

## THE COMPARISON OF TWO TEST ITEM FORMATS IN ASSESSING STUDENTS' READING COMPREHENSION ABILITY

FIDALIA MR  
SMAN 1 Kampar Utara, Kampar, Riau  
fidalia\_MR86@yahoo.com

### ABSTRACT

This study discussed the comparison of the two kinds of test in assessing students' reading comprehension skill. The sample of this research was 36 eighth grade students in one of the SMP-s in Riau. It employed quantitative research. The data collection technique used was reading comprehension test. Furthermore, the data obtained from both test item formats were analyzed, compared and calculated in form of paired samples t-test analysis by using IBM SPSS statistics version 22. The research findings showed some points; there are 4.5833 point differences between two mean scores result (short answer and multiple choice test), the correlation between the result of the two test item formats is .626 which means that there is high correlation between them, and based on the comparison of t value and t table and the P value, it was found  $H_0$  is accepted which means that there is no effect of the use of those two test item formats in students' score in assessing students' reading comprehension ability.

**Keywords:** comparison, testing, multiple choice, short answer, reading comprehension

### A. Introduction

Testing is one of aspects in evaluation of students' learning. It is used to collect the information related to the students' achievement and learning outcomes. The main goal of classroom testing and assessment is to obtain valid, reliable, and useful information concerning student achievement (Miller et al., 2009). Further, testing is one of the sections in the teaching and learning process. Test may be constructed primarily as devices to reinforce learning and to motivate the student or primarily as a means of assessing the student's performance in the language (Heaton, 1988). A language test which seeks to find out what candidates can do with language provides a focus for purposeful, everyday communication activities. A well-constructed classroom test will provide the students with an opportunity to show their ability to perform certain tasks in the language.

In conducting test, the testers need to be able to choose the appropriate test item formats used. Miller et al. (2009) stated that there are two kinds of classroom test item formats; objective test items and performance assessment. He added that objective test items are highly structured and require the students to supply a word or two or to select the correct answer from alternatives. The examples of objective test are multiple choice test, short answer test, completion, matching, true-false or Alternative Response. While performance assessment require the students to construct responses (e.g., write an essay) or perform a particular task (e.g., measure air pressure). However, the objective test items, particularly multiple choice test is the most widely used by the teachers. This is in line with the statement of Cheung and Bucat (2002:1) that the multiple-choice (MC) item is one of the most popular item formats used in educational assessment. Moreover, another test item format that is frequently used in classroom assessment is short answer test.

Furthermore, these two test item formats have their own strengths and weaknesses. Some research and theories have revealed about this case. Based on some research and theories, multiple choice test item format is one of the most popular ones to be used in educational

assessment in which it is simple, valid, reliable, and efficient. On the other side, some researchers also revealed about the use of short answer test, particularly in assessing students' reading comprehension skill. They found that this kind of test item format is an easy test to be constructed and being a good one to make the students answer the questions with minimal possibility of guessing.

These two test item formats can be used in assessing reading comprehension. The multiple choice test offers a useful way of testing reading comprehension (Heaton, 1988:117). Reading comprehension is one of the skills which should be mastered by the students. He added that the extent to which a test is successful in measuring what it sets out to measure depends largely on the effectiveness of each of the items used. Moreover, Miller et.al (2009:172) stated that the short answer test item is suitable for measuring a wide variety of relatively simple learning outcomes, including students' reading comprehension.

In short, this study discussed the comparison of two kinds of test items formats: multiple choice and short answer in assessing students' reading comprehension in a class of eighth grade students in a junior high school in Riau.

## **B. Literature Review**

### **1. Language Testing**

Testing consist of some components which need to be understood. Brown (2010:3) stated that there are some components of testing; (1) method – an instrument that consists of a set of techniques, procedures or items that requires performance on the part of the test-taker; (2) measure – a process of quantifying a test-taker's performance according to explicit procedures or rules (Bachman, 1990:18-19 in Brown 2010:3); (3) individual's ability, knowledge, or performance – testers need to understand who the test-takers are; (4) given domain – for example, in the case of proficiency test, even though the actual performance on the test involves only a sampling of skills, the domain is overall proficiency in a language as the general competence in all skills of a language.

In short, it can be stated that in doing testing, it is needed to consider of the method used, the procedures, the test-takers, and the domains of the test itself in order to get the valid and reliable information through testing.

### **2. Multiple Choice and Short Answer**

Multiple choice as one of the objective test is widely used in assessment. This is in line with the statement of Cheung and Bucat (2002:1) that the multiple-choice (MC) item is one of the most popular item formats used in educational assessment. Or rather, it is stated:

The multiple-choice test is a very flexible assessment format that can be used to measure knowledge, skills, abilities, values, thinking skills, etc. Such a test usually consists of a number of items that pose a question to which students must select an answer from among a number of choices. Items can also be statements to which students must find the best completion. Multiple-choice items, therefore, are fundamentally recognition tasks, where students must identify the correct response ([http://www.flaguide.org/cat/multiplechoicetest/multiple\\_choice\\_test1.php](http://www.flaguide.org/cat/multiplechoicetest/multiple_choice_test1.php))

Multiple choice items can provide a useful means of teaching and testing in various learning situations (particularly at the lower levels) provided that it is always recognized that such items test *knowledge* of grammar, vocabulary, and etc (Heaton, 1988:27). He added that multiple choice items offer a useful introduction to the construction of objective

tests. It means that multiple choice as a kind of objective test can be used in testing students' knowledge, particularly at the lower levels.

Moreover, the characteristics of multiple choice test consists of a problem and a list of suggested solutions (Miller et al., 2009:194). The problem given can be in form of a direct question or an incomplete statement and is called the stem of the item, while the suggested solutions may include words, numbers, symbols, or phrases, and are called alternatives (also called choices or options).

In addition, the initial part of each multiple choice items is known as the *stem*: the choice from which the students select their answers are referred to as *opinions/responses/alternatives* (Heaton, 1988:28). He added that one option is the *answer, correct option* or *key*, while the other opinions are *distractors*. The task of a distractor is to distract the majority of poor students (i.e. those who do not know the answer) from the correct option.

The second types of objective test item format to be compared in this study is short answer test. According to Miller et all (2009:172), the short answer item is supply-type test items that can be answered by a word, phrase, number, or symbol. The short item uses a direct question. It is suitable for measuring a wide variety of relatively simple learning outcomes. He (2009) added some points about the advantages and limitations of short answer The short answer test is one of the easiest to construct, partly because of the relatively simple learning outcomes it usually measures. A more important advantage of the short answer test item format is that the students must supply the answer. This reduces the possibility that the students will obtain the correct answer by guessing. They must either recall the information requested or make the necessary computations to solve the problem presented to them. Partial knowledge, which might enable them to choose the correct answer on a selection item, is insufficient for answering a short item correctly.

On the other hand, there are two major limitation of short answer test described by Miller et al. The first is the unsuitability for measuring complex learning outcomes. This is in line with the principle of the form of short answer test item format in which it only requires simple answer in assessing students' ability. So, it is just able to measure the students' simple learning output. The second one is difficulty of scoring. Unless the question is carefully phrased, many answers of varying degrees of correctness must be considered for total or partial credit. In this case, the teacher needs to decide wisely how to give score of the students' answers.

### **3. Reading Skill**

Reading is one of the four skills in language learning. According to Oxford Learner's Pocket Dictionary (2008: 366), reading is act of reading something that are intended to be read or way in which something is understood. In understanding or comprehending reading material, a good reader actively construct meaning as they read, and they also direct their own comprehension (Pressley, 2000; Snow, 2002; Paris, Lipson, & Wixson, 1983 in Blachowicz and Ogle, 2008:33) by using basic strategies and by monitoring their own understanding (Baker & Brown, 1984 in Blachowicz and Ogle, 2008:33). Strategies in reading activity is one of the elements to be focused on, as stated by Gebhard (2006:194) that reading includes discovering meaning in print, within a social context, through bottom-up and top-down processing, and use of strategies and skills. In short, reading comprehension refers to an act of reading materials which intends to construct meaning of the materials being read by using certain strategies.

People who do reading comprehension activity must be based on certain purposes. A reading course should cover these purposes – reading to search for information (including skimming and scanning), reading to learn, reading for fun, reading to integrate information, reading to critique texts, and reading to write (Nation, 2009:6). Each readers have different purposes depend on what they want to get from the reading materials, whether it is to get information, fun, and etc.

In comprehending reading materials, readers need to connect their background knowledge to what they read. This is in line with the statement of Klingner, Vaughn, and Boardman (2007:8) that reading comprehension is a multicomponent, highly complex process that involves many interactions between readers and what they bring to the text (previous knowledge, strategy use) as well as variables related to the text itself (interest in text, understanding of text types). When readers do the process of connecting their own background knowledge to the reading material, it will be easier for them to get involved more and comprehend the text or reading material more.

Reading comprehension involves taking meaning to a text in order to get meaning from the text (Alexander, 1989 as cited by Kusnadi (2009:13). He added that an individual may be said to comprehend a text completely when he/she can: (1) recognize the meaning of words and sentences of the text; (2) associate meanings, both denotative and connotative, from personal experiences with the printed text; (3) recognize how all these meanings and/or his perceptions of them fit together contextually; (4) make value judgement about, and based on, the reading experience. Those points show that in comprehending reading materials, a reader should actually get the meaning of the materials being read in depth through sequential steps.

### **C. Research Methodology**

This study employed quantitative research. Maykut and Morehouse (1994) in Bergman (2008:11) stated that quantitative research is based on observations that are converted into discrete units that can be compared to other units by using statistical analysis. This is in line with this study in which it was aimed to compare the use of two test item formats (multiple choice and short answer tests) in assessing students' reading comprehension by using statistical analysis, t-test analysis. The calculation of t-test analysis was conducted by using IBM SPSS statistics version 22.

This study was conducted on the third week of November and the first week of December of 2013 in one of the junior high school in Riau. The samples of the research were 36 of eighth grade students in that school which consist of 21 female and 15 male students. The samples of the research were asked to take two kinds of test item formats; multiple choice and short answer test in two separate days, two weeks apart. The tests were written based on the materials that they have learned in the first semester in eighth grade.

In line with the research design employed and the topic of the study, the data collection technique used is in form of reading comprehension ability test. Emilia (2011:12) reported that test is usually administered at the beginning of the research before the implementation of a teaching program and at the end of the program to test the effectiveness of the program or a technique used.

There are two test item formats that were conducted in this research; multiple choice and short answer test. Each test item formats consists of 30 questions. These questions were written based on the materials taught in the first semester for eight grades of junior high school based on school-based curriculum. There are three materials which include;

descriptive text (10 questions), recount text (10 questions), and invitation card (10 questions). The tests were conducted in two separate days, a week apart. This was aimed to ensure that there was no or at least decrease the possibility of any practice effect by taking the same answer in both tests.

The data in form of students' reading comprehension mean scores in both test item formats were analyzed and compared in order to find out which one of the two formats is more effective to be used in assessing students' reading comprehension, particularly in the research site. In finding this comparison, the researcher conducted a statistical analysis. One of ways to do this is by using t-test. The t-test is probably the most widely used statistical test for the comparison of two means because it can be used with very small sample sizes (Hatch & Farhady, 1982:108)

One way of doing t-test analysis is by using an application which is called SPSS. It originally named Statistical Package for the Social Sciences. In this study, the researcher used the IBM SPSS Statistics version 22. Through this application, it is found the t-test analysis which described the comparison of the mean scores of the two constituents to be mainly focused in this research; short answer and multiple choice test item formats.

This analysis would result in the testing of the hypothesis about the comparison of the two test item formats. So, after doing the calculation, it would be found the comparison of the effect between the two test item formats; multiple choice and short answer test.

The following are the hypotheses of the research:

H<sub>0</sub> : There is no effect of the students' score in the use of short answer and multiple choice test item formats in assessing students' reading comprehension ability.

H<sub>1</sub> : There is effect of the students' score in the use of short answer and multiple choice test item formats in assessing students' reading comprehension ability.

#### D. Research Findings and Discussions

The following are the result of the t-test analysis by using IBM SPSS Statistics Version 22.

**Table 1. Paired Samples Statistics**

	Mean	N	Std. Deviation	Std. Error Mean
Pair 1 SA	75.1667	36	17.88642	2.98107
MC	79.7500	36	13.67468	2.27911

Table 1 shows the paired sample statistics which compared in this research; Short Answer (SA) and Multiple Choice (MC). There are some points pointed out; mean, sample number (N), Standard Deviation, and Standard Error Mean. In this case, it is found that there are 4.5833 point differences between those two mean scores.

**Table 2: Paired Samples Correlations**

	N	Correlation	Sig.
Pair 1 SA MC	36	.626	.000

Table 2 shows the paired samples correlation. There are two main points in this part; the correlation and the significant. The first point is about the correlation. The correlation refers to the relationship between the pair. It is found that the correlation between two test item formats is 0.626.

In Table 2, it can be seen that the correlation between the result of the short answer and multiple choice test item formats is 0.626. It means that there is high correlation between those test item formats (Sugiyono, 2012). Furthermore, the correlation above (0.626) is significant. It is because Sig. is smaller than 0.05 and the observed r value (0.626) is bigger than the critical r value (0.361) with 5 percent level of significance.

**Table 3: Paired Samples Test**

	Paired Differences		t	df	Sig. (2-tailed)
	Mean	Std. Deviation			
Pair 1 SA -MC	-4.58333	14.16153	-1.942	35	.060

Table 3 shows some important points; mean (-4.58333), standard deviation (14.16153), standard error mean (2.36025). Further, Confidence interval shows the area in which there is a kind of difference students' grade in 95% of confidence level. In this case, at the lower level, the confidence level is as -9.37490 and at the upper level, it is as .20824. Besides, the degree of freedom was 35 and the t values was 1.942.

In another column, it is found the t value of the t-test analysis. In an article (in a web) which was written by Widhiarso), it is explained that there is such kind of t-test rule:

- a. Sig : p 0.05, means there is a difference result of the two variable in 5% significance level.
- b. Sig : p 0.01, means there is a difference result of the two variable in 1% significance level
- c. Sig : p > 0.05, means there is no difference result of the two variables.

In this case, t table distribution is calculated and  $\alpha = 5\% : 2 = 2.5\%$  (2 tailed) with df (n-1) or  $36-1 = 35$ . With 2 tailed test in  $\alpha$  is .025, so it is found that t-table is 2.030108. When we compare t value and t table and the probability, it is found that:

$$-t \text{ value} > -t \text{ table} \quad \text{and} \quad P \text{ value}$$

$$-1.942 > -2.030108 \quad \quad \quad 0.060 > 0.05$$

Based on the rules in point 6, it is found that H0 is accepted which means that there is no effect of the students' score in the use of short answer and multiple choice test item formats in assessing students' reading comprehension ability.

**E. Conclusions**

Based on the research findings, it is concluded that:

- 1. There are 4.5833 point differences between two mean scores of the testing result (short answer and multiple choice test item format).
- 2. The correlation between the result of the short answer and multiple choice test item formats is 0.626. It means that there is high correlation between those test item formats.
- 3. t table distribution is calculated and  $\alpha = 5\% : 2 = 2.5\%$  (2 tailed) with df (n-1) or  $36-1 = 35$ . With 2 tailed test in  $\alpha$  is .025, so it is found that t-table is 2.030108.

When we compare t value and t table and the probability, it is found that:

$$-t \text{ value} > -t \text{ table} \quad \text{and} \quad P \text{ value}$$

$$-1.942 > -2.030108 \quad \quad \quad 0.060 > 0.05$$

It means that there is no effect of the students' score in the use of short answer and multiple choice test item formats in assessing students' reading comprehension ability.

## References

- Bergman, Manfred Max. 2008. *Advances in Mixed Methods Research*. UK: Sage Publication Ltd
- Blachowicz, Camille & Ogle, Donna. 2008. *Reading Comprehension Strategies for Independent Learners*. NY: The Guilford Press
- Bull, V (Ed). 2008. *Oxford Learner's Pocket Dictionary Fourth Edition*. UK: Oxford University Press.
- Brown, H. Douglas & Abeywickrama, Priyanvada. 2010. *Language Assessment Principles and Classroom Practices Second Edition*. NY: Pearson Education, Inc
- Cheung, Derek & Bucat Robert. 2002. *How Can We Construct Good Multiple-Choice Items?*. Hongkong: Science and Technology Education Conference
- Gebhard, J. G. 2006. *Teaching English as a Foreign or Second Language a Self-Development and Methodology Guide*. USA: The University of Michigan.
- Hatch, Evelyn & Farhady, Hossein. 1982. *Research Design and Statistics For Applied Linguistics*. USA: Newbury House Publishers, Inc
- Heaton, J. B. 1988. *Writing English Language Tests*. UK: Longman Group
- [http://www.flaguide.org/cat/multiplechoicetest/multiple\\_choice\\_test1.php](http://www.flaguide.org/cat/multiplechoicetest/multiple_choice_test1.php) accessed on 12th December 2013
- Klingner, Janette K., Vaughn, Sharon & Boardman, Alison. 2007. *Teaching Reading Comprehension to Students with Learning Difficulties*. NY: The Guilford Press
- Kusnadi. 2009. *Improving Students' Reading Skill Through Interactive Approach*. Bandung: Indonesia University of Education
- Miller, M. David, Linn, Robert L., & Gronlund, Norman E. 2009. *Measurement and Assessment in Teaching Tenth Edition*. New Jersey: Pearson Education, Inc
- Nation, I.S.P. 2009. *Teaching ESL/EFL Reading and Writing*. New York: Routledge, Taylor & Francis
- Simkin, Mark G. & Kuechler, William L. 2005. *Multiple-Choice Tests and Student Understanding: What is the Connection?* USA: Decision Sciences Journal of Innovative Education
- Sugiyono. 2012. *Statistika untuk Penelitian*. Bandung: Penerbit Alfabeta
- Widhiarso, Wahyu. *Bab II Uji Hipotesis Komparatif*. Fakultas Psikologi UGM. Accessed in [http://widhiarso.staff.ugm.ac.id/files/membaca\\_t-tes.pdf](http://widhiarso.staff.ugm.ac.id/files/membaca_t-tes.pdf) on 0th of December 2013