

An Examination of Free Written Recall Tasks
as Listening Comprehension Tests

Hideki Sakai
Shinshu University

JABAET Journal

Number 9

October 2005 所載論文

Published by

THE JAPAN-BRITAIN ASSOCIATION FOR ENGLISH TEACHING

An Examination of Free Written Recall Tasks as Listening Comprehension Tests

Hideki Sakai
Shinshu University

This small-scale study examined the concurrent validity and the construct validity of free written recall tasks as listening tests. Tasks have been used as an instrument to measure comprehension ability in reading and listening. In the tasks, test-takers are told to write everything they remember about the passage(s) they read or heard. Tasks are considered to have the potential to clarify the comprehension process. At the same time, tasks have been criticized as being merely a test of memory ability. A multiple-choice task and a free written recall task were administered to 25 Japanese learners of English. Results showed a high correlation coefficient between the two tasks, and in addition, the free written recall task revealed a variety of listening processes, suggesting that the free written recall task is a useful tool as a test of listening comprehension ability.

1 Introduction

1.1 Listening comprehension

Recently much emphasis has been placed on the assessment of second language (L2) performance as well as the examination of L2 processes (e.g., Alderson, 2000; Buck, 2001). However, this is a relatively new research trend, and so it was only a decade ago that Thompson (1995) pointed out that “listening comprehension testing ... continues to remain somewhat of a neglected area” (p. 24). This small-scale study focuses on the assessment of listening comprehension ability.

Listening comprehension involves complex processes which work simultaneously. Peterson (2001, p. 88) divided listening comprehension into the following three stages: the perceptual processing phase, the parsing phase, and the utilization phase. In the perceptual processing phase, the listener recognizes “meaningful sound units, to determine syllable boundaries, and to identify words.” In the parsing phase, the listener “works with the words and phrases he or she has decoded to form meaningful units, which are stored in short-term memory.” In the utilization stage, the listener “searches long-term memory for ideas that relate to the new information; when a match is made between old and new information, comprehension occurs.” Similarly, Buck (2001, pp. 15-21) listed (a) understanding words (recognizing the word and understanding its meaning), (b) processing idea units (taking the meaning of individual words, and combining them to construct the meaning of complete utterances), (c) processing connected discourse, and

(d) using world knowledge (making inferences). These processes are carried out in the two directions, that is, bottom-up processing and top-down processing, at the same time (Morley, 2001).

The main purpose of this study is to examine the characteristics of free written recall tasks and discuss the validity of the tasks as listening comprehension tests. Thus, the question is whether the listening process can be tapped by using free written recall tasks.

1.2 Free written recall tasks

Several frameworks for different types of listening comprehension tests have been proposed (Buck, 2001; Thompson, 1995). For example, Thompson (1995) proposed two types of listening comprehension tests on the basis of the response type: selected responses or constructed responses. Tests using selected responses “do not require test-takers to create a response, merely to select an the [*sic*] most plausible option” (Thompson, 1995, p. 27). Multiple-choice questions and true-false questions are categorized as selected responses. On the other hand, tests of constructed responses, such as open-ended questions and free written recall tasks, “require test-takers to produce their own answers” (Thompson, 1995, p. 27). This study focuses on a type of listening test with constructed responses, that is, free written recall tasks.

Free written recall tasks have been used as an instrument to measure reading ability (e.g., Carrell, 1985; Lee, 1986) and listening ability (e.g., Baccouche, Castro, Messman, & Zakletskaia, 1996; VanPatten, 1990). This testing type is also called recall protocols (Lund, 1991; Thompson, 1995), immediate recall protocol (James, 1987), written recalls (Baccouche, Castro, Messman, & Zakletskaia, 1996), free written recalls (Lee, 1986; VanPattern, 1990), free-recall tests (Alderson, 2000), immediate-recall tests (Alderson, 2000), or listening recall¹ (Henning, Gary, & Gary, 1983). Buck (2001, pp. 80-82) referred to the testing type called *translation*, which basically resembles the procedure for free written recall tasks except that the translation task contains pauses between the sentences (or other units) during which test-takers are supposed to write down their translation of the passage. Translation may be regarded as a variation of free written recall tasks. For this study, I will use the term *free written recall tasks* for the sake of consistency.

Thompson (1995) listed the procedure for free written recall tasks as follows: “(1) a brief listening passage is recorded at normal speed; (2) a list is prepared of all facts or propositions contained in the passage; (3) students listen to the passage, (4) they are asked to write down everything they remember from the passage” (p. 28). Some variations are found in the previous studies as to the genre², the speed, and the length of the listening passage. VanPattern (1990) used a 274-word, three-minute passage (i.e., an expository text) in Spanish as a listening passage, which was recorded at a slow speed with pauses at clause boundaries. Baccouche, Castro, Messman, and Zakletskaia (1996) used a 505-word, three-minute short story (i.e., a narrative text) as a listening passage. Judging from the rate (168.3 words per minute), their listening passage was recorded at a normal speed. As for the fourth procedure, two tasks are possible: writing in L1 or writing in L2. In a study of reading comprehension, Lee (1986) found statistically significant effects of the language used in the recall protocols on the performance in the free written recall tasks. Alderson (2000) argued that “the recall needs to be in the first language, otherwise it becomes a test of writing as well as reading” (p. 230). Thus, in most cases, the learners’ native language

is used (Baccouche, Castro, Messman, & Zakletskaia, 1996; Lund, 1991; VanPatten, 1990).

So far researchers have used a variety of scoring methods (Alderson, 2000, pp. 230-232). This study used a scoring method called *idea unit analysis*. The passage is divided into idea units (see the second procedure in Thompson's list), and one point is provided for recall of each idea unit. One problem inherent in the scoring procedures is that "scoring of recall protocols is labor-intensive and requires training to ensure inter-rater reliability" (Thompson, 1995, p. 28). Another is the definition of idea units. Baccouche, Castro, Messman, and Zakletskaia (1996) pointed out, "the definition of idea units is not consistent across studies" (p. 14). For the present study, I followed the previous studies (Baccouche, Castro, Messman, & Zakletskaia, 1996; VanPatten, 1990) which adopted a definition of idea units proposed by Carrell (1985). The definition will be explained thoroughly in 2.3.2.

Both advantages and criticisms of the free written recall tasks have been pointed out. For instance, Alderson (2000, p. 230) argued that "this technique is often held to provide a purer measure of comprehension, since test questions do not intervene between the reader and the text." Some other forms of listening tests involve writing or reading abilities; for example, multiple-choice tasks which have written options. Another advantage is that "it [the free written recall task] is also claimed to provide a picture of learner processes" (Alderson, 2000, p. 230). Thus, free written recall tasks will provide researchers or teachers with useful information about how the learners listen in the L2.

On the other hand, one major criticism of free written recall tasks, among others, concerns memorization. Thompson (1995) stated that "critics of this technique [recall protocols] argue that it confounds listening comprehension with memory ability" (p. 28). Alderson (2000) argued that "it might be objected that this is more a test of memory than of understanding, but if the task follows immediately on the reading, this need not be the case" (p. 232). Memory is involved in any type of listening test: Test-takers listen to the passage(s) and memorize the information for questions. As Thompson (1995) pointed out, "a listener may have comprehended what was being said at the time of listening, but by the time he or she got to the question(s), the memory trace may have been erased by subsequent information in the text, and by having to read the question and answer options" (p. 27). In free written recall tasks, this memory problem may play an influential role on the performance.

1.3 Research questions

Two questions were posited for this study:

Research Question 1: Is the free written recall task a test of memory ability or a test of listening comprehension?

Research Question 2: What aspects of the listening process can free written recall tasks measure?

Generally speaking, these questions are regarded as validity issues. So far, little research has been conducted to test the validity of the free written recall task as a measure of listening comprehension ability.

Validity is defined as "the appropriateness of a given test or any of its component parts as a measure

of what it is purported to measure” (Henning, 1987, p. 89). Among several types of validity, this study mainly concerns construct validity and concurrent validity. First, concurrent validity shows “the strength of relationship with some external criterion measure” (Henning, 1987, p. 96). Brown (1988) stated that “the demonstration of criterion-referenced validity usually involves setting up an experiment in which the subjects take two tests: the test being developed and a well-established test of the construct involved” (p. 104). Usually, the correlation coefficient between the two tests is calculated. According to Alderson, Clapham, and Wall (1995), “most concurrent validity coefficients range from +.5 to +.7 --- higher coefficients are possible for closely related and reliable tests” (p. 178). Second, construct validity provides “evidence that underlying theoretical constructs being measured are themselves valid” (Henning, 1987, p. 98). Several methods for construct validation have been proposed. One of the methods is the use of the factor analysis. Alderson, Clapham, and Wall (1995) stated “the reason that these two factor analytic approaches [exploratory factor analysis and confirmatory factor analysis] belong to the class of construct validation procedures is that the factors that emerge are explained or predicted on the basis of theory” (p. 183). This study will deal with these two types of validity.

2 Method

2.1 Participants

The participants for this study consisted of 25 Japanese learners of English from two universities: 19 students from two classes in a national university and 6 students from one class in a private university. They were 11 male students and 14 female students. Most of them were in the 2nd year ($n = 20$), but there were two 3rd year students and three 4th year students among the participants.

2.2 Procedures

The data were gathered using two separate listening comprehension tasks: a multiple-choice task and a free written recall task. The tasks were administered as part of class activities with different intervals between the two tasks in the three classes (one week for one class in the national university and about eight weeks for the other two classes). Immediately before the free written recall task, participants were provided with a practice session for the recall tasks.

2.3 Instruments

2.3.1 The multiple-choice task

The multiple-choice task consisted of ten question items, which were derived from the listening part of the pre-2nd grade and 2nd grade STEP tests³ (Obunsha, 2002, 2004). For each question, a dialogue and a question were played one time, and participants were asked to choose one of the four options printed

on the answer sheet as the most appropriate answer. Ten second intervals were inserted between questions. The total time for the multiple-choice task was 6 minutes and 35 seconds.

In general, the questions require test-takers to process details of the passage in order to arrive at the correct answers (Nos. 1, 2, 4, 5, 6, 8, 9, and 10) while some questions require test-takers to make an inference based on the information in the passage (Nos. 3 and 7). The former questions were of the fact-finding type. This is an example:

No. 1

A: Mom, I'm playing golf with Sarah this afternoon. Do you think it'll rain?

B: The weather report said it should clear up by one o'clock.

A: Oh, good. But if it does rain, we plan to go bowling instead.

B: Well, Joe, you'd better go pick her up now. Drive carefully.

Question: What does Joe plan to do if it rains?

1. Go for a drive.
2. Go golfing with his mom.
3. Go to Sarah's house.
4. Go bowling with Sarah.

Test-takers must comprehend the third line (*if it does rain, we plan to go bowling instead*) to answer the question. For No. 3 and No. 7, the questions are inference questions. Nevertheless, the inference can be made easily if the details of the passages are decoded.

According to the item analysis, one question item (No. 2) gained a negative corrected item-total correlation and was eliminated from later analyses. The Cronbach alpha was .704, which indicated an acceptably high reliability.

2.3.2 The free written recall task

The passage used in the free written recall task was a 55-word narrative text from the listening part of the pre-2nd grade STEP test (Obunsha, 2004), the topic of which was a newspaper delivery. The time of the recording was 30 seconds, and therefore the passage was read at a speed of 110 words per minute. No special pauses were inserted in the text. Participants were asked to listen to the passage for comprehension and to write down everything they remembered immediately after listening. They were not allowed to take notes while listening. The time for recall was not limited. On average, it took about two minutes to complete the writing.

The recall protocols were scored depending on whether idea units were recalled or not. As to the definition of idea units, Carrell (1985) stated that:

Basically, each idea unit consisted of a single clause (main or subordinate, including adverbial and relative clauses). Each infinitival construction, gerundive, nominalized verb phrase, and conjunct was also identified as a separate idea unit. In addition, optional and/or heavy prepositional phrases

were also designated as separate idea units. (p. 737).

According to the example shown in Carrell (1985), prepositional phrases functioning as adjectives which modify noun phrases were not separated as idea units. For this study, heavy prepositional phrases were defined as prepositional phrases consisting of more than three words (Kadota & Noro, 2001, p. 301).

These criteria resulted in comparatively long units and turned out to be too strict for the analysis of protocols of listening tests. As Thompson (1995) pointed out, "people recall less information from listening than from reading in terms of both quantity and quality" (p. 24). Participants will easily lose points with longer units if the evaluation method simply requires scoring protocols as right or wrong. Therefore, for this study, some revisions concerning the criteria for determining idea units were made in order to make idea units shorter: Adverbials and non-heavy prepositional phrases as adverbials were determined as separate idea units.

Based on the above criteria, the passage for this study was divided into 15 idea units as follows.

(1) Mike has a part-time job / (2) as a newspaper delivery boy. / (3) Early every morning, / (4) he rides his bike / (5) around the neighborhood / (6) and throws newspapers / (7) into people's front yards. / (8) But [(9) last Saturday], Mike threw one newspaper / (10) too hard, / (11) and it broke a window. / (12) Now / (13) Mike must use some of the money / (14) he makes / (15) to repair the window.

Each idea unit was numbered. The criteria revision resulted in five additional idea units. Idea unit 5, which consisted of three words, was designated as an idea unit. Also, idea units 3, 9, 10, and 12, which were adverbial phrases, were isolated and considered as units. I asked one researcher to divide the text according to the above criteria to confirm my division.

The highest possible score was 15. I scored all the recall protocols. To assess inter-rater reliability for the scoring, another rater scored all the recall protocols. Agreement reached 99.5% (373 out of 375).

According to the item analysis, one idea unit (No. 9) showed a negative corrected item-total correlation and was eliminated from later analyses. The Cronbach alpha was .722.

2.3.3 Data analyses

The analyses of the validity of the free written recall tasks were based on the descriptive statistics for the two tasks, the correlation coefficient between the two tasks, and the cluster analysis as well as qualitative analyses of the protocols.

Concurrent validity will be discussed on the basis of the correlation coefficient between the two tasks. This analysis mainly concerns the first research question. The assumption was that if memory is involved in the performance for the free written recall task, the correlation coefficient between the two tasks will be low since the multiple-choice task does not require so much memory. Construct validity will be discussed by analyzing the item characteristics using cluster analysis as well as by analyzing the recall protocols qualitatively. I performed cluster analysis instead of factor analysis because the data were

binary (1 or 0) and were not suitable for factor analysis. The clustered items and the processes observed in the protocols will be compared with the theoretical constructs of listening comprehension.

3 Results & Discussion

3.1 Descriptive statistics

Table 1 shows descriptive statistics of the multiple-choice task and the free written recall task.

Table 1 Descriptive statistics of the tasks

	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>
Multiple-choice task (9 items)	25	6.280	1.882	1	9
Free written recall task (14 items)	25	4.400	2.517	1	9

The multiple-choice task consisted of the test items from the pre-2nd grade STEP and 2nd grade STEP tests, whereas the free written recall task used a passage from the pre-2nd grade STEP test. Since the 2nd grade of STEP is a higher level than the pre-2nd grade, it was expected that the learners' scores in the multiple-choice task would be lower than those of the free written recall task. However, the data did not confirm this expectation. That is, participants performed better on the multiple-choice task than on the free written recall task even though the multiple-choice task used more difficult passages, suggesting that the task format itself may have helped participants gain higher scores.

3.2 Concurrent validity issue: The correlation between the two tasks

In order to examine the relationship between the multiple-choice task and the free written task, I calculated the correlation coefficient. A high coefficient ($r = .626, p = .000811$) was obtained, which indicated a positive relationship between the two tasks (see the coefficient range from +.5 to +.7 by Alderson, Clapham, & Wall, 1995, p. 178). The coefficient of determination (the squared value of a correlation coefficient) was .392, indicating that the free written recall task overlapped by about 39.2% with the multiple-choice task which was used in the STEP listening test. The high correlation suggests that the free written recall task may measure the learners' ability to understand the details spoken in the passage.

As described in the method section, the multiple-choice task consisted of a dialogue and a question with four options for the answer. In this format, participants responded to the question immediately after the dialogue and may not have depended on their memory ability so heavily. Taking into account this point and the high correlation coefficient between the two tasks, it can be inferred that there would be less involvement of memory in the free written recall task. If the performance on the free written recall task was influenced by memory ability, which is a construct different from language ability, the correlation

between the multiple-choice task and the free written recall task would be low. In fact, this was not the case. Therefore, it is suggested that the free written recall task is not a test of memory ability, but rather a test of listening comprehension.

As for the reading recall task, Alderson (2000) stated “it might be objected that this is more a test of memory than of understanding, but if the task follows immediately on the reading, this need not be the case” (p. 232). Since the passage for this study was not a long passage (55 words, 30 seconds) as compared to VanPatten (1990) and Baccouche, Castro, Messman, & Zakletskaia (1996), memory may not have influenced the performance on the task so strongly.

3.3 Construct validity issue

3.3.1 The cluster analysis

Figure 1 shows a tree diagram of the cluster analysis using the Ward method (Yamagiwa & Tanaka, 1997). Three clusters were identified. Group 1 contains idea units 8, 10, 13, 12, 6, 7, and 14; Group 2 consists of idea units 5, 11, 3, 15, and 4; and Group 3 has idea units 1 and 2.

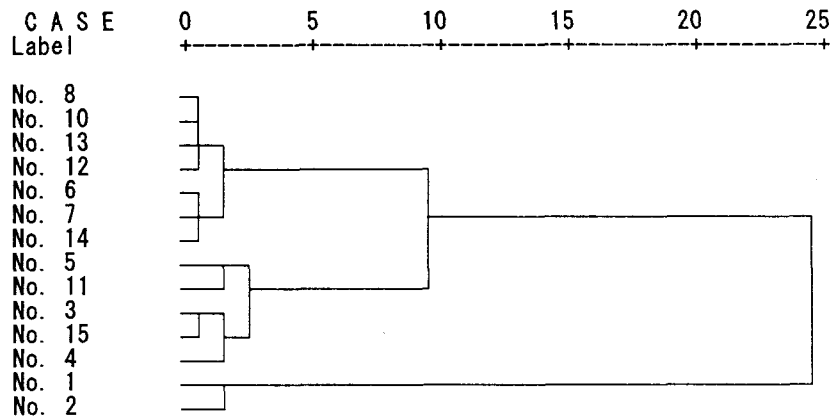


Figure 1 Dendrogram of cluster analysis using the Ward method

In general, the three groups can be regarded as the details of the story, the gist of the story, and the introduction of the story respectively. The idea units of Group 1 are (8) *But ... Mike threw one newspaper, (10) too hard, (13) Mike must use some of the money, (12) Now, (6) and throws newspapers, (7) into people's front yards, and (14) he makes.* Those for Group 2 are (5) *around the neighborhood, (11) and it broke a window, (3) Early every morning, (15) to repair the window, and (4) he rides his bike.* To consider the relationships between the idea units of the two groups, it can be argued that the idea units of Group 1 describe details whereas those of Group 2 show the more general line of the story. For example, the idea unit *and it broke a window* of Group 2 can be explained more concretely by idea units 8 and 10. Likewise, the phrase *to repair the window* of Group 2 is related to idea units 13 and 14.

The third group consisted of idea units 1 and 2. The facility values for idea units 1 and 2 were .68 and .96 respectively, which indicated that they were the easiest units to recall among the 15 idea units.

This suggests that the test-takers in the free written recall tasks may perform better in the recall protocols of the first part of the listening passage. Thus, there seems to be a positioning effect of idea units on the performance.

3.3.2 Qualitative analysis of listening processes

In this section, I will analyze the protocols for idea units 8 to 11: (8) *But* [(9) *last Saturday*] *Mike threw one newspaper / (10) too hard / (11) and it broke a window*. Two of the participants provided no protocols for these idea units. Thus, analyses were based on the protocols of 23 participants. In the following section, participants' protocols will be shown in their native language (Japanese) with the English translation.

3.3.2.1 Identifying chunks

Two of the participants were successful in recalling this phrase even though they missed most of the details. For example, one participant recalled the passage as follows:

マイクは新聞配達のバイトをしている。バイクに乗り、隣人の。先週の土曜日。マイクは貯めたお金で。

(Mike has a part-time job as a newspaper delivery boy. He rides a bike. Neighbors. Last Saturday. Mike ... with the money he saved.)

In total, the participant got six points: idea units 1, 2, 4, 5, 9, and 14. She did not recall anything for the idea units around idea unit 9. This suggests that she was successful in grasping only the phrase *last Saturday* in the listening input. Thus, idea unit 9, *Last Saturday*, seemed to be memorized as a chunk. Skehan (1998) argued that the knowledge of chunks may be helpful for the processing of incoming input and for making time for analysis of the input. The above example suggests that L2 learners listen for and pick up chunks in the stream of sounds.

3.3.2.2 Recognizing words

As for idea unit 9, four of the participants interpreted the phrase as *last Sunday*, but no one recalled this idea unit as *last Monday*, *last Tuesday*, *last Wednesday*, and so on. This was probably because the participants perceived the first part of the word *Saturday* and made use of inferences to compensate for their inability to do complete processing.

前の日曜日、彼は窓を割ってしまった。

(Last Sunday, he broke a window.)

3.3.2.3 Making Inferences

It is interesting that five of the participants wrote *one day* for idea unit 9. Maybe the participants comprehended the situation of the event in the input, but because they were not able to catch the phrase

indicating the time, they compensated for the missing information with a more general phrase to denote the time.

Like the example of the misinterpretation of *last Saturday* as *one day*, very often participants added some information to make the story more logical. Look at the following examples.

しかし、先週の日曜日、彼は新聞を配っていた時、強く投げすぎて、窓を割ってしまいました。

(But last Sunday, when he delivered newspapers, he broke a window because he threw the newspaper too hard.)

ある日、新聞配りをしていると、窓が割れてしまいました。

(One day, when he delivered newspapers, he broke a window.)

先週の土曜日、新聞を配っているときに人の家の窓ガラスを割ってしまい・・・

(Last Saturday, when he delivered newspapers, he broke a window of a house.)

These three examples show that the participants added the information *when he delivered newspapers*, which was not in the original passage. These examples indicate that participants utilized their background knowledge. In most cases, this informational addition was observed when the bottom-up processing was not successful, as in the second and third examples in which participants did not recall idea unit 8.

3.3.2.4 Making unsuccessful inferences

In some cases, participants made use of inferences, but the inferences did not lead to the correct interpretation.

先週の日曜日、マイクは1つの新聞のところに2つの新聞を置いてしまいました。窓が壊れた。

(Last Saturday, Mike placed two newspapers in the place which one newspaper should be put on. The window was broken.)

This participant's interpretation of idea unit 8 was wrong. Also, this example shows that the participant may have interpreted *too* as *two*. This suggests that this participant was able to pick up words, but had difficulty in processing the input syntactically. Look at a similar case in the following example.

ある日、配達がとても大変に感じた。窓を割った。

(One day, (he) felt very hard at delivering. (he) broke a window.)

This participant was able to perceive the phrase *too hard*, but was not able to catch the part immediately before this phrase: *Mike threw one newspaper*. When phrases have several meanings, learners may have

difficulty in interpreting the phrase correctly because of a lack of listening comprehension ability. The following examples also indicate this point.

土曜日に、マイケルは新聞を見すぎて、ガラスを割った。

(Saturday, Micheal read the newspaper too much, and he broke the window.)

ある朝、マイクはとても強く新聞を投げつけてしまいました。すると、その家の二枚のガラスを割ってしまいました。

(One day, Mike threw a newspaper too hard. And he broke two glasses of the house.)

In the first example, the participant may have picked up the word *too* and may have made some inferences to compensate for the missing information. The third example indicates that the participant was able to perceive the word *too*, but interpreted the word wrongly.

Some participants interpreted the word *window* as *wind* as follows:

先週の日曜に、2つの新聞を風で飛ばしてしまった。

(Last Saturday, [he] blew two newspapers off in the wind.)

The above examples suggest that the participants perceived sounds in the listening input and, when they were not able to identify and understand words, they depended on inferences with the help of their world knowledge.

3.3.3 Summary of the cluster analysis and the protocol analysis

The results of the cluster analysis suggest that three types of listening performance were observed on the free written recall task: listening to the first part of the passage, listening to the main line of the passage, and listening to the details of the passage. The differential performance between the main ideas and the details relates to the process of connected discourse (Buck, 2001). In other words, participants may have distinguished the important information of the story from the others while listening. According to the analyses of the protocols, the free written recall task revealed several listening processes used by the participants: dependency on chunks, perception of sounds to identify the word, inferences based on the information which they could grasp, and inferences for the missing information utilizing background knowledge. These processes were carried out in a bottom-up manner and a top-down manner. Also, the analyses suggest that inferences are not always successful. In general, the listening processes tapped by the free written recall tasks cover the processes proposed by researchers (e.g., Buck, 2001).

It must be pointed out that while misinterpretation based on inferences does not lead to a score, inferences leading to a correct interpretation gain points on both multiple-choice tasks and free written recall tasks. Thus, it remains unclear if successful performance is brought about by correct listening comprehension. However, the free written recall tasks may be better in that the performance can be analyzed even when the test-takers do not get points. On multiple-choice tasks, it is quite difficult for the

researcher to infer the reason why the test-taker has chosen a certain option.

4 Conclusion

The findings of this study are summarized as follows. The first finding is related to the nature of the free written recall task as a test of listening comprehension (Research Question 1): The latter three concern the aspects of the listening processes which free written recall tasks can measure (Research Question 2).

1. The free written recall task is not a test of memory ability, but a test of listening comprehension (based on the high correlation coefficient between the free written recall task and the multiple-choice task).
2. The free written recall task can provide information concerning the processes by which test-takers arrive at misinterpretation (see the analysis of the protocols).
3. The free written recall task may measure several aspects of the listening process: (a) the ability to understand the details of the passage (refer to the high correlation coefficient between the free written recall task and the multiple-choice task), (b) the abilities to grasp the gist and catch the details which relate to processing connected discourse (see the cluster analysis), and (c) the ability to perceive sounds, detect chunks in the spoken passage, and make inferences based on the information obtained (see the analysis of the protocols).
4. There may be a positioning effect of idea units: The idea units which appear in the introduction of the passage seem easier to recall.

These findings do not lead to conclusive generalization about the validity of the free written recall task because this study dealt with a small number of learners, just one narrative text as a listening passage, and the multiple-choice task from the STEP test as a well-established test. Also future research needs to examine, for example, the effects of a variety of passages of different difficulty, length, and genre on performance. In addition, it is necessary to investigate the relationships between free written recall tasks and other types of listening tests and the effects of different scoring procedures for free written recall tasks.

I hope that this study provided some evidence that free written recall tasks are of great value for researchers and teachers because they can provide a picture of learners' listening processes.

Notes

1. Henning, Gary, and Gary (1983) used the term *listening recall* to refer to a different type of listening testing. Content words in the passage are deleted in the sheet. The missing words are chosen "in such a way that it would be highly unlikely that the students could supply the missing words from context alone without attending to the aural stimulus" (p. 289). Test-takers are told to read the passage and then to fill

in the blank spaces while listening. Therefore, their listening recall is not a type of free written recall task, but rather a type of listening dictation.

2. James (1987) reported the effects of text type (a dialogue, a news broadcast, and a song) on performance on free written recall tasks. However, because he did not provide the levels of the three texts, it seems difficult to make a direct comparison of the recall protocols of the three texts.

3. The Society for Testing English Proficiency (STEP) tests are widely known in Japan as a test of practical English proficiency. The STEP tests have seven grades: 1st (the highest level), Pre-1st, 2nd, Pre-2nd, 3rd, 4th, and 5th (the lowest level). Find more information about the STEP tests at <http://www.eiken.or.jp/english/index.html>.

References

- Alderson, J. C. (2000). *Assessing reading*. Cambridge: Cambridge University Press.
- Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation*. Cambridge: Cambridge University Press.
- Baccouche, E., Castro, C., Messman, E., & Zakletskaia, L. (1996). *How comprehensible is input when attention is focused on form?* Paper presented at the Annual Meeting of the American Association for Applied Linguistics 1996 Conference, Chicago, IL. (ERIC Document Reproduction Service No. ED396532)
- Brown, J. D. (1988). *Understanding research in second language learning*. Cambridge: Cambridge University Press.
- Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press.
- Carrell, P. L. (1985). Facilitating ESL reading by teaching text structure. *TESOL Quarterly*, 19, 727-752.
- Henning, G. (1987). *A guide to language testing*. MA: Heinle & Heinle Publishers.
- Henning, G., Gary, N., & Gary, J. O. (1983). Listening recall --- A listening comprehension test for low proficiency learners. *System*, 11, 287-293.
- James, C. J. (1987, August). *The immediate recall protocol as an elicitation technique in the listening modality of university-level students of German*. Paper presented at the World Conference of Applied Linguistics, Sydney, Australia. (ERIC Document Reproduction Service No. ED299801).
- 門田修平・野呂忠司編著. (2001). 『英語リーディングの認知メカニズム』くろしお出版.
- Lee, J. F. (1986). On the use of the recall task to measure L2 reading comprehension. *Studies in Second Language Acquisition*, 8, 201-212.
- Lund, R. (1991). A comparison of second language listening and reading comprehension. *The Modern Language Journal*, 75, 196-204.
- Morley, J. (2001). Aural comprehension instruction: Principles and practices. In M. Celce-Murcia (Ed.), *Teaching English as a second or foreign language* (3rd ed.) (pp. 69-85). MA: Heinle & Heinle.
- 旺文社. (2002). 『2002年度版英検2級全問題集 CD』旺文社.
- 旺文社. (2004). 『2004年度版英検準2級全問題集 CD』旺文社.

- Peterson, P. W. (2001). Skills and strategies for proficient listening. In M. Celce-Murcia (Ed.), *Teaching English as a second or foreign language* (3rd ed.) (pp. 87-100). MA: Heinle & Heinle.
- Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.
- Thompson, I. (1995). Testing listening comprehension. *AATSEEL Newsletter*, 47, 24-31.
- 山際勇一郎・田中敏. (1997). 『ユーザーのための心理データの多変量解析法』教育出版.
- VanPatten, B. (1990). Attending to form and content in the input: An experiment in consciousness. *Studies in Second Language Acquisition*, 12, 287-301.