The Written Recall Protocol as an Alternative Reading Comprehension Test

Rafiah Abd Rahman
Zaitun Abd Majid
Arshad Abd Samad, Universiti Putra Malaysia

Available at: https://works.bepress.com/arshad_abdsamad/17/
The Written Recall Protocol as an Alternative Reading Comprehension Test

Rafiah Abdul Rahman, Maktab Perguruan Batu Pahat
Zaitun Abdul Majid, Maktab Rendah Sains Mara
Arshad Abdul Samad, Universiti Putra Malaysia

Reading comprehension is a complex behaviour involving conscious and unconscious use of various strategies. Its assessment, therefore, cannot be taken for granted as it may require specialised test formats that can effectively measure the use of these strategies. According to Alderson (2000), no single test method can fulfil all the varied purposes for which we might test. Certain methods are commonly used merely for reasons of convenience and efficiency, often at the expense of validity. Therefore, it is naïve to assume that a single method commonly applied to all students and situations can be the most valid. Several test formats have been suggested for assessing reading comprehension. These include the multiple choice format, the cloze test, and summary writing. According to Riley and Lee (1996), valid measures of reading comprehension consist of integrative tasks that reflect the constructive processes involved in reading as opposed to discrete-point tasks which are narrow in scope and may reflect a fragmented, compartmentalized reading of a text.

The difficulty in selecting an appropriate measure of reading comprehension is further compounded by student response to task types. A few studies have been conducted to directly compare readers’ performance across various reading comprehension tasks. Shohamy (1984), for example, investigated the effect of testing method, level of proficiency and language assessment on second language (L2) reading comprehension of readers of English as a foreign language (EFL). She used multiple-choice and open-
ended questions presented in the subjects’ first language and L2. The results showed that learners performed best on multiple-choice questions presented in their L1 with greater effects for students with a low level of reading proficiency. Lee (1987), examined the effects of testing method on L2 readers of Spanish and found significant effects for task types and language experience. He found no difference in performance on cloze tasks and questions asked in the first language (L1). There was also no difference on readers’ performance on the recall tasks regardless of whether it was recalled in the L1 or L2. In another study, Wolf (1993), investigated task type, language assessment and L2 experience. There were no significant effects for all three variables on student performance. The students, however, performed best on multiple choice questions, followed by open-ended questions and then the cloze task. Shohamy, Lee and Wolf all conclude that the method of assessing reading comprehension influences, and perhaps even determines, how readers perform on a reading comprehension test. These factors, together with the need to accurately measure reading comprehension, underscore the importance of appropriate test formats for testing reading comprehension.

While the written recall protocol is increasingly being used in L2 reading research as a measure of comprehension (Horiba, 1996; Berndhart, 1991; Carell, 1984; Riley & Lee 1996; Fauziah, 1999), it is seldom used as a test instrument. This paper, therefore, attempts to examine the potential of the written recall protocol as an alternative to the more commonly used reading comprehension test in secondary schools. Students’ responses in their written recall protocols are expected to provide insights into the
appropriateness of the method as a test of reading comprehension, especially with respect to issues of rubrics, scoring procedure, and construct validity.

Traditional Modes of Testing Reading

Among the more common formats used in reading tests are the multiple choice and cloze tests. Multiple choice formats are the most familiar test format in the Malaysian education system. A number of studies have been conducted on testing reading using the multiple choice format (Drum et al., 1981; Shohamy, 1984; Wolf, 1993; Freedle and Kostin, 1999). Several concerns regarding multiple-choice reading comprehension tests, especially in terms of construct validity, were raised in these studies. Drum et al. (1981), for example, studied the predictability of multiple-choice item difficulty of children’s reading comprehension tests. Their results suggest that the multiple-choice reading comprehension test may lack construct validity because the item predictors appeared to be more prominent than the text predictors. Ideally, a language comprehension test should primarily assess the difficulty of the text itself in which the item structure is only an incidental device for the purpose (Freedle and Kostin, 1999).

Although multiple-choice questions are easy to score and are economical, they have their own limitations. From the perspective of reading comprehension, the main limitation is passage dependency. Students may be able to correctly answer the questions by calling upon their prior knowledge or even by guessing without having to read the passage. Moreover, the answer to an item may also be revealed by another item due to the inter-
relatedness of the items (Fauziah, 1999). Another limitation is that the technique tends to restrict what can be tested to only lower order skills, in particular the recall level.

While the cloze test is another option in testing reading comprehension that has often been used, language testing researchers have long argued about what cloze tests actually measure. The cloze test was initially intended to measure reading ability but was later considered to be a test of general language proficiency (Alderson, 2000). Cloze tests are very useful in many situations because they are easy to prepare and score. However, their validity as tests of reading comprehension is somewhat controversial.

Some studies (Alderson, 1979; Shanahan et al., 1982) report that cloze tests do not tap processing abilities beyond the clause level (Sasaki, 2000). Shanahan, Kamil and Tobin’s (1982) study provides evidence that cloze tests do not support processing at the intersentential level, much less at the inter-paragraph level where different types of information need to be linked and integrated. However, according to Sasaki (2000), empirical research has demonstrated that cloze tests can measure both lower (clause level) and higher level (intersentential) reading abilities depending on the particular word deleted and on the proficiency level of the participants.

On the other hand, others claim that a cloze task taps the reader’s ability to make use of syntactic and semantic knowledge, an ability that is fundamental to comprehension. Bachman (1985) and Langer (1985), for example, argue that a cloze task is sensitive to the integration of information across sentences. Langer states that integrating local
understandings across different parts of a text is a very significant aspect of developing an understanding of a text as a whole.

Despite their limitations, both the multiple choice and cloze test formats remain popular formats in assessing reading comprehension. The concerns raised with regard to these formats, however, require investigation into the use of other appropriate formats. In this paper, the potential of the written recall protocol as a test of reading comprehension will be examined.

**The Written Recall Protocol**

Since the early 1980’s, the use of the recall protocol has been recommended as a means of assessing reading comprehension (Johnston, 1983). It is seen as a more valid test format for assessing reading comprehension compared to the multiple choice and cloze test formats (Cohen, 1994). The recall protocol is recommended because, unlike the more traditional tests, it provides no leading information pertaining to passage content and in responding, the reader is expected to demonstrate comprehension by integrating the components of the reading passage (Bernhardt, 1991). Bernhardt further asserts that the recall protocols provide data that reflect the nature of the reading process in terms of encoding, restructuring and analysing information and is a more valid measure of reading comprehension because it conforms to current L2 reading research-driven theories.

In written recall protocol tests, students are asked to read a text, to put it aside, and then to write down everything they can remember from the text. The recall test is an example
of what Bachman and Palmer (1996) call an extended production response type task. Bernhardt (1991) considers two main benefits of recall protocols. First, they show where a lack of grammar interferes with comprehension without focusing the readers’ attention on linguistic elements. Second, they do not influence the reader’s understanding of the text unlike in multiple choice formats where the questions form another “text” for comprehension.

Johnston (1983) states that recall protocols provide information about the extent to which the lack of competence in grammar may have affected comprehension and they are therefore regarded as the “most straightforward assessment of the result of the text-reader interaction” (p.54). In a recall protocol, a direct relationship is assumed between what readers comprehend from a text and what they are able to recall.

Despite the benefits discussed here, there remain criticisms regarding the use of recall protocols for assessment purposes. Among the issues that remain controversial are whether the protocols are written in the students’ L1 or the target language; the rubrics given to the students and whether the task was performed immediately or delayed.

The issue of having the protocols in the students’ first language (L1) has often been debated. Lee (1987) carried out a number of studies on L1 and L2 recall protocols and concluded that “assessing comprehension with a target language task may limit learners’ ability to demonstrate what they comprehended” (p.353). The current and more acceptable view is that the use of L1 should not be totally dismissed since the purpose of
the test is reading comprehension and not writing (Fauziah, 2002). Previous studies using recall protocols as research instruments have allowed the respondents to recall in their L1 (Block, 1986, 1992; Davis and Bistodeau, 1993). This is to avoid poor language competence in the L2 affecting the assessment of what information was recalled. Upton (1993), however, found that the language of recall for more advanced learners does not influence the quantity of ideas recalled.

A second issue concerns the rubrics that the respondents are given in completing the task (Swaffar et al., 1991). The quantity and quality of ideas recalled depend on the instructions that were given to the respondents. Therefore, respondents must be very clear on the level of detail required so that an accurate representation of what is recalled can be achieved.

Apart from rubrics, the issue of whether a recall protocol is a test of memory or comprehension needs to be given consideration. The quality of recall may be affected by the time lapse between when the text is read and when the recall takes place. It is therefore recommended that the respondents recall immediately after the task is completed so that the information recalled is not influenced by knowledge outside the text. Thus, written recall protocols will not be a test of memory if the task follows immediately after the reading (Alderson, 2000). This suggestion is especially appropriate for test situations as tests are constrained by time and require immediate response.
According to Alderson (2000), good reading tests are likely to employ a number of different test formats, possibly even on the same text, but certainly across the range of texts tested. Therefore, the written recall protocol should be considered as an alternative for assessing reading comprehension. By using this format, test questions do not intervene between the reader and the text and the format is therefore seen as a purer measure of comprehension. However, issues of test rubrics, the use of the L1 for recall, and time lapse discussed here need to be examined before the written recall protocol can be effectively used in testing reading comprehension.

Design
This section discusses the design of the study which focuses on the instruments, scoring procedure and the administration of the written recall protocol. Thirty-three Form Four students of a secondary school participated in this study.

Materials and Instrumentation
The instruments employed in this study were the students’ written recalls. The reading passage used for the recall task is a narrative text which is an adapted version of a Japanese story (Appendix A). A narrative text was chosen because the narrative structure can be considered the most facilitative to comprehension as story telling is an important component of all cultures (Hatch, 1992). A neutral, non-culturally biased topic was chosen to control for the effects of background knowledge in reading. Based on scores on the Fry Readability Index, the reading passage used can be considered relatively easier
than texts normally found in English language textbooks for students of the same age group.

**Scoring**

The quantity of information recalled is generally used as the operational definition of reading comprehension in studies using the recall task (Barnett, 1986 and Lee, 1987) with Meyer’s (1985) scoring system frequently used to score the written recall protocol. Texts are divided into idea units and relationships between idea units are then coded. The reading passage used in the study consists of 8 main ideas and 1 supporting detail (See Appendix B). A recall protocol is scored according to the presence or absence of the idea units represented in the scoring template. However, this scoring system is time consuming. This study uses an alternative technique suggested by Block (1986) in which the text is analyzed into two categories of idea units – main ideas and supporting details. The total score derived is based on the total number of main ideas and supporting details correctly recalled. In this study, therefore, the maximum score is 9.

**Procedure and rubrics**

The students in this study were given 5 minutes to read the passage. After the time allocated was over, the students were asked to put away the passage and to write down, in either English or Bahasa Malaysia, everything they had understood. Students were allowed to write in point form as the focus of the instrument was to gather information on what they had comprehended from the passage and not to assess their writing skills. The written protocols were collected after 10 minutes.
Results and Discussion

The results of the written recall protocols were analysed using descriptive statistics. Mean recall scores and standard deviations are presented in Table 1.

Table 1: Mean and Standard Deviation of Number of Ideas Recalled

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main ideas (k = 8)</td>
<td>4.67</td>
<td>1.38</td>
</tr>
<tr>
<td>Supporting details</td>
<td>0.72</td>
<td>0.46</td>
</tr>
<tr>
<td>Total ideas</td>
<td>5.39</td>
<td>1.43</td>
</tr>
</tbody>
</table>

The mean for total ideas recalled is 5.39 (SD = 1.43). This illustrates that, on the average, students recalled about 59.9% of the 9 ideas in the text. The mean score for main ideas recalled is 4.66 while the mean score for supporting details recalled is 0.72.

Type of ideas recalled. Analysis of the results showed that the first and second main ideas were recalled the most often (i.e. by 91% of the students). This could be due to the fact that these two ideas were the lead in to the passage and helped establish the context of the passage. The next most frequently main idea recalled was the fifth idea (88%). The fifth idea is rather significant in the text as it acts as the climax to the passage and probably that was the reason it was recalled by a relatively large number of the students. The only supporting detail in the passage - ‘its face looking ready to cry’ - was also found to have been recalled by more students (82%) than the other main ideas. The
students may have found this detail significant by associating it to the thief’s unsuccessful plan.

Another interesting finding is the types of verbs found in the protocols. The students were found to recall verbs and phrases that indicated physical action such as *crying*, *opened its eyes* and *climb* more often than less action-oriented verbs such as *sleeping*. This indicates that action verbs may be more salient in the reading process among second language learners compared to less action-oriented verbs.

**Recall according to proficiency level.** Students in this study were also categorised as higher proficiency or lower proficiency according to their performance on the Lower Secondary Assessment Examination, a national level standardised examination, taken the previous year. It is interesting to note that although both the higher proficiency and lower proficiency groups inferred and extended upon the main ideas in the original passage, they did so in different ways. For example, the higher proficiency group ended their protocol by inferring that the baby saved the day. They elaborated upon the story line and provided closure to the story through their familiarity with the rhetorical structure of stories.

As for the lower proficiency group, the students elaborated by stating that the baby cried and made the thief run away. This seemed to show that the students relate the passage to their prior knowledge of babies crying when they see strangers. This may indicate that the lower proficiency students may have not fully comprehended the story and therefore
embellished the story by relying on a general knowledge schemata instead of providing closure to the recall consistent with the rhetorical structure of stories. The lack of comprehension of the passage is also demonstrated by their inability to correctly sequence the main ideas.

**Scoring the recall protocol.** The written recall protocol showed that scoring is an important issue that needs to be looked into. Students were found to be using their own words in the written recall protocols and an appropriate scoring system is therefore required in order to assess the use of these words. It is suggested that a scoring template be used whereby scores will be awarded according to the words used. The suggested scoring template for this particular recall protocol is illustrated in Figure 1.

<table>
<thead>
<tr>
<th>2 MARKS</th>
<th>1 MARK</th>
</tr>
</thead>
<tbody>
<tr>
<td>1M* A thief sneaked into the attic of a house. <strong>Other acceptable answers:</strong> A thief crept into the roof space of a house. A thief slipped into the upper floor of a house.</td>
<td>A thief came into the attic of the house. A thief came into the roof of the house.</td>
</tr>
<tr>
<td>2M When he looked down, he saw a father, a mother, and a baby sleeping.</td>
<td>He saw a father, a mother and a baby sleeping.</td>
</tr>
<tr>
<td>3M The father and the mother were sound asleep. <strong>Other acceptable answers:</strong> The father and mother were in deep sleep. The father and mother were fast asleep.</td>
<td>The father and mother were asleep.</td>
</tr>
<tr>
<td>4M The man felt relieved. <strong>Other acceptable answers:</strong> The man felt thankful.</td>
<td></td>
</tr>
<tr>
<td>5M The baby who was sleeping between its parents opened its eyes wide.</td>
<td>The baby opened its eyes.</td>
</tr>
<tr>
<td>6M The man hurriedly climbed back into the attic. <strong>Other acceptable answers:</strong> The man quickly climbed back into the attic.</td>
<td>The man climbed back into the attic.</td>
</tr>
<tr>
<td>7M The baby was looking up towards the man</td>
<td></td>
</tr>
<tr>
<td>1D** its face looking ready to cry <strong>Other acceptable answers:</strong> its face was prepared to cry</td>
<td>its face shows that it is going to cry</td>
</tr>
</tbody>
</table>
its face was about to cry

N/B. * Main idea, ** Supporting detail

**Figure 1. Suggested Scoring Template**

The total scores for the written recall protocol will be based on the sum of main ideas and supporting ideas correctly recalled. To interpret the scores, it is suggested that a summary scoring template as in Figure 2 below be used.

<table>
<thead>
<tr>
<th>Student</th>
<th>Main idea</th>
<th>Supporting detail</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1M</td>
<td>2M</td>
</tr>
<tr>
<td></td>
<td>3M</td>
<td>4M</td>
</tr>
<tr>
<td></td>
<td>5M</td>
<td>6M</td>
</tr>
<tr>
<td></td>
<td>7M</td>
<td>TOTAL</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 2. Summary of Scores Template**

**Administration of the written recall protocol.** In preparing for this study, the rubrics was considered a particularly important aspect in the administration of the written recall protocol. In order for the written recall protocol to measure reading comprehension, the rubrics must state clearly that the protocol is testing reading comprehension and not testing writing. Teachers must be very clear of the objective of the test and they should avoid penalising students for errors that are normally associated with writing such as grammatical errors and organisation.
It is also suggested that the written recall protocol be written in the target language. One may argue that comprehension can be demonstrated without the use of the target language and its use may even confuse the assessment of comprehension with that of writing. Furthermore, as mentioned earlier, research involving the use of recall protocol often allows the use of the first rather than the target language in the recall. However, while both research and testing involve accurate measurement of a particular variable, testing situations often differ from research. This is especially true in the Malaysian context where tests are often high stakes and standardised. In such a situation where the test is highly controlled in terms of time limit, procedures and scoring, it is more appropriate that the target language be used in the written recall protocol. Allowing students to use their first language would also be too drastic a departure from normal practice and may introduce an unknown effect into the measurement.

**Effect on teaching and learning.** Tests are often discussed in terms of its washback or influence on teaching and learning. How would teachers help students to prepare for written recall protocol? It is expected that teachers would train students to look out for propositions - the basic units of meaning in the text. This represents a positive washback on teaching and learning as focusing on propositions will help students to better comprehend texts. As Yang (2000) argues, as long as the student understands the meaning of a certain proposition, he/she is said to be involved in comprehension. Texts are also expected to be carefully selected, with those that are more facilitative to comprehension – involving more obvious propositions expressed through vocabulary that is more action-oriented - used earlier than texts which are more demanding.
Conclusion

In order for the written recall protocol to be used as a measure of reading comprehension, it must have construct validity. Previous studies indicate that the written recall protocol has already been accepted in research on reading as it is able to access reading processes. While the written recall protocol has not yet been frequently utilised in testing situations, the results of this study indicate that it has the potential to be used as a reading comprehension test. Several issues, however, need to be first resolved so that students and teachers can prepare for the test in a more organised and directed manner. Some, such as student and teacher familiarity to the test task and format as well as the objectives of the written recall protocol may be achieved through washback. Students and teachers should be familiar with the test, its objectives as well as format. Once adopted, the continued use of the written recall protocol as well as test preparation will ensure familiarity.

The written recall protocol also needs to be more specified especially in the scoring procedure. The marking scheme needs to be very detailed to ensure that students are fairly scored. Additionally, the scores should reflect students’ ability to comprehend. In this study, it was also noted that the rubrics are also important as it gives the expectation of what students need to do. The rubrics given prior to the task will also allow students to focus on comprehending rather than memorising. While rubrics in assessment is sometimes considered as irrelevant to the construct that is being tested or assessed, nevertheless, it can be argued that the ability to accurately measure reading processes
through the written recall protocol far outweighs the disadvantages of using lengthy rubrics.

In general, the analysis of students’ written recall protocol indicate that the protocol is a viable alternative for assessing reading comprehension in secondary schools. This is a positive development for the testing of reading comprehension as students need to be exposed to different test techniques besides multiple choice and cloze tests. Good reading tests are likely to employ a number of different test formats rather than rely on only one (Alderson, 2000) and the written recall protocol may be considered as an additional format available for this purpose.

References


**Appendix A**

**Reading Passage**
Once upon a time, a thief sneaked into the attic of a house. When he looked down, he saw a father, a mother, and a baby sleeping. Both the father and the mother were sound asleep. “Good. They’re all sound asleep.” As the man, feeling relieved, was about to climb down, the baby, who was sleeping between its parents, opened its eyes wide. “Oh, oh.” The man hurriedly climbed back into the attic. The baby was looking up towards the man with a face looking ready to cry.

Adapted from Horiba (1996)
When he looked down, he saw a father, a mother, and a baby sleeping.

The father and the mother were sound asleep.

The man felt relieved.

The baby who was sleeping between its parents opened its eyes wide.

The man hurriedly climbed back into the attic.

The baby was looking up towards the man.

its face looking ready to cry