Educational and Psychological Measurement, 53*, 771-778. https://doi.org/10.1177/0013164493053003021

21. Lee, H., & Winke, P. (2013). The differences among three-, four-, and five-option-item formats in the context of a high-stakes English-language listening test. *Language Testing*, *30*(1), 99-123. https://doi.org/10.1177/0265532212451235

22. Masters, J. C., Hulsmeyer, B. S., Pike, M. E., Leichty, K., Miller, M. T., &Verst, A. L. (2001). Assessment of multiple-choice questions in selected test banks accompanying textbooks used in nursing education. *Journal of Nursing Education, 40*(1), 25-32.

23. McCoubrie, P. (2004). Improving the fairness of multiple-choice questions: A literature review. *Med Teach,26*(8), 709-712. https://doi.org/10.1080/01421590400013495

24. Owen, S.V., & Froman R.D. (1987). What's wrong with three-option Multiple-Choice items. *Educational and Psychological Measurement*, *47*, 513-522. https://doi.org/10.1177/0013164487472027

25. Paharia, P. (2019). A scale development for student satisfaction: A perceived dimension of quality education. *Humanities & Social Sciences Reviews*, *7*(1), 360-366. https://doi.org/10.18510/hssr.2019.7142

26. Rao, M. M., & Haque, M, I. (2019). A study on the impact of testing on English as a foreign language teaching in a Saudi Arabian university. *Humanities & Social Sciences Reviews*, *7*(2), 58-71. https://doi.org/10.18510/hssr.2019.727

27. Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests* (Expanded ed.). University of Chicago Press.

28. Redmond, S. P. 1., Hartigan-Rogers, J. A., & Cobbett, S. (2012). High time for a change: Psychometric analysis of multiple-choice questions in nursing. *International Journal of Nursing Education Scholarship, 9*(1), 1–16. https://doi.org/10.1515/1548-923X.2487

29. Rodriguez, M. (2002). Review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education,15*(3), 309-334. https://doi.org/10.1207/S15324818AME1503_5

30. Rogers, W.T., & Harley, D. (1999). An empirical comparison of three-and four-choice items and tests: susceptibility to test wiseness and internal consistency reliability. *Educational and Psychological Measurement, 59*(2), 234-247. https://doi.org/10.1177/00131649921969820

31. Schneid, S.D, Armour C., Park Y.S., Yudkowsky, R., &Bordage, G. (2014). Reducing the number of options on multiple-choice questions: Response time, psychometrics, and standard-setting. *Med Educ*., *48*, 1020–1027. https://doi.org/10.1111/medu.12525

32. Shizuka, T., Takeuchi, O., Yashima, T., & Yoshizawa, K. (2006). A comparison of three- and four-option English tests for university entrance selection purposes in Japan. *Language Testing*, *23*(1), 35-57. https://doi.org/10.1191/0265532206lt319oa

33. Sidick, J. T., Barrett, G. V., &Doverspike, D. (1994). Three-alternative multiple-choice tests: An attractive option. *Personnel Psychology, 47*, 829-835. https://doi.org/10.1111/j.1744-6570.1994.tb01579.x

34. Sujinah, S., Setyorini, D., Yunianti, S., N., &SaviraIsnah, E. (2019). Developing reading comprehension assessment-based Adobe Flash CC2015 for BIPA Intermediate Level. *Humanities & Social Sciences Reviews*, *7*(3), 460-466. https://doi.org/10.18510/hssr.2019.7367

35. Susanto, A. (2017). Assessing the relationship between the Vocabulary Level Test (VLT) and Reading Comprehension. *Studies in English Language and Education*, *4*(2), 157-171. https://doi.org/10.24815/siele.v4i2.5118

36. Tarrant, M., Ware, J., & Mohammed, A. M. (2009). An assessment of functioning and non-functioning distractors in multiple-choice questions: a descriptive analysis. *BMC Medical Education*, 9(1), 40-47. https://doi.org/10.1186/1472-6920-9-40

37. Tarrant, M., & Ware, J. (2010). A comparison of the psychometric properties of three-and four-option multiple-choice questions in nursing assessments. *Nurse Education Today*, *30*(6), 539–543. https://doi.org/10.1016/j.nedt.2009.11.002

38. Tversky, A. (1964). On the optimal number of alternatives at a choice point. *Journal of Mathematical Psychology*, *1*, 386-391. https://doi.org/10.1016/0022-2496(64)90010-0

39. Vyas, R., & Supe, A. (2008). Multiple choice questions: A literature review on the optimal number of options. *The National medical journal of India,21*(3), 130-133.

40. Wolfe, E. W., & Smith, E. V., Jr. (2007). Instrument development tools and activities for measure validation using Rasch models: Part II – Validation activities. *Journal of Applied Measurement*, *8*, 204-234.