

# A Comparative Study Of Efl University Students' Written Performance On Cloze And Standardized Test

<sup>1</sup> Maha Ahmed Yaseen, <sup>2</sup> Wassan.A. Rasheed

<sup>1</sup> *University of Mosul, College of Basic Education, Dept.of English, Iraq, amaha8756@gmail.com*

<sup>2</sup> *University of Mosul, College of Basic Education, Dept.of English, Iraq, wassan.adnan@uomosul.edu.iq*

## Abstract

This study aims at investigating EFL university students written performance in cloze tests compared to their performance in standardized tests in relation to the grammatical mistakes, vocabulary, and discursal mistakes of the texts they write. Exploring whether or not the cloze test can be successfully used to measure the grammar, discourse aspects and vocabulary of EFL writing texts. The researcher used a causal-comparative design. There were 275 third graders at Mosul University who made up the study's population. A random sampling method was used to collect samples for this study. During the academic year 2021–2022, a total of 110 male and female 3rd grade university students from four distinct classes made up the sample. Two groups made up of 55 students each are used for each test to divide the sample into. In this study, a test was the instrument used. Specifically, cloze tests and standardized tests were the two test types offered. The results of the current study demonstrated a significant difference between the students' written performance on the cloze test and their written performance on the standardized test, which was used to measure written performance among third-grade students at Mosul University.

**Keywords:** Cloze test, Standardized test, Written performance, Vocabulary, Grammar.

## Preliminary

One of the most difficult skills for nearly all learners is writing in a foreign language. Compared to speaking, listening, and reading, the development of writing skills is assumed to be very complex, perhaps the most complex. Writing requires certain criteria of acceptability in relation to various aspects of writing, including language use, content, organization, vocabulary, spelling, punctuation, and precise capitalization and paragraphing. This makes writing a very challenging task for EFL learners (Hamadouche, 2010). Writing is becoming more important to students' university because this one is the primary standard they utilize to complete tasks in various topic parts and respond to exam requests. Actually, students who struggle with their writing are more likely to struggle academically because they lack the tools to effectively communicate what they know and find it challenging to complete the required essays (Elachachi, 2006 and 2010). Since writing skills are transferable, learning how to write effectively will help learners maintain accurate documentation and clear communication throughout their careers.

According to Celce- Murcia (2001) states that in order to attempt to deliver an excellent written text, one must simultaneously be able to control a variety of linguistic systems and have the mental capacity to consider how the discourse must always be formed for a special purpose (Celce Murcia 2001). Contrarily, a lot of theoretical and methodological research has been done on the cloze test. One of the most widely used testing methods today is the Cloze, particularly for determining the overall language ability of EFL/ESL students. Heaten (1990) claimed that cloze test performance is highly correlated with reading, writing, and speaking skills (p.17). According to Hughes (2003), "When taking cloze tests, students use the skills that underlie all of their language performance. In this study, cloze passage is considered as a prewriting exercise. In an effort to measure students' proficiency with some other language skills such as writing and speaking, Carter and Nunan (2001) found a strong correlation between students' achievement on integrative tests such as the cloze test and the sub-components of numerous other test batteries (p.139).

The best method for teaching language abilities is an integrated one that emphasizes language abilities holistically because writing is a skill that needs a variety of components in combination. One integrative test that evaluates multiple sub-skills simultaneously is the Cloze test. These problems sparked a creative idea: developing creative writing activities and teaching strategies. Even though some worthwhile initiatives have been taken to help learners become better writers, the results have not been satisfactory. The need to adapt writing instruction through engaging methods seems to be pressing. Application of the cloze test as a supplemental exercise in the classroom is one of the so-called activities. Utilizing the cloze test to supplement in-class practice is only one of the many such activities. Cloze tests are different from other tests in that they have piqued learners' interests more than others.

### **Theoretical Background of the Study**

Psychology had altered the manner in which students were tested by the twentieth century. Psychologists altered the testing system by providing standardized tests (Giordano, 2005). Testing has developed into a significant science or field of study for educators and psychologists, and it is frequently used not only because it is simple to do so but also because it is effective in determining students' academic standing and depth of understanding of the material being taught, as well as their potential to advance to higher educational levels. The cloze test is one of many additional test types that are available.

Because it involves taking out roughly every fifth word from a passage or sentence and afterwards asking the student to replace them, a cloze test is an essential tool for assessing test-takers' comprehension. The fill in the gap exercise is its common name as a result (Keisuke, Yuki, and Mamoru, 2013). Since the 1950s, such a teaching method has been in use. The Gestalt School of Psychology's concept of The theoretical background for this kind of test was established by "closure," which asserts that the brain perceives things as both a complete unit and one that would normally as well as merely restore missing elements (Walter, 1974). To put it another way, when details or words are omitted, a participant will fill in the gaps with the help of their experiences, prior knowledge, and abilities like critical thinking and reasoning (Melody Ann Ross, 2017).

According to the Gestalt school's theory of closure, a learner will naturally fill in the gaps based on their previous experiences and prior knowledge because the brain perceives language as a complete unit during the learning process. The emphasis of the cloze procedure is really to stratify the omission to a full paragraph of naturally, properly structured text. In the cloze tests, there are 2 techniques used to have students fill in the blanks with phrases and words based on their own interpretation. Giving students a curated list of items and phrases and asking them to choose the proper ones and correctly insert them in texts or sentences is another way to use the deletion system (Tobias and Torsten, 2014).

When there is cloze exercise in the deletion area, there are two different methods to take. The evaluation of cloze activities is impacted by both the systematic deletion method and the unsystematic deletion method (Bachman, 1985). Language tests, according to McNamara (2000), are crucial for assessing students' abilities. Additionally, he asserted that the cloze procedure, would be the better test to evaluate students' abilities in language skills. How to score cloze tests is one of the most active areas of study in this field. Two of the five scoring methods tested by Oller (1972) using a large number of subjects were the exact word scoring method and the acceptable word scoring method. In contrast to the latter, which allows substitute words to be used in place of the deleted words, the former restrains the utilization of any word other than that which was removed from the original text. The acceptable-word scoring method accepted responses with similar meanings as long as they were grammatically as well as contextually correct, regardless of how they varied from the original word in a part of speech.

There are different kinds of cloze tests, such as the fixed-rate deletion cloze, which tends to focus on removing each seventh word, and the multiple-choice cloze, in which learners are given three to four options and asked to circle the answer that is significantly more accurate to fill in the blanks. The C-test, in which the second half of each item is omitted, and the Cloze elide test, where the gaps are already filled in with elements that are incorrect and inappropriate for the text, when learners are needed to fill those with the correct items. Selective deletion, also known as the rational cloze method, requires

students to select text items that best support the test's objectives (Hadley, 2004).

### **The definitions of Testing**

Testing is an evaluation metric used by teachers to collect data from students. The term "test" is derived from the Latin word "testum," which refers to a device used to gauge or assess activity that takes place in a specific location. A test is a collection of questions that must be satisfactorily answered. If testing is included in daily classroom instruction, is designed to be an element of the learning process, and is integrated into daily teaching, testing can be a significant component of teaching and learning. It enables teachers to daily revise their lesson plans while also enabling students to assess their own progress. A test is a tool for evaluating students' progress in a more organized manner using a numerical scale or specific category systems. According to Tinambunan (1988, p.3), a test is helped facilitate as a tool for measuring how well students are doing. Before creating a test, the teachers need to be aware of the tests that are appropriate for their class.

### **Types of Tests**

Achievement tests, diagnostic tests, placement tests, proficiency tests, and aptitude tests are the five categories of tests that have been identified in the literature (Brown, 2019, p.10, Heaton, 1991, p.171).

#### **1- Achievement Tests**

The most frequent reason for a teacher to administer a test in the classroom is to gauge the level of understanding of a unit, lesson, or even the entire curriculum among the students. They are typically referred to as "achievement tests". They are (or should be) restricted to particular curriculum materials covered in a specific amount of time and are administered after a course has focused on the aforementioned objectives. Achievement tests' main purpose is to confirm if either course goals have been achieved or the necessary skills and knowledge have been obtained by the end of a specific period of instruction, though they can also play a diagnostic function to show what the learner needs to work on moving forward.

#### **2- Diagnostic Tests**

A diagnostic test is a particular kind of test designed to identify students' areas of weakness so that later on, the proper corrective action can be taken. The goal of a diagnostic test is to pinpoint language skills that the student must improve or that the teacher should cover in the course. For instance, a pronunciation test would perhaps identify the phonological aspects of English that are challenging for students and, as a direct consequence, should be covered in a curriculum. Such tests typically offer a checklist of characteristics the manager (often the teacher) can use to spot problems.

#### **3-Placement Tests**

Some aptitude and achievement tests can also be represented as placement tests, and their goal is to place the student in a specific level or area of the school's or curriculum's language program. The placement test could well typically, but not always, contain a sample of the topics that are going to be covered in the various courses of the education system; students' performance on the test could also refer to the point at which they will discover the material to be neither very simple nor too complicated, but at an appropriate level of difficulty.

#### **4- Proficiency Tests**

Testing students' language skills in relation to a specific task they would be asked to complete is known as a proficiency test. The purpose of proficiency tests is to measure particular skills in light of education requirements that will later be presented to the student through a course or job, rather than to assess general attainment in the subject area. Proficiency tests are unrelated to every curriculum item or instructional program. Common components of proficiency tests include multiple-choice questions with standardized answers for vocabulary and grammar, auditory comprehension, and reading comprehension.

#### **5- Aptitude Tests**

An aptitude test is a particular kind of exam created to gauge a student's potential proficiency in a foreign language in which he or she has not yet started learning; it evaluates aptitude for language acquisition. By assessing how well students perform in a made-up language, aptitude tests aim to identify their potential

strengths and flaws in teaching a foreign language. The aptitude test is designed to assess a person's potential or all-around capacity to master a foreign language in order to move up (before enrolling in the course) and their likelihood of passing this certification in the end. To learn any language, language aptitude tests were supposedly intended to be used in the classroom.

### Types of Scoring Tests

There are two different kinds of scoring tests. They are both subjective and objective tests.

#### 1- Subjective Tests

An essay is the typical format for a subjective text. It necessitates a detailed response and a thorough justification. According to Arikunto (2005, p. 164), a subjective test has the following characteristics: Fewer questions, preceded by "why," "explain," "compare," "how," and "conclude," are asked. In order to express their thoughts as fully as possible and to the best of their abilities, test-takers are required to use creativity.

Subjective tests have the advantages of being simple to administer, being able to prevent the potentiality of cheating, being able to assist the teacher in determining the students' comprehension, and being able to train the test taker to respond based on their own judgment. On the other hand, this test's shortcomings include the time it spent asking questions, which means that it did not cover the entirety of the material. The test-writing taker's abilities are improved by subjective test items. The subjective test item is a great way to gauge language proficiency if that is the test's goal. The objective of any test is to measure the performance of the test taker, but the administrator of the subjective test typically emphasizes using knowledge rather than memorizing it (Aldridge 2014).

#### 2-Objective Tests

There are typically more questions on the objective test than on the subjective one, and it is frequently multiple-choice formatted. In his or her opinion, only one item should be selected by the test taker as the best response to the question. Examiners typically spend relatively little time establishing the question but too much time making Angesti (2011) referenced objective

testing, which is reported to necessitate more precise preparation than objective testing (1975, p.26). Each test item is carefully and accurately created by the test-taker, who invests a lot of time in it. The ability of students to review and apply principles is another area in which an objective test is ideal for assessing their abilities.

### Cloze Test

The Cloze procedure is frequently described as a passage or text of the proper length and level of difficulty with every  $n$ th word removed. The meaning of this is pretty obvious. However, some safety measures must be taken concerning the ideas of suitable length, difficulty, value, and frequency of  $n$ . Here's a quick explanation of each of them (Hossein Farhady, 1996).

The number  $n$  represents how far apart two deletions are from one another. That is, the distance between two omissions is calculated by giving  $n$  a value. The frequency of  $n$  denotes how many deletions there were in the text. We go into great detail about the distance between omissions as well as the quantity of deletions because they have an impact on the test's length and difficulty.

The distance between omissions will be 1 if  $n$  is determined to be one and a paragraph is clarified. This necessitates the test's creator starting to eliminate words one at a time. In such a scenario, all text would be eliminated, leaving nothing behind. Consequently,  $n$  cannot be regarded as one. In other words,  $n$  must always be at least 2. In other words, one must start counting from the very first word and eliminate each subsequent word.

Another aspect of  $n$  to consider would be its frequency. That also represents the number of omissions. In general, because each deletion is regarded as an item, the more deletions there are, the more items there are, and as a result, the test will be longer. A great amount of omissions, on the other hand, would then lead to an absurdly long test. The text's length, and then the omissions number, must be earmarked within a reasonable domain. Some researchers believe that there should be no more than 50 blanks in a deletion, which is the maximum number that they would accept. For some, 20 to 25 blanks are sufficient (Rand, 1978; Cohen, 1980). But the ideal number of omissions in a cloze procedure is typically thought to be between 30 and 35. The

passage length can be established once the distance between omissions as well as the number of omissions have been decided. The passage would total approximately 210 words long, for instance, if there were 30 omissions and every seventh word was eliminated.

However, another consideration must be made in terms of passage length. According to research on reading comprehension, text structure is an effective element in helping to process the material to be read. The largest portion of the data is spread across two distinct sentences in the paragraph. The topic sentence for the first of those certain sentences contains the paragraph's main idea. The paragraph is summarized and ended in the subsequent sentence. According to some academics, these two sentences shouldn't be deleted because doing so might alter how the language reader naturally processes the sentences.

A precise definition of such a cloze test could be given as follows when the deletion number and the distance between omissions are known, and consequently of the paragraph length: The cloze test consists of a passage of about 240 words that has every seventh word removed, provided the initial and final sentences are left alone.

### **Standardized Tests**

The test must have certain characteristics to be a true standardized test. Standardized tests are advanced in accordance with agreed standards for test improvement. Each gets tested on the similar arrange of questions. They are tested under the same situations. They get the same instructions and they have the equal quantity of time to finish the test. The methods' test is also the same. Standardized tests may be offered in a group or face-to-face setting.

Most standardized tests in education are offered in group settings. However, some tests are given face-to-face with a single test taker and an administrator (Holly Dolezalek, 2009). A good standardized test is the produce of a detailed empirical research and improvement process that may expand behind purely establishing criterion or standards. In addition, standardization mean the use of methodical steps for management and scoring. Furthermore, several standardized tests, particularly extensive tests, are norm-referenced, the purpose of which

is to place test-takers in a continuum across area of marks and to distinguish between test-takers by their proportional classifications.

### **Standardized Testing Frameworks**

In addition to taking into calculation the design for apprehensions about reliability, validity, and rightness, test publishers design a standardized test according to how to report and use their results. The numeral of questions responded correctly on the test, the learner's row mark, is meaningful only in the situation of the test's explicatory frame. here are various types of interpretive frameworks.

#### **1- Norm-referenced Testing (NRT)**

The standardized test prepared in the NRT explanatory frame can be used to parallel the results of test takers with the results of a reference group that took the same test. To norm the test so that outcomes can be paralleled, the test producer collects normative data from the test area judgments with agent general sample of learners. To compare big groups such as whole school systems, standard reference tests are usually intended to overlay a wide domain of what test takers are predictable to know and be capable to do in a subject part.

#### **2- Criterion-referenced Testing (CRT)**

Instead of comparing a learner's test outcomes to those of a reference group, criterion-referenced tests aim to measure mastery standard against a particular series of performance criteria. Therefore, the purport of a criterion-referenced test often contains a more intensive topic matter than a norm-referenced test. The students mark matches to a standard of performance, like elementary, skill, or progressive. CRT requires the plan or determination of a valuation that produces outcomes that can be used to class learners into performance standards for the conformable educational topic.

#### **3- Standards-based Testing**

By integrating the components, including both norm-referenced testing and criterion-referenced testing, standards-based testing enables both objectives (NRT and CRT) to be achieved simultaneously. Standards-based testing is both aligned with or consistent with such a set of performance standards and normed

to a reference group. This framework, also known as the Enhanced NRT model, enables the reporting of performance levels (cut off scores), standards-based information (content criteria scores), and percentile rating data for each student. An existing norm-referenced test can be expanded by a test publisher, for instance, using the state's academic standards, allowing test taker results to be compared to a reference group and performance levels to be assigned. The test reference group is typically created using the statewide outcomes in the first year a standards-based test is given. The test's validity to assess students' academic standards mastery is guaranteed by the test publisher's precise design.

Standards-based tests that are properly designed are especially well-suited to satisfy NCLB requirements because the NCLB mandates that educators' levels of performance be reported while comparing the outcomes of particular groups on campus with those of previous years (Holly Dolezalek, 2009; Sasha Zucker, 2003).

## Methodology

### 3.1 Population and Sample

#### 3.1.1 The population of the research

According to Lodico (2006, p. 13), "the population is the "substantial number with which the researcher would prefer the findings of the research to be generalizable". This study's population consists of all third-year English major college students studying in the mornings at the Department of English/College of Basic Education/University of Mosul during the second semester of the academic year 2021–2022. There are 275 students in total, A, B, C, D, and D comprise the five sections. To validate the study's objectives, they were specifically chosen.

#### The Sample of the Research

According to LoBiondo-Wood and Haber (2018), "Sampling is the method by which a subsection of a population is just a representative of that population." To decrease the possibility of bias, the sampling method used in the current study was randomized. "It represents students who are chosen at random from the general population to make up the study sample and who are given a comparison treatment "(Cumplings, Grady, & Hulley, 2013, p. 137).

#### Instrument of the Research

The majority of research instruments are tools for gathering data that is pertinent to the study (Wilkinson and Birmingham, 2003, p. 3). Two different types of testing methods for written performance serve as the research instruments. To evaluate the students' written performance, cloze and standardized tests were used.

#### The Cloze Test

The cloze test includes two different argumentative essays which were taken from the students' textbook entitled. In the first essay, the deletions covered the content and function words. In addition, each type of deletions consists of 40 deletions. The second essay covered the deletions (linking words). 40 deletions are also included.

Three different cloze tests were used to alter the text, including:

1. the omission of every fifth ( $\pm 4$ ) function word.
2. the omission of every fifth ( $\pm 4$ ) content word.
3. the omission of every fifth ( $\pm 4$ ) linking word.

Prepositions, articles, modal verbs, auxiliary verbs, pronouns, and conjunctions were listed as the researcher's definition of function words.

The researcher identified the following words as the content words: nouns, verbs, adjectives, and adverbs.

#### Standardized Test

The learners' written exam, as usual, included a writing test, which was given to them. The test was administered to the students in their regular classroom and had a 50-minute time limit. An essay writing test was used to administer the test. The assignment required students to write an entire essay. An instructor who did not have lectures with the test-takers proctored the exam. The invigilator collected the papers and turned them over to the researcher just after the allotted 50 minutes had passed.

## Results and Discussion

### Discussion

The main results of this study exposed that students who took the cloze test performed significantly better than those who took the standardized test. These results suggest that the cloze test format is superior to standardized

testing for assessing grammar, vocabulary, and discursive aspects in essay writing.

Regarding the grammatical aspect, compared to their performance on content words, students' performance on function words had a higher mean score, indicating that there was a number of errors associated with choosing content words. On the other hand, students' written performance in standardized test in related to the grammar aspect was very poor. This means that there were a great number of grammatical errors made by the students in composing written texts.

The most prevalent grammatical mistake was verb usage, which was frequently discovered in the essays of the students. The students' average ability to use verb tenses when writing their essays may be the cause. A verbal error occurs when the word forms used to indicate an action are incorrect. The mistakes are in the form of irregular or regular verbs as well as subject-verb agreement. Additionally, some students used the incorrect verb tenses and couldn't construct simple past sentences. This indicates that the students struggled to correctly apply the grammar rules.

Regarding vocabulary, the results showed that students' written performance in cloze tests is superior to that in standardized tests in relation to vocabulary. This indicates that the students preferred taking cloze tests over standardized tests because they were more familiar with them. Word lists, random words etc. form part of studying vocabulary. In reality, students have weaknesses in vocabulary. It is believed that lack of vocabulary, at least to some extent, may be involved in their writing deficit. It is also possible that this could be the other way around – that writing performance impairments may lead to lack of vocabulary. In fact, how it can be seen is that the context is the key. The word lists or isolated words do not make so much meaning.

Regarding the discursive components, it was discovered that students performed better on a multiple-choice cloze than on a standardized test. The use of cohesive devices, particularly linking words, was the focus of the test. It was found that students' performance was better in other multiple choice cloze tests (i.e. content and function words). It can be observed that students were not aware of linking words. It is evident that students were not capable of differentiating

between the distinctive functions of linking words. This in turn made their scores less compared to the other deletion types of cloze test.

According to Bachman (1982:61–70), numerous scholars and researchers have discovered that specific cloze test types, particularly the selective deletion cloze tests, are useful for testing grammatical knowledge, including cohesion, syntax, and strategic textual comprehension. In addition, Selective deletion cloze, according to Bowen (1985: 376), is the ideal for testing vocabulary and grammar. According to McNamara (2000), it is essential to assess students' language proficiency using tests. He also argued that such cloze test provides the most effective way to gauge students' language skills. According to Aitken, Cloze tests are valid and reliable for evaluating second languages (1977).

## Reference

- [1] Aitken, K. G. (1977). Using cloze procedure as an overall language proficiency test. *Tesol Quarterly*, 11(1), 59-67.
- [2] Aldridge (2014). Adult 12 learners need a reading comprehension test. *International Journal of Linguistics*, 7(5), 1948-5425. doi:10.5296/ijl.v7i5.8447.
- [3] Angesti Palupiningsih. (2011). *Testing reading comprehension using cloze tests and cloze-elide tests among the year-10th students of sman 1 depok in the academic year of 2010/2011*. (Bachelor degree dissertation). Yogyakarta university, Yogyakarta.
- [4] Arikunto, Suharsimi (2005). *Manajemen Penelitian*. Jakarta: Rineka Cipta.
- [5] Bachman, L. (1982). The trait structure of cloze test scores. *TESOL Quarterly*, 16, 61-70.
- [6] Bachman, L. F. (1985). Performance on cloze tests with fixed-ratio and rational deletions. *TESOL Quarterly*, 16, 61-70.
- [7] -Bowen, JD., Madsen H, and Hilferty, A. (1985). *TESOL: Techniques and Procedures*. Rowley, MA: Newbury House Publishers.
- [8] Brown (2019). *Language assessment principles and classroom practices*.
- [9] Carter, R. & Nunan, D. (2001). *The Cambridge guide to teaching English to speakers of other languages*. Cambridge University Press.
- [10] Murcia, M. (2001). *Teaching English as a second or foreign language*. (3rd Ed).
- [11] Cohen, A. (1980). *Testing language ability in classroom*. Massachusetts: Newbury, House Publishers, Inc.

- [12] Cummings, S. R.; Grady, D. & Hulley, S. B. (2013). " Designing a Randomized Blinded Trial ". In S, Hulley, S., Cummings, W., Browner, D.
- [13] Wilkinson, D & Birmingham, P (2003). Using Research Instruments. A Guide For Researchers.
- [14] Dolezalek, H. (2009). **Standardized Testing In Schools**. United States.
- [15] Giordano, G. (2005). How testing came to dominate America Schools. New York. NY: Peter Lang.
- [16] Haladyna, T. M. (2004). *Developing and Validating Multiple-Choice Test Items*. (3<sup>rd</sup> ed.). Lawrence Erlbaum Association Publishers.
- [17] Hamadouche, M. (2010). Developing the Writing Skill through Increasing Learners ' Awareness of the Writing Process. M.A. thesis. University of Constantine.
- [18] Elachachi, H. (2006). An exploration into the strategies used for essay writing across three languages: The case of EFL university students. Unpublished doctoral thesis, University of Tlemcen, Algeria.
- [19] Elachachi, H. (2010). Development of a writing curriculum for academic purposes at tertiary level: The case of Algerian EFL university students. *ESP World*, 1 (27), 1-9.
- [20] Heaton, J. B. (1990). *Writing English Language Tests* (New Edition, J. Harmer & R. Kingsbury, Consultant Eds.). London: Longman.
- [21] Heaton, J. B. 1991. *Writing English Language Tests*. New York: Longman Group Limited.
- [22] Farhady, H. (1996). Varieties of Cloze Procedure in EFL Education. Iran University of science and Technology. Available at: <https://www.researchgate.net/publication/254003849>.
- [23] Hughes, A. (2003) *Testing for Language Teachers* Cambridge: Cambridge University Press.
- [24] Keisuke S, A Mamoru, K. Y. (2013). Discriminative Approach to Fill - in - the - Blank Quiz Generation for Language Learners. Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, pages 238-242, Sofia, Bulgaria, August 4-9 2013. c 2013 Association for Computational Linguistics
- [25] Wood, L. G., & Haber, J. (2018) *Nursing Research: Methods and Critical Appraisal for Evidence - based Practice* (9th Edition). Missouri: Elsevier.
- [26] Lodico, M.; Dean, T. & Katherine, H. (2006). *Methods in Education Research*. San Francisco: Jossey-Bass.
- [27] McNamara, T. F. (2000). **Language Testing**. Oxford: Oxford University Press.
- [28] Ross, A. M. (2017). What is a Cloze Test? Cloze Deletion Tests and Language Learning. <https://www.clozemaster.com/blog/cloze-test/>.
- [29] Oller, (1972). Scoring methods and difficulty levels for cloze tests of proficiency in English as a second language. *Modern Language Journal*, 56,151-157.
- [30] Rand, E. (1978). The effects of test length and scoring method on the precision of the cloze test scores. *Workpapers in TESL*. UCLA.
- [31] Tinambunan. 1988. Evaluation of Student Achievement. Jakarta: Departemen Pendidikan dan Kebudayaan.
- [32] Tobias H. and Torsten Z. (2014) . Towards automatic scoring of cloze items by selecting low - ambiguity contexts. Proceedings of the third workshop on NLP for computer - assisted language learning. NEALT Proceedings Series 22 / Linköping Electronic Conference Proceedings 107: 33-42.3.
- [33] Walter, R. B (1974). Historical overview of the cloze procedure. New Jersey: Kean College. ERIC Document Reproduction-Service No. ED 094 337).