

A Comparative Study on the Measuring Capacity of the Three Types of Cloze Tests

Jing-tao LIU

Heze College, Heze, China

This paper conducts a comparative study on the measuring capacity of the three types of cloze tests where the comparison variables of difficulty, discriminability, reliability, and validity were calculated and analyzed. The findings from the research are drawn as follows: Firstly, the test-takers have a significant difference, on the results of the three types of cloze tests, that the test-takers perform much better on the multiple-choice cloze while having the worst performance on open cloze. Secondly, based on the comparison of the variables of difficulty, discriminability, reliability, and validity, it can be concluded that both open cloze and banked cloze have the desirable measuring capacity with strong discriminability power, high reliability and validity, while the most frequently used multiple-choice cloze does not perform a satisfied result on all of the comparison variables. Thirdly, open cloze yields the most effective and reliable results in measuring the test-takers' overall language proficiency.

Keywords: multiple-choice cloze, open cloze, banked cloze, measuring capacity

Introduction

According to Gestalt psychology, Taylor, the German psychologist, designed cloze tests in 1953. The cloze test was firstly used to test readability of native speakers. Based on this theory, psychologists stated that when people observe the shape of a familiar object without some parts of it, they will have the tendency that they regard it as a whole and add the missing parts so that they can see it as a complete one. And, as we all know, natural languages contain much redundant information for the sake of full understanding among people's communication. Cohen (1980) argued that people may still be able to understand the main idea of the whole text and even can recover the missing parts by using linguistic knowledge, textual knowledge, and knowledge of the world if some parts with redundant information are cancelled in a written text. Therefore, people's reading ability can be measured in a cloze test in which people can decode those interrupted messages by making the most acceptable substitution from all the contextual clues.

Cloze test has been widely used at home and abroad since it was created by Taylor. Its convenient operation and facilitating standards of measurement ensure its wide application in various tests. Its broad application also arouses researchers' interests. Many linguists have turned their attention to the topics like categories of cloze, validity of cloze, language proficiency that cloze measures, and test-takers' cognitive processes in a cloze test.

Significance of the Study

Some linguists and experts on language tests have reached the identical viewpoint that the cloze test could be used as an assessing tool to measure the language ability of the second language learners. However, more scholars still have different ideas about what type of cloze test is the most effective and efficient and about what influence the different cloze tests will have on the test-takers' performance.

The researcher hopes that this study can bring some new light on English tests, teaching, and learning. Firstly, it can deepen our understanding about the three types of cloze test formats and the cloze procedure. Secondly, it can draw some conclusions on the three types of cloze tests based on the scientific experiment to help test designers, users, or administrators to decide which type of cloze testing format is more suitable for their purpose. Thirdly, it is hoped to contribute to improving the quality and effectiveness of the English cloze tests in China.

Research Questions

The focus of this paper is to research on the measuring capacity of three types of cloze tests in the general English proficiency. Three research questions are put forward in this study:

- (1) Are test-takers' performances different when different test formats of cloze are utilized?
- (2) Are there any variances in difficulties, discriminabilities, validities, and reliabilities of multiple-choice cloze, open cloze, and banked cloze?
- (3) Which kind of cloze test can yield the most effective and reliable results of testees' overall language competence?

Research Methodology

Subjects

The researcher solicited 30 subjects from the juniors of Qufu Normal University. They were chosen randomly from the four parallel classes of a total number of 160 students. The four classes are all ordinary classes with students in all language proficiencies. All of the 30 students have had experienced many English tests, such as CET4 (College English Test Band-4) and CET6 (College English Test Band-6) which contain the multiple choice cloze format, so they are familiar with the MC (Multiple-choice) cloze. For open cloze and banked cloze, although they do not often do these kinds of tests, they are familiar with the formats and know how to do the tests. Besides, all the subjects share the same social and educational background. They are all native Chinese speakers, live under the same social customs, and receive similar cultural values. All of the above ensure the homogeneity of the subjects. One of the most important factors is their willingness to participate in the experiment, which makes sure the cloze tests reflect their real language proficiency and the questionnaire and retrospective report are reliable and valid. In the research, they are required to take the cloze test which consists of multiple-choice cloze, open cloze, and banked cloze and finish the questionnaire about the three types of test.

Retrospective Report

The retrospective report in this study will be implemented. All of the subjects will take the cloze test, which consists of three types of cloze tests: multiple-choice cloze, open cloze, and banked cloze. The three types of cloze tests are developed from a 230-word passage chosen from College English Test Band-4 which requires that

the test-takers should have studied college English for three years at least and mastered no less than 5000 vocabulary. Since the subjects are in their third years in college and most of them have passed CET4, the difficulty level of the passage is suitable for them. The same passage is developed into three types of cloze tests with 20 blanks. For the multiple-choice cloze, there are 20 blanks and each contains four options with one correct answer and three other distractors. For banked MC cloze, 25 options are provided in a word bank in which five words are distractors. For open cloze, there are only 20 blanks for subjects to fill in according to the content of the text without options provided.

All of the test-takers were informed that the purposes of the two experiments were research on language-testing and that no grading or placement would be made. Furthermore, their names and scores would be kept confidential.

After answering a series of questions like the above ones, the testees will present their explanations, which are derived from their reading and comprehending process of the passage. Then by analyzing their explanations, the researcher can get a clearer knowledge of how the testee reads and decodes the given passage and to what extent he understands it. Moreover, the researcher will find out to what extent each of the three cloze tests measures the testees' language ability underlying the context, that is to say, the research will figure out the construct validity of the three cloze tests by the retrospective report.

Questionnaire

Questionnaires are used mostly to collect data on phenomena which are not easily observed, such as the learner's attitude, motivation, etc. The questionnaire is carried out after the cloze test. It is designed to retrieving the testees' general view on the three types of cloze tests after they finish the test to serve for the research of the validity of the three cloze tests. The questionnaire was composed of 16 items, each of which consisted of a short statement and three or four options expressing different degrees of the student's confirmation of the statement. Four items were designed to investigate into their attitudes toward three cloze types' testing ability. The other 12 were supposed to extract from the subjects' information about their opinions and reflections to the testing effects of different cloze tests. Before question items, brief introductions to MC cloze, banked cloze, and open cloze are available so that students can get a general idea about these three cloze types since they rarely see the later two in their study or exams as a rule. Under the directions and explanations of the classroom teacher, each student finished all the 18 items of the questionnaire by circling the options which best represent their own opinions or attitudes toward these three different cloze tests.

Administration Procedures

In order to testify the research purposes, quantitative and qualitative analysis will be used. For quantitative analysis, test scores were processed with the the SPSS14.0 (Statistical Product and Service Solutions). And for the qualitative analysis, a questionnaire and a retrospective report analysis were constructed.

The experiment consisted of the following components: Cloze Test administration including Open Cloze, Multiple-choice Cloze, and Banked cloze, the questionnaire survey, and the retrospective report. In order to facilitate a clear understanding of the task and to ensure the validity of the experiment, before distributing the test papers to the students, the researcher clarified that the purpose of the experiment was for research only. The examinees were informed that the scores of the test would not be taken into their academic record by the

university and cheating in the exam would be forbidden. In addition, the participants were asked to finish the test papers without referring to any electronic dictionaries and the questions of the questionnaire should be answered honestly. And in order to make a clear understanding of the task, the instructions were given to the students verbally as well as in written form.

All of the three experiments were administered to the junior students in the regular class. One hour after the subjects finished the cloze test, the questionnaire was given out and all of the 30 questionnaire papers were collected. Then, the retrospective report was carried out. Because the introspection approach was new to the students, a pre-training about this method is necessary.

Results and Discussion

Comparison of the Descriptive Statistics

The following table shows the basic descriptive statistics of the three types of cloze tests.

Table 1

Comparison of Means and S.D.

	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance
OC	30	12.00	3.00	15.00	9.6000	2.76181	7.628
MC	30	11.00	9.00	20.00	15.1667	2.23860	10.489
BC	30	10.00	9.00	19.00	14.1667	2.56076	6.759

Notes. OC: Open cloze; MC: Multiple-choice Cloze; BC: Banked Cloze.

From Table 1, it can be clearly seen that the mean of the MC cloze test is much higher than the other two, but its standard deviation is the lowest in the three types of cloze items, which is just 2.23860. While the mean of the open cloze test is a little lower than that of banked cloze, but its standard deviation is the biggest one among the three and much closer to that of banked cloze. In the banked cloze test, the mean is 14.1667 which is the best situated one among those three and its standard deviation is the middle one of the three. Since standard deviation is a measure that shows how much the scores are spread out on each side of the mean, a high standard deviation may be an indication that the test discriminates effectively between candidates. So we can conclude that multiple-choice cloze is the easiest one and has the lowest discriminability; open cloze is the most difficult one and have the highest discriminability; banked cloze is the middle one on the difficulty and discriminability.

Table 2

Paired Samples Test

	Paired Differences						t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference					
				Lower	Upper				
Pair 1 OC - BC	-4.56667	2.68692	.49056	-5.56998	-3.56335	-9.309	29	.000	
Pair 2 OC - MC	-5.00000	3.64833	.66609	-6.36231	-3.63769	-7.506	29	.000	
Pair 3 BC - MC	-.43333	3.37008	.61529	-1.69174	.82508	-.704	29	.487	

From Table 2, it shows that the value of Sig. between open cloze and banked cloze is “.000 < .01”, which proves that there exists a significant difference at the 0.01 level between the mean score of open cloze and that of

banked cloze. And the value of Sig. between open cloze and multiple-choice cloze is also “.000”. The double values of Sig., “.000”, indicate that the testees’ performance between open cloze and the other two types is significantly different. However, the value of Sig. between banked cloze and multiple-choice cloze is “.487 > .05”, which means that there is not a significant difference between the performance of the two type of tests.

Correlation Analysis of Open Cloze, Multiple-Choice Cloze and Banked Cloze

After collecting every participant’s total scores respectively in open cloze, multiple-choice cloze, and banked cloze, SPSS 14.0 was employed to conduct a correlation analysis.

Table 3

Correlations Between OC, BC, MC Cloze and the Total Scores

		OC	BC	MC	Total
OC	Pearson Correlation	1	.492(**)	.258	.775(**)
	Sig. (2-tailed)		.006	.168	.000
	N	30	30	30	30
BC	Pearson Correlation	.492(**)	1	.332	.754(**)
	Sig. (2-tailed)	.006		.073	.000
	N	30	30	30	30
MC	Pearson Correlation	.258	.332	1	.741(**)
	Sig. (2-tailed)	.168	.073		.000
	N	30	30	30	30
Total	Pearson Correlation	.754(**)	.775(**)	.741(**)	1
	Sig. (2-tailed)	.000	.000	.000	
	N	30	30	30	30

Note. ** Correlation is significant at the 0.01 level (2-tailed).

Table 3 shows that all these three types of cloze are positively correlated with the total score with the value of Sig., “.000 < 0.01”, which indicates that the correlation is significant at the 0.01 level (2-tailed). The correlation coefficients of the open cloze and the banked cloze with the total score are respectively “.775(**)” and “.754(**)”, which are much higher than that of MC cloze with total score, “.741(**)”. The correlation coefficients of the open cloze and total score is the highest one, “.775(**)”, which indicates that they are the most closely correlated; while the MC cloze is the least correlated cloze type to the total score.

Construct Validity

The cloze test is an integrative test; it not only measures students’ grasp of language knowledge including the vocabulary discrimination, phrases usage, and grammar mastery, but also effectively measures their context ability which refers to their reading comprehension and language usage ability. The test-takers are required to comprehend the whole article including the theme, the orientation, the writer’s attitude, the context, between sentences, within the sentences and the envisions, and the inferences based on their language knowledge. They are then required to make the proper choices and complete the blanks through analysis and judgments.

Table 4

Retrospective Report Performances of MC

	Comprehend the context information	Discriminate the four options	Frequency	Proportion
Correct Answer	√	√	400	82%
	√	×	61	12%
	×	√	17	4%
	×	×	11	2%
Total frequency			489	100%
Incorrect Answer	√	√	0	0%
	√	×	53	48%
	×	√	39	35%
	×	×	19	17%
Total frequency			111	100%

According to the definitions given by Bachman (2000) and Cohen (1984) and their research on the construct validity of TEM-4 (Test for English Majors-4) cloze test, CHEN Xiao-kou and LI Shao-shan (2006) tentatively proposed a formula to calculate the construct validity. The formula is “ $CVI = CTP/TTP$ ”. Where, CVI is short for Construct Validity Index; CTP is short for Correct Test-Taking Process; TTP is short for Total Test-Taking Process (CHEN & LI, 2006). According to the results, there are 400 times that the test-takers comprehend and employ the relevant context information and discriminate the options correctly, which accounts for 82% of the total correct answers and there are total 600 times test-taking process. Therefore, the construct validity of the multiple-choice cloze is “.67 (400/600)”.

To sum up, the construct validity of the open cloze is the highest among the three types of cloze tests, while the construct validity of the multiple-choice cloze is the lowest. That is because both the multiple-choice cloze and banked cloze provide options including correct answers to test-takers and the test-takers can not only employ the relevant context information to comprehend the text, but they can apply test-wise strategies to guess the correct answers. Therefore, the construct validity of multiple-choice cloze and banked cloze becomes lower than that of open cloze.

Conclusion

According to the descriptive results and paired sample t-test, it can be concluded that the three types of cloze tests have great effects on the performance of the subjects. It can be known that the multiple-choice cloze has the highest mean scores but lowest standard deviation; open cloze has the lowest mean scores but highest standard deviation; banked cloze is the middle one both in mean scores and standard deviation. The results indicate that the multiple-choice cloze is the easiest test while the open cloze is the most difficult cloze test to the subjects and the banked cloze is the middle one. However, known from the standard deviation, the multiple-choice has the lowest discrimination ability while the open cloze has the greatest discrimination ability and banked cloze is also the middle one.

Some linguists and experts on language tests have reached the same opinion that the cloze test could be used as a tool to measure the language proficiency of the second language learners. Heaton (2000, p. 132) supported that the cloze test is effective in assessing readers' global comprehension ability. However, scholars still have different ideas about what type of cloze test is the most effective and efficient one to measure the language

learners' overall language proficiency. In the present study, correlations between OC, BC, MC cloze and the total score and correlations between OC, BC, MC cloze and Public English Test Band 2 (PET2) proved that open cloze can measure the students' overall language proficiency. The correlation results show that open cloze has the highest correlation coefficient of ".775" with the total score of the cloze test and the significant value is ".000". It can be seen that the open cloze has the greatest correlation coefficient with the scores of PET2 and the correlation is significant at the 0.001 level. It can be known from the two correlation test that the results from the cloze test can best reflect the general language proficiency of the subjects.

References

- Bachman, L. F. (2000). Modern language testing at the turn of the century: Assuring that what we count counts. *Language Testing*, 17(1), 1-42.
- CHEN, X. K., & LI, S. S. (2006). On the construct validity of TEM-4. *Modern Foreign Languages*, 29(1), 71-77.
- Cohen, A. D. (1980). *Testing language ability in the classroom*. New York: Newbury House.
- Cohen, A. D. (1984). On taking language tests what the students report. *Language Testing*, 1, 70-81.
- Heaton, J. B. (2000). *Writing English language tests*. Beijing: Foreign Language Teaching and Research Press.