Towards Enhanced Second Language Reading Comprehension Assessment: Qualitative and Quantitative Methods

MME MEBARKI Zahia
Université Ferhat Abbas-Sétif

Abstract:

The article reports the results of an investigation conducted at Ferhat Abbas University. The general research questions addressed in the study were: (1) To what extent does the written recall reflect reading comprehension? (2) To what extent do the results of a reading comprehension test correlate with those of the written recall?

Data for this study came from two sources: a reading comprehension test and a written recall protocol of eight fourth year microbiology students

Results of the study revealed that the written recall is an authentic method that helps measure the reading comprehension level of test takers, both in qualitative and quantitative terms and that there is a positive correlation between the test score and the recall scores.

الملخص:

يقدم المقال نتائج بحث ودراسة تناولت تقويم مستوى فهم النصوص الإنجليزية لعينة تتكون من ثماني طلبة من معهد البيولوجيا اختصاص ميكروبيولوجيا في جامعة فرحات عباس-سطيف.

الدراسة قائمة على موازنة بين نقطيين للتقويم: -الأولى: فحص تقليدي. - والثانية: تدعى«written recal»، مكن أن نقول أنها قراءة بنائية تقوم على قراءات سابقة.

النتائج تبين بكل وضوح فعالية التقنية الثانية في إبراز المستوى الفهمي للطلبة، بالإضافة إلى التطابق الإيجابي بين نتائج التقنيين.

Introduction

During the last decades, higher education has witnessed substantial changes that consist in the development of new teaching methods which emphasize the importance of the learning process and which give assessment a major role. Furthermore, second language (L2) reading comprehension

assessment has also witnessed a shift away from the classical quantitative, product-oriented measurement techniques (i.e., multiple choice, filling the gaps and cloze) to both quantitative and qualitative product-oriented measurement techniques (written recall) and also to process-oriented ones (think-aloud).

The present article sheds light on two reading comprehension assessment methods, a quantitative reading comprehension test, and a qualitative measurement technique, namely immediate free written recall. The study aims at answering the following questions:

1. To what extent does the written recall reflect reading comprehension?

2. To what extent do the results in of a reading comprehension test correlate with those of the written recall?

To answer the above questions, we have chosen eight, four-year microbiology students for whom reading in English is of primordial importance to get access to the latest and up-to-date information in their domain of study, and for whom the text has moved from being, to use Johns & Davies' (1983) expression, a linguistic object (TALO) to a vehicle of information (TAVI). To assess their reading comprehension level, we have analysed results from a classical reading comprehension technique and a written recall task.

1. The Nature of Reading:

Before discussing the assessment of reading comprehension, it would be useful to appeal to some concept of what it means to read texts and to understand them. In short, those who need to test reading clearly need to develop some idea of what reading is. We need to start somewhere, and we will do so by considering the nature of reading then the reading process and product.

Reading is a complex process which has been extensively studied across a wide range of different disciplines. This is manifestly reflected in the vast literature on reading and on the teaching and testing of reading in both mother tongue and foreign language classrooms contexts.

Reading can be viewed as decoding process of reconstructing the author's intended meaning via recognizing the printed letters and words, and building up a meaning from the text.

1.2. Reading Process Vs Reading Product:

It is a common ground to make a distinction between the process of reading and the result of that process, the product i.e. comprehension. Thus, we can view reading as a process or examine the product of that process. As a matter of fact, in any theory of reading the focus is placed on process or product.

Process models may be sequential, that is, they model the reading process as a series of stages, each of which is complete before the next stage begins, or interactive, that is every component in the reading process can interact with any other component.

Reading product (or comprehension) is the goal of the reading process. It is what reading is about; i.e., getting information from written texts. For Rayner and Pollatsek (1989), comprehension equals ' the meaning of the text' that is being read. While reading, readers maintain constant focus on constructing the meaning throughout the process, always seeking the most direct path to meaning (Carrell et al, 1988). Any reader's comprehension is variable depending on the semantic background brought by the reader to any given reading task; that is, the process taking place between the reader and the text in which background knowledge and various types of language knowledge interact with information in the text to contribute to text comprehension.

2. Methods for Assessing Reading:

The term assessment refers to a variety of ways of collecting information on a learner's language ability or achievement.

Reading comprehension assessment methods can be broadly categorized into product-oriented or process-oriented. Product-oriented methods can further be categorized into quantitative measurement techniques (multiple choice, filling the gaps and cloze), or qualitative

measurement techniques (written and orall recall protocols). On the other hand, process-oriented techniques involve think-aloud protocols.

2.1. Tests

There are three main kinds of tests: proficiency, achievement and diagnostic tests. Proficiency tests are designed to measure people's ability in a language regardless of any training they may have had in that language. The content of the proficiency test is based on a specification of what candidates have to be able to do in the language in order to be proficient. Achievement tests are directly related to language courses, their purpose being to establish how successful individual students, groups of students, or the courses themselves have been in achieving objectives. Diagnostic tests, on the other hand, seek to identify those areas in which a student needs further help. These tests can be used to identify students' strengths and weaknesses.

2.1.2. Some Test Techniques:

There is a variety of test techniques used to assess reading comprehension, namely multiple-choice items, matching, gap filling, information transfer, and the summary test.

Multiple-Choice items are a common device for testing students' text comprehension. The candidates provide evidence of successful reading by making a mark against one out of a number of alternatives.

Matching: By "matching", we mean items where students are given a list of possible answers, which they have to match with some other list of words, phrases, sentences, or paragraphs.

Gap-filling refers to tests in which the candidate is given a short passage in which some words or phrases have been deleted. The candidate's task is to restore the missing words.

Information Transfer technique is associated with graphic texts. Candidates usually have to transfer material from the text on a chart, table, form, or map.

2.2. The Recall Protocol Procedure:

2.2.1. Definition:

The method of assessing reading comprehension through written recall involves a reader recalling what she/he has read without further recourse to the text. The basic assumption behind the use of recall protocols as a measure of reading comprehension is that there is a direct relationship between what readers comprehend from a text and what they are able to recall. Consequently, those who comprehend a text better will recall it better.

Since the early 1980s, the use of recall protocols has been recommended not just as an instructional tool but also as a means for assessment of reading comprehension. Its proponents (Bernhardt, 1991; Johnston, 1970) see it as a more valid method for assessing reading comprehension than most of the other current methods.

Recall protocols can be classified as either oral or written in terms of the language mode, or either immediate or delayed in terms of time of the recall, or either free or probed; i.e., with or without cues for recalls. First, a text is analyzed in terms of idea units (or propositions) and this analysis becomes a template for scoring recalls. The number of propositions recalled after listening or reading will be counted as scores

Th basic procedures of the immediate –recall protocol includes:

a) **selecting an unglossed text,**
b) **telling students that they may read the text as often as they wish, and that when they are finished you will ask them to write down everything that they remember from the text,**
c) **giving students sufficient time to read the text several times,**
d) **Asking the students to put the text out of sight and to write down everything that they can remember, and**
e) **Collecting the protocols written by the students. (Bernhardt, 1991: 187)**

2.2.2. Advantages: Recall as an authentic assessment task

The recall protocol procedure is viewed as an authentic task and clear alternative to use in conjunction with more traditional tests such as the multiple-choice questions (MCQ) and cloze tests. It is a procedure that has several distinct advantages with respect to foreign-language testing.

Bernhardt (1983) considers the main benefits of recall protocols as follows:

- The recall procedure shows where a lack of grammatical skill interferes with comprehension, without focusing the reader's attention on linguistic elements; i.e., it interferes with the student / text communication.

- The recall procedure does not influence the reader's understanding of the text (whereas reading comprehension questions form another "text" for comprehension).

- The procedure stresses the importance of understanding. Students cannot simply guess at answers; they must attempt to form an understanding of the text.

2.2.3. Disadvantages of the recall protocol:

As with any assessment measure, limitations to the recall protocol exist and must be acknowledged. Objections have been raised regarding the use of recall protocols for assessment purposes. For example, such recalls provide quantitative information about how much students know but do not suggest how they think about what they know (Herman, 1991). Another frequently listed disadvantage of the recall protocols concerns production difficulties in L2. If subjects were required to produce their written recall in the L2, the result may be confounded by their production ability. To avoid this shortcoming, most studies have required subjects to recall in their native language so as not to interfere with their ability to demonstrate comprehension. Still this solution raises the problem that readers are more likely to include interference from their own background when they write in their mother tongue.

Furthermore, Alderson (2000) and Lee (1986) take the issue that the recall protocol may be more of a test of memory rather than a measure of comprehension and since the recall occurs immediately after reading these objections are diminished.

Alderson (2000) and points to the major disadvantages of the recall protocol procedure, practical in nature, that traditional scoring is very time consuming and rater-reliability may be open to question; thus, traditional studies using the recall protocol were necessarily limited to small groups.

2.2.4. Different Types of Recall Presentation:

Fransson (1984) proposes three different types of recall presentation.

- The mentioning type of recall presentation, which gives a more mention of the fact that the author discussed a certain problem or concept in the text, is common among subjects who have treated the text as some kind of catalogue where they looked for correct answers.

- The description type of recall presentation is the most common type among surface-level readers. Subjects read with only a general idea of the demands of an expected test. They seem to organize the different parts of the text as a sequence of separate pieces of information.

- Subjects who discuss recalls relating to a specific problem or concept and provide some kind of conclusion give the conclusion-oriented type of recall. This type is found exclusively among deep-level readers. It reflects the research for meaning characteristic of this type of comprehension.

2.2.5. Scoring the Written Recall Protocol:

The intent of measuring recall is to see what the individual comprehended. It reflects the ability to organize information into a schema; thus, recall of both the main topic and supporting details are important. For example, in a passage of 5 main topics, recall of the 5 main topics is scored using a 0, 1, 2, scale. A '0' is given if the subject does not identify the main topic; a '1' is given if the subject can identify the gist of the main topic but cannot clearly state the main topic and a '2' is given if the subject can clearly identify

the main topic. A total of 10 points is possible for the recall of all five main topics. Additionally, recall of supporting details is measured. Although recall of supporting details is an important indication that comprehension occurred. If a participant only remembered specific details and did not remember the main topics, the organization of information is more fragmentary and true comprehension still is not achieved. Recalling two supporting details is sufficient for measuring comprehension without taking away from the importance of also recalling main topics. For each of the five sections, a maximum of two supporting details is scored. If a participant recalls no supporting details in a section, he/she will have a score of '0' for that section. If a participant recalls one supporting detail in a section, he/she will have a score of '1' for that section. If a participant recalls more than one supporting detail in a section, he/she will have a score of '2' for that section. A total of 10 points is possible for the recall of supporting details in the five main topic sections.

Recall takers are also assessed in terms of their recall structures, which can be identified mainly as those which followed the original passage structure, and those which deviated from the original structure.

3. Method

3.1. Subjects

Eight subjects participated in this experiment; they are fourth year Micro-biology students. The subjects are all in their final year of study and they have to prepare a dissertation on their speciality; much of the documentation is written in English; thus, they need to read extensively in English.

The choice of the subjects for the recall task was based on the scores they obtained in the reading comprehension test. The subjects were arranged in a rank order based on their achievements in the reading comprehension test into: high achievers and low achievers (4 high-achievers and 4 low- achievers).

3.2. Materials:

The experiment is based on an authentic text, i.e. one whose original form has been kept intact, the reason being that it is the sort of texts the subjects

would read in their speciality and eventually for their dissertations. Furthermore, the topic of the text was not random. In fact, a list of topics was proposed to the subjects including: environmental microbiology, clinical microbiology and food microbiology. Five students chose a text from the second topic. It is entitled: "Contamination, Infection, and Disease". (See attached appendix)

The topic, being chosen by the subjects, would motivate them and generate more interest in the information contained in the text. Hence, we believe that the more motivated and interested they are, the more they will understand and eventually recall.

The length of the text -424 words- was motivated by the fact that it should be long enough to contain sufficient idea units- propositions- to recall. A shorter text would contain less idea units to be recalled; and a longer text would be tiresome and boring which would affect the students' motivation for reading and for recalling.

3.3. Procedure:

The experiment was divided into two parts. In the first part, students were submitted to a reading comprehension test which is part of a Ph. D thesis undertaken by the writer of the present article. The test is made up of two reading passages followed by the following tasks: guessing words from context, determining word function, finding opposites, finding synonyms, reading for main ideas, reading for details, information transfer, multiple-choice exercise, matching, and gap-filling[(1)].

In the second part, on which our experiment is based, students were asked to think-aloud while reading the text. Then, after they gave back the text, they were asked to recall what they have understood. They were not told beforehand that they would be required to recall the text afterwards. Participants were left the choice to recall in whatever language they feel at ease with, using their own words or words from the original text.

3.4. Idea Units and Scoring:

Information of the text was reduced to 13 idea units (main ideas+ supporting details) for the purpose of marking.

Main ideas:

Definitions of the concepts of:

1. **Contamination.**
2. **Infection.**
3. **Disease.**
4. **The process through which contamination occurs.**
5. **Contamination develops into infection.**
6. **Infection develops into disease.**
7. **Reversible diseases.**
8. **Irreversible diseases.**

Supporting Details:

1. **The difference between commensals and parasites.**
2. **What an infestation is.**
3. **The procedure followed to avoid getting through the three different states (aseptic procedures).**
4. **Cases of reversible disease.**
5. **Cases of irreversible disease.**

The recall protocols were marked. For each unit recalled one mark was given, the total being 13.

If the protocol offers an organised description about the topic, it is classified as a collection of description type. If it mentions the fact that the author discusses a certain concept, it is classified as a mentioning type..And if it provides some kind of conclusion, it is a conclusion-oriented type of representation. One mark is given for the structure and organisation of ideas; the maximum, thus, being 14.

3.5. Results and Analysis:

Research Question One: To what extent does the written recall reflect reading

comprehension?

In order to answer the first research question, we proceeded by analysing the written recall protocols, first qualitatively, then quantitavely as follows.

3.5.1. Qualitative Analysis:

The type of the recall presentation is on the one hand a collection of description, with four recalls of our subjects. And two subjects followed the mentioning type of presentation. On the other hand, two of the recalls fit into the description of the conclusion-oriented type. The conclusions drawn were often introduced by speculative ways of making statements which give the subject freedom to make his/her own understandings and conclusions. Instead of writing everything they could remember from the passage, they chose to recall the passage by giving a summarised account of it, seeking help from their background knowledge which reflects the research for meaning characteristic of this type of comprehension for deep-level readers. Whereas, subjects who try to re-create the author's message and seem to organize their recalls as separate pieces of information or try to mention that the author discussed a certain concept in the text are surface-level readers (submissive readers).

The way subjects organize their recalls reflects their level of comprehension and the extent to which they have comprehended and grasped the meaning in the text. Thus, as far as scoring is concerned, it is the text's level of comprehension, as reflected by the recall presentation which counts. Moreover, the recall provides a rich sample of their individual construction of the text, without taking into account grammatical and orthographic mistakes.

3.5.2. Quantitative Analysis:

The idea units recalled by the whole group are as follows:

Subject	The Idea Units Recalled	%
LM	1- 2- 3- 4- 5- 6- 7- 8- 10- 11- 12- 13	92.30%
TI	1- 2- 3- 5- 11- 12	46.15%
LN	1-2-4-5-6-11	46.15%
KS	2-4-5-6-7-8-9-11	61.53%
BD	2-11	15.38%

OM	1-5-11	23.07%
KF	11	07.67%
DH	2-5	15.38%

Table 1: The Idea Units Recalled by the Subjects

NB The numbers 1, 2, 3, 4, etc., refer to the idea units mentioned above.

The number of idea units recalled by each subject indicates the score obtained. For example, LM has recalled idea units: 1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, so, the score is 12.

The table below indicates each idea unit frequency of occurrence.

Idea units	Frequency of occurence	%
1. Definition of contamination	4	10.25%
2. Definition of infection	5	12.82%
3. Definition of disease	2	05.12%
4. The process through which contamination occurs	3	07.69%
5. Contamination develops into infection	6	15.38%
6. Infection develops into disease	3	07.69%
7. Reversible disease	2	05.12%
8. Irreversible disease	2	05.12%
9. The difference between commensals and parasites	1	02.56%
10. What an infestation is?	1	02.56%
11. Aseptic procedures	7	18.42
12. Cases of reversible disease	2	05.12%
13. Cases of irreversible disease	1	02.56%
Total	39	100%
Mean	4.87	

Table 2: The frequency of occurrence of the idea units

The figures in the table suggest that the most comprehended idea units are idea units 11, 5, 2, and 1 being the most frequently recalled since the more comprehended an idea is the more it is recalled.

On the whole, we can say that there is an average 4.87 idea unit recall per student which can be considered below average. In other words, the subjects'comprehension of the text can be said to be poor.

To sum up, we can consider the written recall, from what has been discussed in both qualitative and quantitative analyses, as an authentic method for assessing reading comprehension since it reflects comprehension.

Research Question Two: To what extent do the results in a classical reading comprehension test correlate with those of the written recall?

To answer the question, we did compare between the test scores and the written reacll scores as follows.

Concerning the relationship between the reading comprehension test and the recall task, the tables below indicate the scores obtained in both the reading comprehension test and the recall task of the text:

Students'Codes	Test Scores /60	Recall Scores /14
LM	48	13
TI	43.5	07
LN	43.50	07
KS	42.50	09
BD	16	02
OM	15.50	03
KF	12.50	01
DH	13.50	02
Mean	29.37	5.5

Table 3: Reading Comprehension Test and Written Recall Scores Analysis:

As can be seen from table 3, the scores range from 12.5 to 48 out of a maximum of 60 for the reading comprehension test; whereas, in table 4, the scores range from 1 to 13 out of 14 for the recall task.

The scores of the students can be analysed thoroughly by calculating the means then comparing them. The mean, for the test is 29.37; that is, below average. The mean, for the recall is 5.5; that is, below average.

We notice that, the reading comprehension test's mean is higher than the recall's mean; which means that, subjects did better in the reading comprehension test. But still, they are slightly convergent and below average which hints that there is a certain correlation between the results of both tests. To futher detect this correlation, we consider the mean of high-achievers, then we will compare it with the mean of low-achievers.

High-achievers	Test scores /60	Recall Scores /14	Low-achievers	Test scores /60	Recall Scores /14
LM	48	13	BD	16	02
Ti	43.50	07	OM	15.50	03
LN	43.50	07	KF	12.50	01
KS	42.50	09	DH	13.50	02
Mean	44.37	09	Mean	14.37	02

Table 4: High and Low- achievers' Test and Recall Scores

Analysis:

As can be seen from table 5, the level of scores is ranging from 42.5 to 48 out of a maximum of 60 for the reading comprehension test with a mean of 44.37; that is, high and from 07 to 13 out of 14 for the recall procedure with a mean of 09; that is, high as well.

Table 6, indicates that the level of scores is ranging from 12.5 to 16 with a mean of 14.37; that is, low, and from 01 to 03 with a mean of 02; that is, low as well.

In other words, subjects who have obtained the highest scores in the reading comprehension test (high-achievers) have also obtained the highest scores in the recall task. In the same way, subjects who have obtained the lowest scores in the reading comprehension test (low-achievers), have also obtained the lowest scores in the recall task. As shown by the scores, the high/low level of reading comprehension is much the same in both tests which shows that there is a positive correlation between the reading comprehension test and the recall.

Conclusion

The present study demonstrates that the procedure of combining two assessment techniques will necessarily provide a better picture of the reading comprehension level of test takers and lead to enhanced reading comprehension test development. Therefore, and in order to report on a variety of the readers' ability to understand through reading tests, one should seek to use multiple techniques. However good a given test might be, a single score will always mislead (Spolsky, 1994).

Acknowledgements

I am grateful for the students Bourenane Cherif Bochra and Abderrezak Abir who helped in the realization of this study. Thanks should also go to the anonymous microbiology students who participated in the study by making the written recall protocols.

Bibliography

Alderson, C. (2000). Assessing Reading. Cambridge University Press.

Bernhardt, E. B. (1983a). Testing foreign language reading comprehension: The immediate recall protocol. Die Unterrichtspraxis, 16, 27-33.

Bernhardt, E. B. (1991). Reading development in a second-language. Norwood, NJ: Ablex.

Carrell, P, L.; Devine, J.; & Eskey, D. (1988). Interactive Approaches to Second Language Reading. Cambridge University Press.

Fransson, A. (1984). 'Cramming or Understanding? Effects of Intrinsic and Extrinsic Motivation on Approach to Learning and Test performance'. In J. C. Alderson and A.H. Urquhart (eds.), Reading in a foreign language. London: Longman.

Johns, T., and Davies, F. (1983). "Text as a Vehicle for Informatio: the Classroom Use of Written Texts in Teaching Reading in a Foreign Language." Reading in a Foreign Language, 1, 01-19.

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.

Lee, J.F. (1986) 'On the use of recall task to measure L2 reading comprehension'. Studies in Second Languge Acquisition 8, (1), 83-93.
Rayner, K. and A. Polatsek (1989). The Psycholody of Reading. Englewood Cliffs, NJ: Prentice-Hall.

Spolsky, B. (1994). 'Comprehension Testing or can Understanding be Measured?' In G. Brown, A. Politt and J. Williams (eds.) Language and Understanding (pp.139-153) Oxford: Oxford University Press.

[1] For more details about the test, see the unpublished Ph.D thesis by the author of the article.

Appendix 01: Text used for written recall testing technique

Contamination, Infection, and Disease[1]

Contamination, infection, and disease can be viewed as a sequence of conditions in which the severity of the effects microorganisms have on their hosts increases. Contamination means that the microorganisms are present. Inanimate objects and the surfaces of skin and mucous membranes can be contaminated with a wide variety of microorganisms. Commensals do not harm, but parasites have the capacity to invade tissues. Infection refers to the multiplication of any parasitic organism within or upon the host's body. (Sometimes the term infestation is used to refer to the presence of larger parasites, such as worms or arthropods, in or on the body.) If an infection disrupts the normal functioning of the host, disease occurs. Disease is a disturbance in the state of health wherein the body cannot carry out all its normal functions.

Both infection and disease result from interactions between parasites and their hosts. Sometimes an infection produces no observable effect on the host even though organisms have invaded tissues. More often an infection produces observable disturbances in the host's state of health; that is, disease occurs. When an infection causes disease, the effects of the disease range from mild to severe.

Let us look at some examples to understand the differences among contamination, infection, and disease. A health care worker who fails to follow aseptic procedures while dressing a skin wound contaminates her hands with staphylococci. However, after she finishes her task, she washes her hands properly and suffers no ill effects. Although her hands were contaminated, she did not develop an infection. Another worker performing the same task on another patient

---

[1] From 'Contamination, Infection, and Disease'. By Black, J. G. (1999) Microbiology: Principles and Explorations.
John Wiley&Sons, INC. p 377.

fails to wash his hands properly after treating the patient, and the organisms gain entrance to the body and infect a small cut. Soon the skin round the cut becomes reddened for a day or so. This worker was contaminated and infected. In similar situation, a third worker develops a reddened area on her skin; she ignores it and in a few days has a large boil. This worker has experienced contamination, infection, and disease.

Disease, or illness, is characterized by changes in the host that interfere with normal function. These changes can be mild, severe but reversible or irreversible. For example, if you become infected with one of the viruses that cause the common cold, you may have just a runny nose for a few days. Or you may have a severe cold with a sore throat, cough, fever, and headache, but the disease runs its course in a week or so without any permanent effects. The changes in your state of health are reversible. But if you develop trachoma, a bacterial infection of the eye, without treatment scarring of the cornea can occur, leading to permanent vision impairment and sometimes to blindness. Likewise, if you fail to get proper treatment for streptococcal infections, you might suffer irreversible damage to your heart or kidneys.