

Feature Article

Three Dynamic Methods of Assessing the Reading Comprehension of ESL/EFL Learners

Aaron David Mermelstein National University of Kaohsiung

Abstract

Reading is a complicated cognitive procedure of deciphering symbols to create meaning. It's a complicated relationship that the reader has with the text and is lead by a reader's previous knowledge, experiences, attitude, and culture. Second language (L2) reading comprehension is perhaps a more complicated procedure and there's a need to ensure that L2 reading comprehension is being properly and accurately assessed. Currently, there are several different authentic assessment methods in practice (i.e. Author, 2015), but most others are similar in nature and used for similar purposed. This article offers a brief introduction and an objective discussion regarding the strengths and weaknesses of three of the most common methods to assessing reading comprehension: the multiple-choice test, the written recall method, and cloze reading tests.

Key words Reading comprehension, multiple-choice test, written recall protocol, cloze test

Introduction

In the field of education, reading is one of the most frequently assessed skills (Collins, & Lindström, 2021; Sweet & Snow, 2013; Wixson & Peters, 1987), with the primary goal of reading being comprehension (Collins, & Lindström, 2021; Farr & Carey, 1986; Sweet & Snow, 2013). Johnstron (1983) suggests that there are two views of comprehension: a process or a product, and describes it as a *process* where comprehension is seen as the change in, or state of, knowledge which has occurred. As a *product*, it is viewed as the process by which the change, or state, comes about.

Most modern assessments of reading comprehension assume that comprehension comes from the outcome of the reader's interrelationship with a text (Meyers, 1991; Perkins & Parish, 1988; Seda, 1989; Sternberg, 1991). This correlates with both the top-down and interactive methods to reading, where the reader, rather than the text, is at the center of the reading activity (Nunan, 1991). A product of reading comprehension can be measured by asking readers to answer questions about the text (Johnston, 1983). However, examining a complicated and rigorous procedures involving second language (L2) reading comprehension is not always an easy task. Bernhardt (1991) states "if a test is to adequately assess L2 reading ability it must acknowledge the status of the reader's knowledge base," and "a successful assessment mechanism must be integrative in nature" (p.193).

Three of the most common methods of assessment in reading comprehension are the multiple-choice test, the written recall method, and the cloze reading test, which are discussed below.

Written Recall Protocol Method

The written recall method is now used more often in L2 reading studies for the method of assessing comprehension (Basse, & Vogt, 2022; Dixon, Oxley, Gellert, & Nash, 2022). The recall procedure is a simple tool where the final tally comes from the summing of the individual propositions properly recalled. Therefore, item and reliability analyses compared with the ones analyzing multiple-choice tests are recommended be done on recall procedures. However, in first language (L1) and L2 reading research, there seems to be an agreement that the recall of texts is a valid and integrative method to assess the process of reading comprehension (e.g. Bernhardt, 1983, 1991; Brantmeier, 2016; Carr, 2011; Hammadou, 1991; Johnston, 1983; Lee, 1990; Heinz, 1993). Bernhardt (1991) contends that the written recall method is a well-grounded measurement of reading comprehension because it reflects modern L2 theories (e.g. the constructivist model).

First language and L2 research has shown that free recall does provide valid information (Carrell, 1993; Clark, 1982, Lee, 1990), and Johnston (1983) believes that the recall method is a direct method of assessing the text-reader interaction. The recall method can also shed light on other ways learners use to remember information and can uncover the methods of reconstruction that the learner uses to code data and words while reading. Further, Bernhardt (1991) states that the recall method can circumvent some of the pitfalls that are connected to multiple-choice tests because they are not providing key information or indications to text content and the reader must also understand the text good enough to remember it clearly and logically. Therefore, the recall method doesn't interfere with a learner's understanding. Instead, according to Bernhardt (1991), it "constitutes a purer measure of comprehension, uncomplicated by linguistic performance

and tester interference” (p. 200). Bernhardt (1991) outlined 9 advantages to the use of the recall method:

1. The reader is unable to guess answers.
2. It does not influence understanding.
3. The technique is easy for teachers to prepare.
4. It induces positive feeling because readers are not penalized for paraphrasing.
5. It provides both quantitative and qualitative data.
6. It reveals grammar weaknesses that interfere with comprehension.
7. It stresses the importance of comprehension.
8. It reveals individual comprehension strategies.
9. Teachers can customize instruction to individual needs.

Bernhardt (1991) also suggested that the best method of measuring reading comprehension is the immediate recall method procedure, and offered a six step procedure:

1. A text is selected of about 200 words.
2. Students are told to read the article as many times as they need and when they're done, they're told to write as much as they can recall from the text
3. Learners are given an adequate amount of time to read a text, more than once.
4. The learners are told to remove the text while writing things down that they recall.
5. The protocols written by the students are collected
6. The data are evaluated.

However, while there are some strengths to using the written recall protocol, there are also some practical and realistic objections to the recall method that need to be considered. A large problem with this procedure is that the rewriting of information is not necessarily the same as comprehension. A reader may understand an idea in the text, but not remember it and include it in the recall. According to Johnston (1983), nothing can be said about the understanding and memory of what is not recalled. And, a reader may also recall something they read, but not comprehend its meaning.

Another limitation of using the recall method is that there are some readers, especially English as a second language (ESL) or English as a foreign language (EFL) readers, who may have difficulty in expressing their ideas in writing (Maarof, 1998). The recall method heavily relies on production skills which learners may or may not have. A bad recall isn't necessarily because they do not understand. Instead, it may just be a reflection of poor writing productions skills. This is especially true in a L2 or foreign language. Some researchers (e.g. Hammadou, 1991; and Lee, 1990; Maarof, 1998) have

suggested a solution to this by using written recall in the readers' native language. Therefore, weak language production skills should not interfere with measuring the learner's comprehension. However, there is no agreement on this point, as Cohen (1994) cited one study that found no significant difference between the recalls in the native language of Japanese and the L2 of English for advanced ESL students in Japan.

Schmidt-Rinehart (1994) has also questioned whether or not the recall method is an authentic method of assessing L2 learners' reading comprehension. She analyzed research that used the recall protocol procedure and came to the following conclusion: "although the advantages of this comprehension measure are well documented ...one would rarely be asked to perform a similar task in real life" (p. 186). And, while other researchers (e.g. Heinz, 2004) have argued that the skill of orally describing what one reads is the same as a written description, this researcher believes that oral production and written production are different skills and should be assessed differently.

Scoring written recalls is probably the biggest problem. One means for evaluating recalls as described by Meyer (1975), involves separating the text into idea units, or units of text that contain one complete idea. These idea units are then weighted according to their importance to the text. Another means for evaluating recalls by Johnston (1970) involves a scoring system. This system uses 'pausal' acceptability units to divide up the text. Each unit is created by asking the learners to indicate where pausing occurs (i.e. the place where the reader pauses when reading orally to catch their breath, highlight something, or to enhance meaning). Each pausal unit is then evaluated by their connection to the content of the text. Generally, this is decided by a group of native (or near native) speakers who have been taught how to rate and work independently. (For a more detailed information regarding the pausal scoring system, see Berkemeyer, 1989; Bernhardt, 1991; Maarof, 1998) Both of these systems of evaluation are still being used today. However, in a cross-validation study between the two systems by Bernhardt (1991), she concluded that Johnson's system appeared to be better because it provides the same information as Meyer's system, but it is less time consuming.

More drawbacks that the recall method has are that traditional scoring is very time consuming (Alderson, 2000). Reading learners' writings could take up a considerable amount of time, especially in large classes. There could also be problems in administering the recall protocol task that could interfere with the results. Alderson (2000) and Lee (1986) both objected to the quick recall procedure, stating that it might just a test of memory instead of assessing reading comprehension.

Multiple-Choice Testing

Multiple-choice testing is probably the most well-known, and used, form of reading comprehension assessment used today (Brown & Abdunabi, 2017; Butler, 2018; Farr, Pritchard, & Smitten, 1990; Perkins & Parish, 1988; Statman, 1988). According to Johnston (1983), they are the “most researched, most maligned, most difficult to construct, most abused, yet most functional of all items (when properly harnessed)” (p.59). In general, a multiple-choice reading tests will consist of a fairly short text, perhaps a few paragraphs, with a set of fixed questions or statements pertaining to the text (Butler, 2018). The test taker then chooses the answer they think matches the question or statement. In theory, if the test taker understands what they have read, they should be able to select the correct answers. As with the written recall method, multiple-choice tests have both numerous advantages and limitations.

Aiken (1987) provides a list of several advantages:

1. They are more reliable than other means because of their objectivity and their ability to be less susceptible to guessing, as compared to true/false tests.
2. They are versatile in measuring specific and complex objectives at nearly all
3. proficiency levels.
4. They can provide an adequate sampling of the domains to be tested.
5. They can be quickly and accurately scored.
6. They can enable an easy and objective item analysis.
7. They can provide diagnostic information through analyses of responses to the
8. alternatives in the items.

While there are many researchers who agree with Aiken (e.g. Farr & Carey, 1986; Oller, 1979), other researchers (e.g. Bernhardt, 1991; Perkins & Parish, 1988) have pointed out several disadvantages of the multiple-choice test. The first is that there can be only one correct answer, and because of this, good items are harder to construct. Since some multiple-choice tests consist of choices stating “all of the above”, it can often be confusing for test takers, especially those with limited English proficiency. Further, if a test taker comes across this style of question and can identify two answers, possibly out of four or five, they can easily select the “all of the above” answer as being correct without actually understanding the other possible answers. Often, a longer response time is needed in order to read through the options. Another disadvantage is that the emphasis is on the recognition of the predetermined correct answer, and not necessarily on recall or organization.

Several researchers believe that the multiple-choice tests actually penalize creative readers (e.g. Perkins & Parish, 1988; Spivey, 1989). In fact, even Aiken (1987) agrees and stated that multiple-choice tests tend to indulge “shrewd, nimble-witted, rapid readers” (p. 44) and they can even encourage inferior instruction and bad study habits. Since there is an emphasis on the recognition of the predetermined correct answer, and not necessarily on recall or organization, it may be easy for test takers to merely use general test taking strategies instead of actually trying to comprehend the text. This point may be illustrated by the enormous amount of English language test preparation schools or courses throughout the world, and the numbers of students who are deemed *qualified* by these exams (e.g. the TOEFL or IELTS) and enter into the higher education system of English speaking countries with almost nonfunctional English abilities.

Cloze Tests

A Cloze test is intended to gage a learners' reading comprehension skills by giving the learner small texts with areas left blank where normally there be a word, and then having the learners to write in the empty spaces with appropriate, and grammatically correct, words. When taking a cloze test one must have the skills to comprehend what the text is about and the vocabulary used. One must be able to identify the correct words, or style or word, which should be entered into the blank spaces of the text. The term *cloze* comes from the spoken abbreviation of the word closure. The cloze test, in many forms, has been widely dispensed for the evaluation of both L1 and L2 learning and teaching.

There are generally five central styles of cloze reading comprehension tests that teachers can utilize: The fixed-rate deletion cloze, the selective deletion cloze, multiple-choice cloze, the cloze elide, and the C-test (Gree, 2022; Ikeguchi 1995;; Klein-Braley & Raatz 1984; Weir 1990; Zulianti, Wiratno, & Novia, 2022). In the fixed-rate deletion cloze, following a few sentences every *nth* word is taken out and left blank. Normally, it's about the sixth word. However, Brown (1983) believes that with larger articles, about the twelfth or fourteenth word should be removed so that learners whose language ability is lower. In a selective deletion cloze test, the test giver selects the words that will be deleted from a text.

The main purpose of implementing this style of test is to control how difficult the text is and/or to measure the precise awareness of grammar items and/or vocabulary items. Multiple choice cloze tests are similar in that there are words that have been deleted, but many different items to select from for every empty space are provided for the test takers. The cloze elide test is quite unique. For this style of cloze test, wrong words have been intentionally placed the text, and the test takers need to locate the wrong words and insert a correct word. The C-test is also unique, in that it is created by taking out a piece of every second word in a text. The test takers must then complete each unfinished word.

The meaning of how successful one is on a cloze test varies because it depends on the larger objectives surrounding the test. The test may be dependent on if the test is an objective test, where the test takers are given a group of words to utilize in the cloze, or a subjective test, where the test takers are supposed to write in the empty spaces with a word that would allow the given sentences to make grammatical sense. The score on a cloze test is generally the percentage of the correct words entered. However, since the intended objective of the cloze test is to test comprehension, synonyms and misspellings are usually allowed.

There are several advantages and disadvantages in using a cloze test. One of the main advantages of using a cloze test is the ease and speed that a language instructor can create it. Virtually any text that has been identified at a language appropriate level can be used to create a cloze test. All a teacher has to do, is omit some of the words in the text. This feature alone makes the cloze test highly popular with language teachers.

The close test is also quite versatile for language learning. Besides comprehension assessment, a cloze test can also be used as a learning activity or for classroom or individual language practice. Due to the relative ease of its design and the advances in technology, there are many cloze tests available online for individual or classroom use. And, there are even specific websites for teachers to create their own cloze tests that match their curriculum.

The procedure for taking a cloze test can also be quite simple and time efficient. In theory, since the test is assessing a learners' reading comprehension, the length of the test will be determined by the learners' reading speed. If the learner comprehends the text, they're likely to fill in the blanks and move quite quickly. If the learner does not understand the text, they would simply leave the blank spaces empty and then continue reading the text. However, theory is not always a reality in the classroom.

It might actually take the learners a long time to complete a cloze test if there are no given time constraints. Learners who do comprehend the text may still need additional time to think of, and select, their answers. In fact, one of the main disadvantages of a cloze test is that it is not merely testing a learners' comprehension, but also their amount of additional vocabulary knowledge. Therefore, the cloze test may provide inaccurate, and invalid, results for reading comprehension because learners may in fact understand the text, but still leave blanks empty because they do not know the appropriate word to place in the blank space. Learners who do understand the text may spend an enormous amount of time trying to think of an appropriate response, and this may also lead to a

learners' frustration and tension level, which may also have a negative influence over the test takers' overall success.

Another disadvantage of the cloze test is the amount of time it might take to rate the test results. This depends on which style of cloze a teacher is using, but since there may be more than one correct answer given, it may take the rater a longer time to assess the answers given by the learner. While word choice may be limited by the test design to a specific grammatical item, like an adjective, the learners' choice of words may require a deeper evaluation. The learners' cultural background may also play a large role in the learners' word selection. It would be important to consider this while assessing the test outcomes.

A final point about cloze tests is that there is disagreement about what the tests can and cannot measure. Some researchers believe that a cloze test can assess the language learner's general communication abilities in the goal language (Ajideh, Ansarin, & Mozaffarzadeh, 2020; Hanania and Shikhani 1986 Suryani, 2014). However, other researchers contend that cloze tests can only assess the simplest of L1 and L2 reading understanding (Davood, Borzabadi, & Farahani, 2011; Gree, 2022; Shanahan, Kamil and Tobin, 1982; Zulianti, Wiratno, & Novia, 2022).

Conclusion

Reading is a complicated cognitive procedure of decoding symbols so that one can generate meaning. It's a complicated involvement with the reader and a text, and it's directed by the reader's prior knowledge, experiences, attitude, and culture. Readers may utilize several reading methods to help them understand words and decode. Therefore, teachers and researchers should also utilize a variety of methods to evaluate and assess a learner's reading comprehension.

While all three of the reading comprehension assessment tools mentioned in this article provide their own unique advantages to both learners and teachers, they also all have their own disadvantages. Based upon these findings, it seems clear that the best approach to assessing learner's reading comprehension would be a multiple-measures approach. Several other researchers (e.g. Aiken, 1987; Farr & Carey, 1986; Johnston, 1983; Wolf, 1993a, 1993b; Zulianti, Wiratno, & Novia, 2022) have encouraged instructors to use several methods of assessing reading comprehension, using a blend of both product and process assessments. Since each student may perform differently on different assessment tools, and since there may also be something different with the quality of information one recalls as a function of text structure, using a multiple-measure approach should provide more accurate and valid information that teachers can use to assess their students and

better design instruction to what their learners need. In doing so, and instructor should be creating more active and interactive student-centered instruction.

References

- Aiken, L. (1987). Testing with multiple-choice items. *Journal of Research and Development in Education*, 20, 45-58.
- Ajideh, P., Ansarin, A., & Mozaffarzadeh, S. (2020). A shift in cloze procedure usage: A new tool for activating ZPD. *Indonesian Journal of Applied Linguistics*, 9(3), 646-656. doi:<https://doi.org/10.17509/ijal.v9i3.23215>
- Alderson, J. C. (2000). *Assessing Reading*. Cambridge, MA: Cambridge University Press.
- Bassey, A., & Vogt. (2022). Diagnostic assessment of academic reading: Peeping into students' annotated texts. *Languages* 7(84), 1-19. <https://doi.org/10.3390/languages7020084>
- Berkemeyer, V. C. (1989). Recall protocol data: Some classroom implications. *Die Unterrichtspraxis*, 21(3), 131-137.
- Bernhardt, E. B. (1983). Three approaches to reading comprehension in intermediate German. *Modern Language Journal*, 67, 111-115.
- Bernhardt, E. B. (1991). *Reading development in a second-language*. Norwood, NJ: Ablex.
- Brantmeier, C. (2016). The effects of language of assessment and L2 reading performance. *The Reading Matrix*, 6(1).
- Brown, J.D. (1983). "A closer look at the cloze: Validity and reliability." In J.W. Oller, Jr. (Ed.) *Issues in Language Testing Research*. (p. 237-250). Rowley, MA: Newbury House.
- Brown, G. & Abdulnabi, H. (2017). Evaluating the quality of higher education instructor constructed multiple-choice tests: Impact on student grades. *Front. Education*, 2(24). doi; 10.3389/educ.2017.00024
- Butler, A. (2018). Multiple-choice testing in education: Are the best practices for assessment also good for learning?, *Journal of Applied Research in Memory and Cognition*, 7(3), 323-331. <https://doi.org/10.1016/j.jarmac.2018.07.002>
- Carr, N.T. (2011). *Designing and analyzing language tests*. Oxford University Press.
- Carrell, P. (1993). Introduction: Interactive approaches to second language & reading. In P.L. Carrell, J. Devine, & D. E. Eskey (Eds.), *Interactive approaches to second language reading* (pp. 1-7). Cambridge: Cambridge University Press.
- Clark, C. (1982). Assessing free recall. *The Reading Teacher*, 35, 434-439.
- Cohen, A. (1994). *Assessing language ability in the classroom*. Boston: Heinle & Heinle.
- Collins, A. A., & Lindström, E. R. (2021).). Making sense of reading comprehension assessments: Guidance for evaluating student performance. *Intervention in School and Clinic*, 57(1), 23–31. <https://doi.org/10.1177/1053451221994806>
- Davood, Borzabadi, & Farahani, (2011). The effect of regular practice with cloze passages on EFL students' reading ability. *Theory and Practice in Language Studies*, 1(8).DOI:10.4304/tpls.1.8.941-946
- Dixon, C., Oxley, E., Gellert, A.S. & Nash (2022). Dynamic assessment as a predictor of

- reading development: a systematic review. *Reading and Writing*.
<https://doi.org/10.1007/s11145-022-10312-3>
- Farr, R. & Carrey, R. (1986). *Reading: What can be measured?* Newark, NJ: International Reading Association.
- Farr, R., Pritchard, R., & Smitten, B. (1990). A description of what happens when an examinee takes a multiple-choice reading comprehension test. *Journal of Educational Measurements*, 27 (3), 209-226.
- Gree, A. (2022). *Cloze Procedure Purpose and Techniques*. Retrieved on September 1, from: <https://study.com/learn/lesson/cloze-procedure-purpose-techniques-what-is-a-cloze-procedure.html>
- Hammadou, J. (1991) Interrelationships among prior knowledge, inference, and language proficiency in foreign language reading. *Modern Language Journal*, 75, 27-38.
- Hanania, E. and Shikhani, M. (1986). 'Interrelationships among three tests of language proficiency: Standardized ESL, cloze and writing.' *TESOL Quarterly*, 20, 97- 109.
- Heinz, P. (2004). Towards enhanced second language reading comprehension assessment: Computerized versus manual scoring of written recall protocols. *Reading in a Foreign Language*, 16 (2), 97-124.
- Ikeguchi, C. (1995) "Cloze testing options for the classroom." in J.D. Brown and S. Yamashita (Eds.) 1995. *Language Testing in Japan* (p. 166-178). Tokyo: The Japan Association for Language Teaching.
- Johnson, R. (1970). Recall of prose as a function of the structural importance of linguistic units. *Journal of Verbal Learning and Linguistic Behavior*, 9, 12-20.
- Johnston, P. (1983). *Reading comprehension assessment: A cognitive basis*. Newark, NJ: International Reading Association.
- Klein-Braley, C. & Raatz, U. (1984). "A survey of research on the C-test." *Language Testing*, 1, 134-146.
- Lee, J. F. (1986). On the use of the recall task to measure L2 reading comprehension. *Studies in Second-language Acquisition*, 8, 83-93.
- Lee, J.F. (1990). Constructive process evidenced by early stage non-native readers of Spanish in comprehending an expository text. *Hispanic Linguistics*, 4(1), 129-148.
- Maarof, N. (1998). *Assessing Second-language Reading*. Selangor, Malaysia: Faculty of Language Studies, Universiti Kebangsaan Malaysia.
- Meyer, B. (1975). *The organization of prose and its effects on memory*. Amsterdam: North-Holland.
- Meyers, S. (1991). Performance in reading comprehension-product or process? *Educational Review*, 43(3), 257-272.
- Nunan, D. (1991). *Language teaching methodology: A textbook for teachers*. Englewood Cliffs, NJ: Prentice Hall.
- Oller, J. W. Jr. (1979). *Language test a school: A pragmatic approach*. London: Longman.
- Perkins, K. & Perish, C. (1988). *What's wrong with reading comprehension tests?* ERIC Document Reproduction Service No. ED 297 305.
- Schmidt-Rinehart, B. C. (1994). The effects of topic familiarity on second-language listening comprehension. *Modern Language Journal*, 78, 179-189.

- Seda, I. (1989). *Assessment format and comprehension performance*. Paper presented at the annual meeting of the International Reading Association, New Orleans, LA. ERIC Document Reproduction Service No. ED 310 362
- Shanahan, T., Kamil, M.L., & Tobin, A. (1982). 'Cloze as a measure of intersentential comprehension.' *Reading Research Quarterly*, 17, 229-225.
- Spivey, N. (1989). *Constructing constructivism: Reading research in the United States* (Report No.12). Berkeley: University of California, Center for the Study of Writing. ERIC Document Reproduction Service No. ED 310 358.
- Statman, S. (1988). Ask a clear question and get a clear answer: An enquiry into the question/answer and sentence completion formats of multiple-choice items. *System*, 16(3), 367-376.
- Sternberg, R. (1991). Are we reading too much into reading comprehension tests? *Journal of Reading*, 34(7), 540-545.
- Suryani, (2014). Teaching reading comprehension by applying cloze procedure technique. *Getsempena English Education Journal*, 1(2).
- Weir, C. (1990). *Communicative Language Testing*. Hemel Hempstead: Prentice Hall International Ltd.
- Wixson, K., & Peters, C. (1987). Comprehension assessment: Implementing an interactive view of reading. *Educational Psychologist*, 22, 333-356.
- Wolf, D. (1993a). A comparison of assessment tasks used to measure EFL reading comprehension. *The Modern Language Journal*, 77, 473-489.
- Wolf, D. (1993b). Issues in reading comprehension assessment: Implications for the development of research instruments and classroom tests. *Foreign Language Annals*, 26, 322-331.
- Zulianti, H., Wiratno, P., & Novia, S. (2022). Using cloze strategy to improve students' reading comprehension. *Journal of Language Intelligence and Culture*, 4(1), 1–14. <https://doi.org/10.35719/jlic.v4i1.78>

Author

Aaron David Mermelstein is a Washington State, U.S.A. certified K-12 teacher with .A. in Teaching and Ph.D. in TESOL. He taught MS and HS ESL before moving to Asia, where he's spent 22 years teaching EFL at the post-secondary level. His specialties include: student-centered teaching and teaching methodologies.

National University of Kaohsiung
 No. 700, Kaohsiung University Road,
 Nanzi District, Kaohsiung, Taiwan, 811
 Office (07) 591-9000 #1102, mobile 0903-452415 merman2@hotmail.com