

Teaching Evaluation System on English Text Reading

Rohani Ganie^a, Syahron Lubis^b, Tengku Thyrhaya Zein^c, ^{a,b,c}University of Sumatera Utara, Indonesia, Email: rohani@usu.ac.id, ronlubis@gmail.com, t.thyrahaya@usu.ac.id

This study discusses the differences in results obtained between three evaluation tools in reading lessons, the comprehension question (CQ), multiple choice (MC), and cloze passage (CP), and ascertains which evaluation tools show the highest results. This study was designed using an alternative measuring tool by connecting the validity of test cloze and tests C as a measure of the ability to read English text. The text is taken from the reference book "Contact U.S.A Reading and Vocabulary Textbooks." The information obtained is useful for developing reading techniques and teaching strategies, especially for remedial purposes. The results of this study are also useful for the development of English teaching and the application of applied linguistics in English research. The findings in this study prove that the differences between the three types of evaluation in this study in the form of pairing, namely CQ-MC, MC-CP, and CQ-CP, are not significant. This shows that the evaluation results of the three types of evaluation tools do not have much difference. Although there is no significant difference between the three evaluation tools used, this does not mean that all three have the same advantages. The research conducted has not been able to provide answers to questions as to which form of evaluation tool is better used in reading and which can show the highest results. This is because each form of evaluation has an unequal emphasis and therefore, the three forms of evaluation should be given variedly to students learning to read. In addition to reducing the level of boredom in students, the use of a variety of evaluation tools gives teachers clarity on students whose performance' are not always uniform. Teachers can then find out the form of exercises that still give difficulties to students in achieving teaching indicators.

Key words: *Evaluation, Reading, Comprehension Question, Multiple Choice, Cloze Passage.*



Introduction

English is an international language and the official language of several countries in the world, like Britain, America, Australia, New Zealand, South Africa, Grenada, Singapore, and so on. With this distribution, English has a number of language variations. Geographically, variations in English are determined by the geographical location and the influence of the language environment. The Australian dialect, for example, absorbs a number of lexical elements from Aboriginal languages in Australia. Likewise, English in Singapore contains several elements of Malay, Chinese, and Indian.

English occupies a very important position in relations between nations because English is one of the most widely used communication tools in the world. According to a survey from the English Proficiency Index, the Netherlands has emerged as a country that has the highest level of English proficiency, with a value of 72. It surpasses the five other northern European countries that are ranked at the top. In fact, the only non-European country included in the top ten was Singapore, which was ranked sixth. In Indonesia English is the first foreign language and is taught officially from junior high school until the first years of tertiary education. In addition, English Language at the university level has long occupied a special position as one of the disciplines offered in the humanities. At the Faculty of Cultural Sciences, Universitas Sumatera Utara, for example, there is a major in English Literature that provides a Language and English Literature curriculum.

One of the English language courses given to students majoring in English Literature at the Faculty of Cultural Sciences, *Universitas Sumatera Utara*, is devoted to reading. This course is taught from semester one for four semesters with a total credit load of eight SKS. In general, it can be said that the reading course aims to provide students with skills to understand the information contained in the text material, such as skimming, scanning and intensive reading.

Skimming is to search text quickly to get the essence of the text, scanning is searching text to find certain pieces of information, while intensive reading is reading short texts to get specific information. The reading activities here are more detailed. To find out whether the curricular goals have been successfully achieved, instructional activities are usually given an evaluation. In this case there are several tools for evaluation that can be used in accordance with the instructional objectives that have been set; because each evaluation has a form and pattern that is not the same, a comparison between the tools used to determine the level of student success is needed in each of these evaluation systems.

Many studies on the validity of cloze tests as a reading comprehension test have tended to focus on test scores; process-oriented approaches have not often been taken up. However,

other researchers have investigated the processes of taking these types of tests by examining verbal reports from test-takers. Each study focuses on different aspects of the test-taking processes and used various types of word deletion techniques and data collection methods.

In one study, (Saeedi, M., Tavakoli, M., Kazerooni, S., R., & Parvaresh, V., 2011) the authors examined a paper about *Do C-Test and Cloze Procedure Measure what they Purport to be Measuring? A Case of Criterion-Related Validity*. This article investigates the validity of C-test and the cloze test, both of which purport to measure general English proficiency. To provide empirical evidence pertaining to the validity of the interpretations based on the results of these integrative language tests, their criterion-related validity was investigated. The English as a foreign language (TOEFL) test, an established, standardised, and internationally administered test of general English proficiency, was used as the criterion measure. Some 90 Iranian English majors participated in this study. They were seniors studying English at a university in Tehran, Iran. The results of analyses shows a statistically significant correlation among participants' scores on the cloze test, C-test, and the TOEFL test. Building on the findings of the study and considering criterion-related validity as the evidential basis of the validity argument, it was cautiously deduced that these tests measure the same underlying trait. However, considering the limitations of using criterion measures to validate tests, no absolute claims can be made as to the construct validity of these integrative tests.

Importantly, Brown's (2013) paper about *My twenty-five years of cloze testing research: So what?* examines one strand of my research (a dozen or so studies on cloze testing) published between 1978 and 2002 in terms of: (a) the results of these studies, (b) the questions each raised that led to further studies, and (c) the answers I now have to the following questions: (1) How do score distributions generally affect cloze results? (2) How do score distributions affect cloze reliability and validity statistics? (3) Does tailoring cloze, and using item analysis, lead to a more reliable and valid test? (4) Are cloze tests basically sentential or extrasentential? (5) Why does K-R21 consistently underestimate cloze reliability? (6) Are there cloze items that are not contributing at all to test variance? (7) How do cloze tests administered to low and high proficiency groups differ overall? (8) How many items are not functioning well in cloze tests administered to low and high proficiency groups? (9) Are different items functioning well in cloze tests administered to low and high proficiency groups? The discussion summarises answers to those nine questions, and the conclusion considers how (a) I learned from my mistakes, (b) I learned more than I set out to learn, and (c) the whole is greater than the sum of the parts.

Saeedi's (2016) paper, *Construct Validity of Multiple-Choice Cloze Test and Cloze-Elide Test in Testing Reading Comprehension among Iranian EFL Learners*, finds that Cloze tests have been extensively used for measuring readability, language proficiency and reading

comprehension. The study attempts to determine and cross-compare the construct validity of multiple-choice cloze tests and cloze elide tests in testing reading comprehension via a differential group design. Two instruments were developed and used: a multiple-choice cloze test and a cloze elide test. The tests were administered to sixty, BA and MA EFL Iranian students. The results indicate that a) multiple-choice cloze tests are valid in testing reading comprehension, b) cloze elide tests are valid in testing reading comprehension, c) BA and MA performances on cloze elide tests were better than BA and MA performances on multiple-choice cloze tests.

Based on the study above, the writer attempts to contribute information for improving teaching English, especially for reading skills. Information might be used to arrange exercises that are appropriate for this skill.

Aims and Scope

This study aims to obtain a picture of the differences in the results of evaluations from three types of evaluation tools or exercises that can be used in reading. In addition, it is also expected to be able to obtain an overview of evaluation tools that can show the highest level of results. Information like this can be used in developing techniques and strategies for teaching reading, especially for remedial purposes.

This research provides benefits for the development of science and teaching English. The results of this study not only provide input that can develop theories of teaching methods in reading in particular, but also contribute to the teaching of English as a foreign language, which has a very important role in the development of science and technology in general. The results of this study can also provide great benefits for the application of applied linguistics and teaching English as a foreign language in Indonesia, especially relating to reading skills problems.

A teacher always gives an evaluation in the form of exercises to students after reading and discussing a text. To find out whether students have understood the information contained in the text, some form of exercise is often used. Thus, it is necessary to know the difference in results achieved between the several evaluation tools used.

This research is limited to analysing the results without regard to factors such as the shape and pattern of the evaluation tools used or other, external factors.

Review of the Literature

In this case, reading is the activity of receiving, analysing, and interpreting that is done by the reader to obtain the message conveyed by the writer in the written media. There are various notions of reading, but in the narrow sense, reading is the activity of understanding the meaning contained in writing. In a broad sense, reading is a critical and creative reading processing that is carried out by the reader to gain a thorough understanding of the reading, which is followed by an assessment of the circumstances, values, functions and effects of the reading (Nurhadi, 2016).

Furthermore, Tarigan (2015) concludes there are broadly two important aspects of reading, namely:

1. Skills that are mechanical (mechanical skills) that can be considered to be in a lower order.

Aspects of mechanical skills include:

- a. Introduction to letter shapes
- b. Introduction of linguistic elements (phonemes / graphemes, words, phrases, clause patterns, sentences, etc.)
- c. Introduction of spelling and sound correspondence / correspondence (the ability to voice written material or "to bark at print")
- d. Reading speed is slow.

2. Comprehensive skills that can be considered to be in a higher order. They include:

- a. Understanding simple terms (lexical, grammatical, rhetorical)
- b. Understanding the significance or meaning (the intent and purpose of the author, the relevance / state of culture, and the reaction of the reader)
- c. Evaluation or assessment (content and form)
- d. Flexible reading speed, which is easily adapted to circumstances.

Broughton et al, (1978: 211) explain that to achieve the goals contained in mechanical skills, the activity most suitable is reading aloud, (or oral reading). For comprehension skills, the most appropriate is silent reading, which can also be divided into:

1. Extensive reading that includes:

- a. Reading survey

- b. Skimming
 - c. Superficial reading
2. Intensive reading is divided into:
- a. Reading content study, which includes:
 - 1) Close reading
 - 2) Reading comprehension
 - 3) Critical Reading
 - 4) Reading for ideas
 - b. Reading a language study, which includes:
 - 1) Foreign language reading
 - 2) Literary reading

According to the explanation above, aspects of reading can include mechanical and comprehensive skills. But before we are able to read well, we must first understand the mechanical aspects. This is so that we know and recognise the shape of letters, linguistic elements and spelling patterns. After that, we develop reading skills that understanding the text.

Some types of texts in general can be text in a novel, short stories, drama scripts, poems, rhymes, letters, advertisements, reports, reference books, food recipes, comics, dictionaries, and others. There are two main reasons why people read them: for pleasure (reading for pleasure), and for obtaining information (reading for information).

Some studies support the validity of cloze tests as a reading comprehension test, Maroko's (2016) paper *Uncovering Cloze Testing Practices in Selected Practice Papers for Primary Schools in Kenya*, focuses on reading comprehension that is probably the most prominently tested skill in upper primary classes in Kenya. Most test papers, for instance, comprise a fifteen-item cloze test, a fictional passage and a factual one, both of which contribute at least twenty items. While the latter two passages mainly focus on the learner's ability to identify correct answers from the passage, a cloze test taker is expected to provide words that have been deliberately omitted. Assessment may depend on whether the exercise is objective (i.e. students are given multiple choices to use in a cloze) or subjective (i.e. students fill in a cloze test with words that would make a given sentence grammatically correct). Consequently, cloze tests require the ability to understand context and vocabulary in order to identify the correct words or type of words that belong in the deleted parts of a text. However, cloze tests are probably the least understood aspect of language testing in schools. Thus, this paper seeks to address the following issues: What aspects of language are tested in cloze tests and how is

this done? How are multiple choice questions constructed? What contextual clues are evident in cloze passages? It is hoped that answers to these and other issues could have implications for testing practices and also the development of essential skills for answering such tests.

Septiarini and Hikmah's (2017) study *Multiple-Choice and Cloze Tests in EFL Reading Comprehension: What Do They Effectively Measure?* focuses on the research objective of figuring out the differences of multiple-choice and cloze tests in EFL reading comprehension and finding out which one is the most effective measurement in EFL reading comprehension. The subjects were 85 adults, EFL advanced learners at UNTIRTA. This research uses a quantitative method. A causal comparative research was conducted to determine the effective measurement, either multiple-choice test or cloze test in EFL reading comprehension. Data were collected by a random sampling technique. Data analysis uses the comparison of independent samples test (T-test) and one-way Analysis of Variance. The independent samples comparison of CT is - 0,06. A significant (sig) of F-test is 0,000. The sig F-test result is 0,000 which refers to $p > 0,05$. This means there is diversity between the multiple-choice test and the cloze test. The sig (2-tailed) T-test result is 0,026 which refers to $p > 0,05$. This means there is diversity of comparison. The mean difference between the multiple-choice test and the cloze test is - 0,89412. This means the multiple-choice test is lower than the cloze test in EFL reading comprehension.

Kleijn, Maat, Pander and Sanders' (2019) paper *Cloze testing for comprehension assessment: The HyTeC-cloze*, found that while there are many methods available for assessing text comprehension, the cloze test is not widely acknowledged as one of them. Critiques on cloze testing centre on its supposedly limited ability to measure comprehension beyond the sentence. However, these critiques do not hold for all types of cloze tests; the particular configuration of a cloze determines its validity. The authors review various cloze configurations and discuss their strengths and weaknesses and propose a new cloze procedure specifically designed to gauge text comprehension: The Hybrid Text Comprehension cloze (HyTeC-cloze). This cloze test employs a hybrid mechanical-rational deletion strategy and semantic scoring of answers. The procedure was tested in a large-scale study, involving 2926 Dutch secondary school students with 120 unique cloze tests.

Methodology

This study is a type of development research which is exploratory in nature. This study was conducted using a mixed-methods design that uses correlational techniques by connecting the test cloze score, test C, with a standard reading test score used as a benchmark, the Reading Comprehension test from the reference book. As Creswell and Plano Clark (2007) define, mixed methods design is a procedure for collecting and mixing both quantitative and qualitative methodologies and methods in a single study to understand a research problem.

The basic assumption is that the use of both quantitative and qualitative methods, in combination, provided a better understanding of the research problems and questions. In addition, the scores of each of the three tests are correlated with the self-assessment scores for each test.

According to Creswell and Plano Clark (2007), an explanatory mixed method design consists of firstly collecting quantitative, and then collecting qualitative data to help explain or elaborate on the quantitative results. The rationale for this approach is the quantitative data and results provide a general picture of the research problem; more analysis, specifically through qualitative data.

So that the objectives of this study can be achieved, in this study the authors use a measuring instrument commonly used to measure the ability to read English text. The cloze test is a type of test in which each n (the number that is used as the basis for eliminating words in reading) is eliminated, and the reader is asked to fill in the word that is omitted. Basically, a C test is a kind of cloze test. The C test also asks the reader to complete partly written words in a text. In the cloze test, the reader is asked to fill in the word that is omitted, while in the C test the reader is asked to complete a rudimentary word, the letter removal follows the rule 2. That is, every second word and multiples are made incomplete by removing the second half of letters of those words and if the number of letters makes it an odd (not divisible by half) word, the letters removed are the second half plus one letter (Brown, 2004).

In this case, according to Alderson (1983), reading ability can be measured in various ways. One way might measure certain aspects of reading. There is no one best way to measure reading ability, because each method has advantages and disadvantages. Therefore, each method is liked by a number of people and not liked by a number of other people.

An overview of the background of the subject of this study can be seen in table 1 below:

Table 1: Educational background of research subjects

No.	ELEMENT	% N = 20
1	Age	20-22 = 100%
2	Background of Senior High School	Government Senior High School = 90 % Private Senior High School = 10 %
3	English Course	Ever = 60 % Never = 40 %

Based on table 1 above, it can be seen that the population of this study is 20 students who were randomly selected as research subjects, all aged between 20-22 years. Of those 20 people, 18 (90%) students came from government high schools in North Sumatra Province,



while two people (10%) came from private high schools. Based on information from these research subjects, 12 people have studied English at the English Language Institute, while 8 people claim never to have studied English in a course or outside the lecture bench.

This study uses three different text titles as research instruments, which text length amounts to around 500 words. The instrument tested was then given to members of the research subject to work on. Comprehension question (CQ), multiple choice (MC), and cloze passage (CP) tests are given to students. After completing the test, students are asked to provide an assessment of their own ability to take the test they have just completed. The assessment they put in a self-assessment sheet.

The texts used in this study are as follows:

Text 1:

Food in America

Many changes are taking place in “food styles” in the United States. The United States is traditionally famous for its very solid and unchanging diet of meat and potatoes. Now we have many different alternatives to choose from various ethnic food, health food and fast food, in addition to the traditional home cooked meal.

Ethnic restaurants and supermarkets are commonplace in the United States. Because the United States is a country of immigrants, there is an immense variety. Any large American city is filled with restaurants serving international cooking. Many cities even have ethnic sections: Chinatown, Little Italy, or Germantown. With this vast ethnic choice, we can enjoy food from all over the world. This is a pleasant thought for those who come here to travel or to work, they can usually find their native specialties, or miso soup. Besides sections of the cities, there are regions which are well known for certain food because of the people who settled there. For example, southern California has many Mexican restaurants, and Louisiana has a strong Creole accent to its food. (Creole is a mixture of French, African, and Caribbean Island food).

Health food gained popularity when people began to think more seriously about their physical well-being. The very term health food is ironic because it implies that there is also "unhealthy" food. Health food is fresh, natural, unprocessed food. It does not contain preservatives to make it last longer or chemicals to make it taste or look better. Most health food enthusiasts are vegetarians: They eat no meat; they prefer to get their essential proteins from other sources, such as beans, cheese, and eggs.



Fast-food restaurants are now expanding rapidly all over the country. In the United States, speed is a very important factor. People usually have a short lunch break or they just do not want to waste their time eating. Fast-food restaurants are places which take care of hundreds of people in a short time. There is usually very little waiting, and the food is always cheap. Some examples are burger and pizza places.

America's attitude toward food is changing, too. The traditional big breakfast and dinner at 6:00 P.M. are losing popularity. People are rediscovering the social importance of food. Dinner with family or friends is again becoming a very special way of enjoying and sharing. Like so many people in other cultures, many Americans are taking time to relax and enjoy the finer tastes at dinner, even if they still rush through lunch at a hamburger stand.

Source: (Contact U.S.A Reading and Vocabulary Textbooks “Chapter 4. page 54”)

Text 2:

A Country of Immigrants

As you walk along the street in any American city, you see many different faces. You see Oriental faces, black faces, and white faces. These are the faces of the United States, a country of immigrants from all over the world. Immigrants are people who leave one country to live permanently in another country.

The first immigrants came to North America in the 1600s from northern European countries such as England and Holland. These people generally had light skin and light hair. They came to live in North America because they wanted religious freedom. In the 1700s and early 1800s immigrants continued to move from Europe to the United States. At this time there was one group of unwilling immigrants, black Africans. These people were tricked or forced to come to the United States, where they worked on the large farms in the south. The blacks had no freedom; they were slaves. In the 1800s many Chinese and Irish immigrants came to the United States. They came because of economic or political problems in their countries. The most recent immigrants to the United States, the Indochinese, Cubans, and Central Americans also came because of economic or political problems in their own countries. Except for the blacks, most of these immigrants thought of the United States as a land of opportunity, of a chance for freedom and new lives.

In the United States, these immigrants looked for assistance from other immigrants who shared the same background, language and religion. Therefore, there are neighbourhoods in each U.S. city made up almost entirely of one homogeneous ethnic group. There are all Italian, all Puerto Rican, or all Irish neighbourhoods in many East Coast cities and all



Mexican neighbourhoods in the Southwest. In Dearborn, Michigan, there is a large group of Lebanese. There are racial neighbourhoods such as oriental Chinatown in San Francisco and black Harlem in New York. There are also neighbourhoods with a strong religious feeling such as a Jewish part of Brooklyn in New York. And, of course, there are economic neighbourhood divisions; in American cities very often, poor people do not live in the same neighbourhoods as rich people.

This diversity of neighbourhoods in the cities is a reflection of the different groups in American society. American society is a mixture of racial, language, cultural, religious and economic groups. People sometimes call America a melting pot and compare its society to a soup with many different ingredients. The ingredients (different races, cultures, religions and economic groups) supposedly mix together to make a smooth soup. But, in reality, there are a few lumps left in the soup.

Source: (Contact U.S.A Reading and Vocabulary Textbooks “Chapter 2. Page 20”)

Text 3:

Impressions of the United States

United States. What is your first thought when you hear these words? Is it an image of something typically American? Perhaps you think of hamburgers and fast-food restaurants. Or perhaps you have an image of a product, such as an American car or Coca-Cola. Some people immediately think of American universities. Others think of American companies. Many Americans think of the red, white and blue flag when they think of United States. There are many images associated with the name of a country.

There are also many ideas or concepts associated with the words United States. Some people think of a positive concept, such as freedom, when they think of the United States. Other people think of a negative concept, such as American involvement in other countries. Many Americans have both positive and negative ideas about their country. When they think of the lifestyle or the scenery (landscapes such as mountains or beaches at the ocean), they feel very positive and proud of their country. But sometimes, when they think about the government, they think about nuclear war and international problems. Then they have negative feelings about the country.

These images and ideas are all impressions of a country, the United States. People form these impressions in many different ways. They see American products and advertisements. They read newspapers and hear people talk about the United States. They probably see American



movies and television shows. These impressions are always changing. As people receive more information, they adjust their images and concepts of a country.

Knowledge of a country includes many things. Typical products and actions by governments are part of this knowledge. But the most important thing in learning about a country is knowledge of the people of that country. What are their customs and lifestyles? How do they raise their children? What are their values and beliefs? How do they feel about work and entertainment, about time, about friendships?

Although technically more accurate, the term North American is not used by the people in the United States to describe themselves. Therefore, the term American is used throughout this book to describe things in the United States.

Source: (Contact U.S.A Reading and Vocabulary Textbooks “Chapter 1. Page 4”)

The three texts above are each followed by two of the three types of exercises used. Text 1 is attached with comprehension questions (CQ) and multiple choice (MC), text 2 with multiple choice (MC) and cloze passage (CP), text 3 with cloze passage (CP) and comprehension questions (CQ).

The research data were analysed descriptively to determine the scores obtained by students for each test and the self-assessment of each test. The descriptive statistics used are standard and standard intersections. To determine the validity of these benchmarks, the *Product Moment Correlation* technique was used, by correlating the cloze test and C test scores with the Reading Comprehension test scores from the reference book. The scores of each self-assessment are also correlated with the scores of each test with *Product Moment Correlation* techniques.

The Pearson Correlation, often called *Product Moment Correlation*, is a statistical test tool used to test the associative hypothesis (relationship test) of two variables if the data is interval or ratio scale. *Product Moment Correlation* was developed by Karl Pearson (Hasan, 1999).

Findings and Discussions

Findings

There are several things that need to be considered using reading comprehension exercises in the results of this study, namely:

1. Until now the training material has been sourced from sentences and smaller parts of

sentences. This can still be justified. The assumption is that a text is a continuation of different sentences that are thematically related and are considered necessary to pay attention to the relationship of structure and meaning of sentences. However, if reading needs to be efficient, the structure of longer sections such as paragraphs or the whole text must be easier to understand. It is not good to study the text as a number of independent units. Assumptions like this led to:

- a. Students become very dependent on the understanding of every sentence in the text. In fact, such actions are not necessarily in accordance with the purpose of reading. The result is students tend to read different texts with different reading goals in the same reading speed.
 - b. Students feel reluctant to conclude the meaning of a sentence or paragraph from the context (or from the sentence / paragraph before or after it).
2. As a consequence of the first assumption, one must start with a global understanding and move towards a more detailed understanding, not work the other way around. The task given as a start should be of a more global type, still within the scope of easy understanding, so more detailed and deep understanding can be done. Likewise, when developing a reading comprehension exercise for a particular text, it is preferable to begin with the global meaning of the text, the function and purpose of the text, not with a vocabulary or more specific ideas.

The foregoing needs to be stated for the following reasons:

- a. The method proposed is an efficient way to build student confidence when faced with authentic (original) texts that often contain difficult vocabulary or more complex structures. If the initial activity starts globally (for example, choosing from a list of functions that the text suggests), students have no difficulty.
- b. The above method also helps develop awareness about how the text is actually arranged. This is done by explaining the main information and its development or giving a chronological sequence of events. This general awareness of the structure of essays is what makes students read more efficiently in the future.

This study uses three different text materials (quoted from the book Contact U.S.A Reading and Vocabulary Textbooks) with three types of comprehension questions, multiple choice, and cloze passage exercises. Comprehension questions are open-ended questions that ask the subject's understanding of the contents of the text in general. The multiple choice is a form of multiple choice that asks the subject's understanding of the contents of the text. Cloze Passage is a form of exercise that tries the ability of the subject to understand the context of

the text used as reading material. Each type of exercise is then abbreviated as CQ, MC, and CP.

Each text used is attached to two different types of exercises. Accordingly, text 1 is attached with CQ and MC, text 2 is attached with MC and CP, and text 3 is attached with CP and CQ. Each group of subjects gets a different set of exercises. A list of exercises for each group can be seen in table 2 below:

Table 2: List of exercises for each group

Groups	Text	Exercises
A	1	CQ and MC
	2	MC and CP
	3	CP and CQ
B	1	CP and CQ
	2	CQ and MC
	3	MC and CP

From the exercises given to the research subjects, quantitative data were found as seen in table 3. This table shows there are several variations between research subjects in the different forms of exercise. Analysis of variations is given in the following sections:

Table 3: Analysis of variations in research subjects

Subject A	CQ/MC	MC/CP	CP/CQ
1	66, 66/70	80/70	70/70
2	66, 66/70	70/70	80/70
3	83, 33/80	80/70	70/80
4	66, 66/80	80/70	80/70
5	83, 33/80	80/80	80/80
6	83, 33/80	80/80	80/70
7	83, 33/80	70/80	70/80
8	83, 33/90	80/80	80/80
9	66, 66/70	70/70	70/70
10	66, 66/70	70/70	70/70
Subject B	CP/CQ	CQ/MC	MC/CP
1	80/70	66, 66/80	80/70
2	70/70	66, 66/80	80/70
3	70/80	83, 33/70	70/70
4	80/70	66, 66/70	70/80
5	80/80	83, 33/80	80/80
6	80/70	83, 33/80	80/80
7	80/80	83, 33/80	80/70
8	70/80	83, 33/80	80/80
9	70/70	66, 66/80	70/70
10	70/70	66, 66/70	70/70

Discussion

In accordance with the number of pairs available, there are three comparisons of average values in the analysis of the results of the training given to the subjects. The quantitative analysis used in this comparison use the cloze test and C test. Statistical formulas are used to see whether the difference between the two mean values of the two different training groups is significant.

The null hypothesis is used in the application of the cloze test and C test. Therefore, the hypothesis proposed in this analysis is that there is no difference between the use of the CQ and MC, MC and CP test forms, as well as CP and CQ.

The significance level chosen was 5% (0.5) to reject the above hypothesis. Based on the cloze test and C test distribution table, the value of r is 2.13. Thus, if the value of r obtained is greater than the value of the distribution of r then the hypothesis is rejected. In other words,

there is a difference given. Conversely, if the r value obtained is smaller than the hypothesis can be accepted.

The results of the analysis can be seen in table 4. In this table it can be seen that the r value obtained for each comparison is CQ-MC = 1.23 MC-CP = 1.45 and CP-CQ = 0.63 because for each value it is below the value of r 2.13. Therefore, the above hypothesis can be accepted. There is no difference between the use of the CQ-MC, MC-CP, CP-CQ test forms in reading.

Table 4: Critical ratio correlation between reading ability tests and correlation of self-assessment with all three reading ability tests

No.	Types of Exercises	Sd	M	t	df
1	CQ-MC	7.29	2.00	1.23	19
2	MC-CP	6.16	2.00	1.45	19
3	CP-CQ	7.18	1.00	0.63	19

The rejection of the above hypothesis implies that there are no differences in the results of the exercise shown by the three types of exercises used. The results of the cloze test and C test above are unable to show whether each exercise used has the same level of difficulty. This information is needed especially in preparing appropriate strategies for the final evaluation in reading learning.

On the other hand, if the value of the exercise results can indicate the level of difficulty of the exercise, the analysis provided in this study can help provide the information needed. In the same type of exercise, this relationship is acceptable. In other words, two types of CQ are given to the subject. If between the two CQ there are different training results, one of them can be said to have a higher level of difficulty.

The results of the *Product Moment Correlation* analysis for the overall data shows that all correlation coefficients were negative. These coefficients indicate that the cloze test and the C test have benchmark validity and can be used as an alternative measure of reading ability at any level. Self-assessment can also be done at any level. The important thing to note is that there is clarity of instructions in carrying out the self-assessment.

When compared between group tests, the correlation coefficient was no different between the two groups. The same correlation is found in group A with a coefficient of 2.00, and group B with a coefficient of 2.00. This situation illustrates what is stated by Cohen (1994: 199), that self-assessment is sometimes subjective and not also because it is influenced by various factors such as past achievements, desires, expectations, comparisons with others, and so on.

In other words, the higher the knowledge of a student does not necessarily correlate with the student's ability to judge his or herself.

To find out whether the correlation coefficient of the test cloze and the test C with the Reading Comprehension test from the reference book as a benchmark and the correlation coefficient of self-assessment with these three tests statistically have only a slight difference based on knowledge, the critical ratio of the correlation coefficients can be used (Furlong, Lovelace, & Lovelace, 2000). Critical ratios can be found by changing the correlation coefficient to the value of z . The results obtained are compared with the critical value of z , which is ± 1.96 . If the critical ratio is greater or equal to the critical value of z , the correlation coefficients are statistically different. Table 4 above shows the critical ratio of correlations between reading ability tests and correlation of self-assessment with the three tests.

Conclusions

From the comparison of the average value of reading exercises provided, it can be seen that there is no significant difference between the three pairs of evaluation tools. The three evaluation tools are in text 1 comprehension questions (CQ) and multiple choice (MC), text 2 with multiple choice (MC) and cloze passage (CP), and text 3 with cloze passage (CP) and comprehension questions (CQ). Thus, the three forms of exercise can be used to add variety to reading exercises.

Although there is no significant difference between the three forms of exercise, this does not mean that all three have the same advantages. It must be acknowledged that the research conducted has not been able to provide answers to questions as to which form of evaluation tool is better used in reading and which can show the highest results. This is because each exercise has an unequal emphasis. Accordingly these three forms of exercise should be given as variations in reading lessons (reading), because to do so can reduce the level of boredom of students, and the use of a variety can provide input that is not always uniform to the teacher. Teachers can find out which forms of practice still provide obstacles to students in achieving teaching indicators.

Understanding the text means filtering the information needed from the text efficiently. For example, we apply different reading strategies when looking at the bulletin boards to see if there are job advertisements to when reading articles in scientific journals. From these two examples, the purpose of reading in each example has been fulfilled. In the first example, a reader filters out the form of advertisement that is relevant to his or her situation. In the second example, a reader reading a scientific article not only understands the main points of the text but also a more detailed understanding is needed.



Acknowledgement

The authors would like to express their appreciation and gratitude to the English Department *Universitas Sumatera Utara* for funding the research and publication of this article.



REFERENCES

- Abraham P. and Mackey D. (1989). *Contact U.S.A Reading and Vocabulary Textbook*. London: Prentice Hall International UK Limited.
- Alderson, J. C. (1983). *The cloze procedure and proficiency in English as a foreign language*. In Oller, J. W., Jr. (Ed.) *Issues in language testing research*. Rowley: Newbury House Publishers, Inc.
- Broughton, G. et al. (1978). *Teaching English as a Foreign Language*. MA: Routledge Education Books.
- Brown, J. D. (2004). *Language Assessment*. New York: Addison Wesley Longman, Inc.
- Brown, J., D. (2013). My twenty-five years of cloze testing research: So what?. *International Journal of Language Studies*, Vol. 7, No. 1, pp. 1-32.
- Cohen, A. D. (1994). *Assessing language ability in the classroom (2nd ed.)*. Boston: Heinle & Heinle Publishers.
- Creswell, John W., & Vicki L. P. C. (2007). *Designing and Conducting Mixed Methods Research*. Thousand Oaks: SAGE Publications.
- Farr, R. & Carey, R. F. (1986). *Reading: what can be measured? (2nd ed.)*. Newark: International Reading Association, Inc.
- Furlong, N. E., Lovelace, E. A., & Lovelace, K. L. (2000). *Research methods and statistics*. Fort Worth: Harcourt College Publishers.
- Hasan, Iqbal. (1999). *Pokok-Pokok Materi Statistika 2 (Statistik Inferensif)*. Jakarta: Bumi Aksara.
- Kleijn, S., Maat, H. Pander., & Sanders, T. (2019). Cloze testing for comprehension assessment: The HyTeC-cloze. *International Journal of Language Testing*, Vol. 36, No. 4, pp. 553–572.
- Maroko, G. M. (2016). Uncovering Cloze Testing Practices in Selected Practice Papers for Primary Schools in Kenya. *Journal of Applied Linguistics and Language Research*, Vol. 3, No. 4, pp. 121-134.
- Nurhadi. (2016). *Teknik Membaca*. Jakarta: Bumi Aksara.



- Saeedi, M. (2016). Construct Validity of Multiple-Choice Cloze Test and Cloze-Elide Test in Testing Reading Comprehension among Iranian EFL Learners. *International Research Journal of Applied and Basic Sciences*, Vol. 10, No. 6, pp. 632-637.
- Saeedi, M., Tavakoli M., Kazerooni. S. R., & Parvaresh, V. (2011). Do C-Test and Cloze Procedure Measure what they Purport to be Measuring? A Case of Criterion-Related Validity. *International Journal of Cognitive and Language Sciences*, Vol. 5, No. 2, pp. 190-199.
- Septiarini, T., & Hikmah, S. (2017). Multiple-Choice and Cloze Tests In EFL Reading Comprehension: What Do They Effectively Measure?. *International Journal of Multidisciplinary and Current Research*, Vol. 5, No. 6, pp. 680-684.
- Suharso. (2008). Validitas Tes Cloze, Tes-C, dan Penilaian Diri Sebagai Alat Ukur Kemampuan Membaca Teks Bahasa Inggris. *Jurnal Penelitian dan Evaluasi Pendidikan*, Vol. 1, No. 11, pp. 101-117.
- Tarigan, H. G. (2015). *Berbicara Sebagai Suatu Keterampilan Berbahasa*. Bandung: Angkasa.