

The Study of Second Language Text Comprehension and Measurement Issues*

Euen Hyuk (Sarah) Jung
(Yonsei University)

Jung, Euen Hyuk (Sarah). (2007). The study of second language text comprehension and measurement issues. *English Teaching*, 62(3), 55-68.

Much research compared various types of reading tasks and showed that task types affect how readers comprehend the text. In contrast, few studies examined the role of task types in L2 listening. Despite the wide use of recall tasks and the usefulness of summary tasks in measuring comprehension, no study, thus far, has compared recall and summary tasks to assess L2 listening comprehension. An attempt to compare the effects of these tasks on text comprehension will provide valuable insights to better understand the assessment of spoken text processing. It is also necessary to examine not only how much information is comprehended, but also what levels of information are comprehended. Consequently, this study compared the use of recall and summary tasks to assess L2 learners' listening comprehension of high-level information and low-level information in the text. 40 Korean EFL university students participated in the study. Of 40 learners, half performed summary tasks, while the other half performed recall tasks. Results show that task types play a major role in L2 listening comprehension. Learners performing summary tasks understood significantly more high-level information, whereas learners performing recall tasks comprehended significantly more low-level information.

I. INTRODUCTION

Research shows that learners' abilities to demonstrate text comprehension vary according to the assessment task types and that the assessment tasks themselves are an influential factor in measuring text comprehension (Bachman, 1991; Brindley, 1998; Buck,

* This work was supported by the Korean Research Foundation Grant Funded by Korean Government (MOEHRD)(KRF-2005-003-B00315)

1997; Shohamy, 1997; Wolf, 1993). A study by Mills, Diehl, Birkmire, and Mou (1995), for example, suggests that task instructions (i.e., the goal or purpose of the task) affect the manner in which readers process and retrieve text information. Similarly, studies by Horiba (2000) and Lorch, Klusewitz, and Lorch (1995) show that different cognitive processing takes place according to the various task demands.

Text comprehension processes involve both top-down and bottom-up processes (Brindley, 1998; Lynch, 1998; Mendelsohn, 1998; Morley, 1995; Rubin, 1994). Top-down processing is driven by comprehenders' expectations and understandings of the text topic and organization, while bottom-up processing is triggered by comprehenders' attempt to decode linguistic input. These top-down and bottom-up processes operate in an interactive way. That is, text features and comprehenders' features interact with each other simultaneously.

As Riley and Lee (1996) maintain, valid measures of comprehension should consist of integrative tasks that can reflect interaction between the text and the comprehender (i.e., constructive processes involved in comprehension), as opposed to discrete-point tasks that may fail to accurately reflect the comprehender's true understanding of the text. Unlike the tests with set formats, such as multiple-choice question tests, recall and summary tasks seem to be more aligned with the interactive text processing perspective (Bernhardt, 1991; James, 1986a; James, 1986b). Recall tasks have been frequently used as an effective method of measuring text comprehension (James, 1986a; James, 1986b; Scott, Stansfield, & Kenyon, 1996). A number of researchers have begun to identify summarization as an important academic skill and to advocate the relevancy of summary tasks to learners as an ecologically valid assessment task (Cohen, 1994; Sung-Ae Kim, 1998; Mayes, 1990).

More specifically, there are several advantages of recall and summary tasks. First, recall and summary tasks are holistic measurements, which assess how information is integrated as a whole. Comprehenders are actively involved in constructing the mental representation of the text on the basis of their true understanding of the text (Bernhardt, 1991; James, 1986a; James, 1986b; Roebuck, 1998). Second, recall and summary tasks are likely to reflect process rather than product because they seem to pinpoint wrong reasoning and the types of errors comprehenders are making (Bernhardt, 1991; James, 1986b; Rinehart & Thomas, 1993). Lastly, recall and summary tasks require comprehenders to produce their own responses to a text and prevent them from making random guesses by providing no external clues to help them understand the text (Bernhardt & Deville, 1991). As Alderson (2000) claims, recall and summary tasks provide a purer measure of comprehension, since test questions do not intervene between the text and the comprehender.

There are also, however, several shortcomings of recall and summary tasks. First, it is time-consuming to divide the text into information units and to construct standards for scoring procedures. It is also time-consuming to score each task (Thompson, 1995).

Second, inter-rater reliability is another concern. It may not be always easy to find competent raters and to train them (Thompson, 1995). Lastly, another potential problem with recall and summary tasks is production bias. It is important to ensure that learners have the necessary production skills to demonstrate their comprehension of the text in recall and summary tasks. As several researchers argued, recall and summary tasks should not be conducted in learners' second language (L2) alone because more evidence of comprehension might be masked by the L2 use in the tasks (Berne, 1992; James, 1986a; James, 1986b; Wolf, 1993).

Despite the pros and cons that recall and summary tasks share, these assessment tasks seem to require different cognitive processing and output demands on learners. Learners set up expectations as to how to approach the particular task in accordance with its purpose, and thus, tend to engage in constructing a mental representation of the text that is appropriate for the given task. That is, a summary task requires learners to get the main points of the text and demands additional, careful evaluations of the relative importance of ideas in the text (i.e., an editing process). In contrast, a recall task requires learners to remember everything from the text.

In fact, the prior argument on the summary versus recall tasks is empirically supported by reading research, including studies by Goldman, Saul, and Cote (1995) and Riley and Lee (1996). In their studies, while learners performing a summary task recalled significantly more high-level information (major ideas) than those performing a recall task, learners performing a recall task recalled significantly more low-level information (minor ideas) than those performing a summary task. Although it has been documented in reading research that summary and recall tasks measure different levels of information in the text, no study has ever looked into this issue from a listening comprehension perspective. An attempt to compare the effects of two differing assessment tasks on learners' listening comprehension will provide useful insights to better understand the assessment of listening comprehension processes.

To measure listening comprehension processes, an increasing number of researchers argued that it is necessary to examine not only how much information is comprehended, but also what levels of information are comprehended (Alderson, 2000; Berne, 1992; Chen & Donin, 1997; Jamieson, Campbell, Norfleet, & Berbisada, 1993; Wolf, 1993). From an assessment perspective, the issue of measuring listening comprehension with respect to differing levels of information is of importance. Merely counting information units that had been correctly recalled is at the risk of providing a researcher with an inaccurate measurement of listeners' comprehension (Alderson, 2000). It is necessary to obtain an accurate assessment of listeners' text comprehension, by differentiating high-level information from low-level information in a text.

To summarize, much research has compared various types of reading tasks and showed

that task types affect how readers comprehend the text. In contrast, few studies have examined the role of task types in L2 listening. Despite the wide use of recall tasks and the usefulness of summary tasks in measuring comprehension, no study, thus far, has compared recall and summary tasks in the assessment of L2 listening comprehension. An attempt to compare the effects of these tasks on comprehension will provide valuable insights to better understand the assessment of spoken text processes. It is also necessary to examine not only how much information is comprehended, but also what levels of information are comprehended. A differentially weighted scoring scheme should be used because simply counting information units accurately recalled risks giving a distorted picture of listeners' understanding. Consequently, the present study aims to compare the use of summary and recall tasks to measure L2 learners' listening comprehension of high-level information (major ideas) and low-level information (minor ideas) in the text.

II. RESEARCH DESIGN

1. Research Questions and the Corresponding Hypotheses

Research Question 1: What is the effect of assessment task types (i.e., summary vs. recall tasks) on L2 learners' listening comprehension of high-level information (i.e., major ideas) in the text?

Hypothesis 1: Based on previous research, it is hypothesized that learners who perform a summary task (i.e., the summary group) will comprehend significantly more high-level information than those who perform a recall task (i.e., the recall group).

Research Question 2: What is the effect of assessment task types (i.e., summary vs. recall tasks) on L2 learners' listening comprehension of low-level information (i.e., minor ideas) in the text?

Hypothesis 2: Based on previous research, it is hypothesized that the recall group will comprehend significantly more low-level information than the summary group.

Research Question 3: What is the effect of assessment task types (i.e., summary vs. recall tasks) on L2 learners' listening comprehension of information from combined high and low levels (i.e., major and minor ideas combined) in the text?

Hypothesis 3: Based on previous research, it is hypothesized that there will be no difference between the summary and recall group in terms of comprehending information from combined high and low levels in the text.

2. Participants

The present study involved 40 Korean students who were taking an English language course at a major university in Korea. Most of the participants were juniors in their early or mid-twenties from various academic disciplines, including law, humanities, education, business administration, and political science. These students had studied English in Korea for seven to eight years. None had previously been abroad either to travel or to study English. Thirty of the forty students scored between 41 and 50 in the Listening Comprehension section of the paper-based Test of English as a Foreign Language (TOEFL) that they completed 1 week prior to the study. The remaining ten students scored between 51 and 60 on the test.

3. Procedures

Of 40 learners, half ($n=20$) performed summary tasks (i.e., the summary group). The other half ($n=20$) performed recall tasks (i.e., the recall group). In order to create maximally homogeneous groups as to the learners' L2 listening proficiency, learners of both groups were carefully matched on the basis of their listening comprehension scores on the TOEFL. Care was taken to balance higher and lower listening proficiency learners in both groups. Fifteen of the thirty learners who scored between 41 and 50 and five of the ten learners who scored between 51 and 60 in the listening section of the TOEFL were assigned to each of the two groups.

Each group completed its task in a language laboratory. The aural input was an audiotaped lecture from an academic lecture series. At the beginning of each session, the researcher briefly explained the purpose and procedure. Each learner received a packet containing (1) a response sheet for note-taking, (2) a topic familiarity survey sheet, (3) a response sheet with instructions for either the summary or the recall task, and (4) a personal information and English education background questionnaire.

Immediately after listening to the lecture, learners indicated their familiarity with the lecture topic on the survey sheet. All learners indicated that they had moderate topic familiarity with the lecture topic. After answering the question on lecture topic familiarity, learners performed either the summary or the recall task, as previously assigned. The learners performing summary tasks received the oral instruction from the researcher (i.e., writing down the important points of the lecture); the learners performing recall tasks also received the oral instruction from the researcher (i.e., writing down everything remembered from the lecture). The task instructions were also written on the response sheet for the tasks. While engaging in the tasks, the learners could refer to the notes that they had taken during the lecture. Once the learners completed the given task using Korean,

English, or a mixture of both, they completed the personal information and English education background questionnaire.

4. Data Analyses

The learners’ responses to the written protocols (summary and recall tasks) were scored for the presence of each of 60 high-level information units and 120 low-level information units in the lecture. These information units were previously designated by two independent researchers, who reached 98% agreement on their relative importance. The learners were given 1 point for each information unit they correctly reported. Phrases or sentences with equivalent meanings were accepted if they represented relevant ideas. While spelling errors were disregarded, distortions of the original text content were not allowed. Two raters independently evaluated the learners’ responses to the summary and recall tasks. They resolved disagreements through discussion. The current data were subjected to three separate *t* tests, using the Statistical Package for Social Sciences (SPSS) 11.0.

IV. RESULTS

The results relevant to research questions 1, 2, and 3, which address the effects of task types on learners’ comprehension of high- and low-level information as well as information from combined high and low levels, are presented first as descriptive statistics. Table 1 presents the means and standard deviations for the summary and recall task. Table 2 demonstrates inferential statistics for each corresponding hypothesis.

TABLE 1
Means (*M*) and Standard Deviations (*SD*) for the Summary and Recall Task

Level of Information	Summary Task		Recall Task	
	<i>M</i> (%)	<i>SD</i>	<i>M</i> (%)	<i>SD</i>
High-level Information	28.6	10.3	22.2	9.5
Low-level Information	20.9	7.9	29.0	10.6
Information from Combined High and Low Levels	23.4	8.3	26.8	9.9

TABLE 2
Results of *T* test

Level of Information	df	<i>t</i> Value	
High-level Information	38	2.051	<i>p</i> < .05
Low-level Information	38	2.772	<i>p</i> < .05
Information from Combined High and Low Levels	38	1.152	<i>ns</i>

Note: *ns* = not significant

Analysis of the participants' written protocols of high-level information revealed that task types play a major role in understanding the text information at a high level. As Tables 1 and 2 illustrate, learners who performed the summary task comprehended more high-level information than those who engaged in the recall task (M summary group = 28.6% vs. M recall group = 22.2%, $t(38 \text{ df}) = 2.051, p < .05$). Hypothesis 1, thus, received support. That is, the summary group understood significantly more high-level information than the recall group.

Concerning the effect of task types on learners' understanding of low-level information from the text, as shown in Tables 1 and 2, learners who performed the recall task comprehended more low-level information than those who engaged in the summary task (M summary group = 20.9% vs. M recall group = 29.0%, $t(38 \text{ df}) = 2.772, p < .05$). This result provided support for Hypothesis 2. The recall group comprehended significantly more low-level information than the summary group.

Regarding the effect of task types on the comprehension of information from combined high and low levels in the text, as Tables 1 and 2 demonstrate, the t test indicated no significant effect for tasks on learners' comprehension of information from combined high and low levels (M summary group = 23.4% vs. M recall group = 26.8%, $t(38 \text{ df}) = 1.152, ns.$). The data were consistent with Hypothesis 3 in that no difference existed between the summary and recall groups with respect to comprehending information from combined high and low levels in the text.

The results of the current study are summarized as: (1) the summary group comprehended significantly more high-level information in the text than the recall group, (2) compared to the summary group, the recall group understood significantly more low-level information in the text, and (3) no difference was found between the summary and recall groups in terms of amount of information comprehended from combined high and low levels in the text.

V. DISCUSSION

The current findings seem to indicate the crucial role that assessment task types play in foreign language learners' listening comprehension. Learners performing a summary task comprehended significantly more high-level information than those performing a recall task. Conversely, learners performing a recall task understood significantly more low-level information than those performing a summary task. These findings are consistent with the claim made by previous research that assessment task types affect learners' comprehension performance (Bachman, 1991; Berne, 1992; Brindely, 1998; Buck, 1997; Goldman et al., 1995; Horiba 2000; Lorch et al., 1995; Mills et al., 1995; Riley & Lee, 1996; Shohamy,

1997; Wolf, 1993).

Based on the previous research cited above, a plausible explanation for the current findings lies in the argument that task instructions set up learners' expectations about how to approach the particular listening task and thus, to some extent, determine the listening strategies they adopt. In other words, the manner in which learners perceive the task demands affects their listening comprehension processes. Learners perceive different output demands for recall and summary tasks. Learners are more likely to make efforts to construct a mental representation of the text that is required for the particular assessment task that they must engage in. For example, those who hear that they will perform a summary task prior to the lecture are likely to focus more on main ideas and generalize over details during the lecture, instead of trying to remember every piece of information.

In the current study, the summary task required learners to identify the main points of the lecture. The task demanded careful evaluations of the relative importance of ideas in the lecture. Such demand of the summary task made it more necessary for learners to be sensitive to the relative importance of ideas and to distinguish high-level from low-level information during the lecture. Learners might have sought more high-level information during the lecture and thus, recalled more high-level than low-level information in the summary task. Conversely, the recall task required learners to remember everything from the lecture. This demand of the recall task made it less necessary for learners to be sensitive to the relative importance of ideas and to identify high-level from low-level information during the lecture. Learners might have tried to comprehend everything by treating all information as equally important during the lecture and thus, recalled more low-level than high-level information in the recall task.

In the current study, although assessment task types were found to play a significant role in listening comprehension when learners' comprehension was examined with respect to both high- and low-level information, the task types were not shown to play a critical role when learners' listening comprehension was examined in terms of amount of information comprehended from combined high and low levels (i.e., no differentiation of levels of text information in comprehension scoring). These findings seem to show an interaction between task types (summary vs. recall) and comprehension of different levels of information units; the sensitivity of the assessment task types to the different levels of information in the text is manifested.

Different tasks appear to measure different levels of learners' listening comprehension. In this study, the summary task seemed to tap into learners' listening comprehension at a higher level (i.e., a high-order thinking task). In order to perform a summary task, learners must be sensitive to the relative importance of ideas in the lecture. The summary task required learners to be selective in attending to different levels of information in the text. Learners needed to engage in deliberate, conscious activities, which involve condensation,

generalization, and deletion to a much greater extent (Cohen, 1994; Goldman et al., 1995). In contrast, the recall task made it less necessary for the learners to be sensitive to the different levels of information in the text, but the task made it more necessary for them to attend to every information unit, irrespective of the different levels of information in the lecture.

In light of the current findings, it is necessary to revisit the issue of the listening construct. As discussed earlier, listening comprehension is a complex, multifaceted process, in which both text variables and listener variables play critical roles simultaneously. As Carrier (1999) noted, however, the current discussion of the listening construct by listening theorists including Bentley and Bacon (1996) and Glenn (1989) has failed to address what different levels of information are processed and measured. As evidenced by the current study, aural text information is processed at different levels (i.e., high- vs. low-level information) and different tasks appear to measure different levels of learners' listening comprehension.

As with reading comprehension, listening comprehension is difficult to assess because of the mental operations involved in the comprehension process. Listening comprehension ability is observable only through indirect, second-hand measurement, from which one must infer the listener's mental activity (Buck, 1997). As demonstrated by the current findings, task type plays a significant role in learners' listening comprehension processes since different tasks seem to tap into different levels of information. Therefore, to obtain a fuller picture of learners' listening comprehension, it is critical that one needs to employ various assessment tasks, considering the pros and cons of each assessment task (Berne, 1992; Brindley, 1998; Shohamy, 1997).

For example, in the summary task in the current study, learners might not have included a certain piece of information they had understood because they thought that particular information was not relevant and thus, simply chose not to include it in their written protocols (Flowerdew & Tauroza, 1995). Conversely, in the recall task, learners simply might not have recalled a certain piece of information due to memory constraints (Thompson, 1995) and thus, could not write down that particular information in the recall protocols even if they had understood it during the lecture. In sum, one will have a greater chance of gaining an accurate picture of learners' listening comprehension by appropriately selecting a combination of various assessment tasks.

It is important not only to use multiple assessment tasks but also to construct valid listening assessment tasks that will allow one to make accurate inferences concerning the listener's text understanding. The current study considered some specific criteria in devising listening assessment tasks based on previous research (Berne, 1992; Buck, 1997; Lynch, 1997; Shohamy & Inbar, 1991). The following criteria might serve as a guideline for test developers as well as for researchers in constructing listening assessment tasks:

First, listening assessment tasks should ensure that learners demonstrate comprehension skills as listeners and that tasks should clearly distinguish listening abilities from reading and writing abilities. To this end, assessment tasks (summary and recall tasks, in particular) should not require learners to complete their tasks in their L2 exclusively, but in their first language or a mixture of both; Second, listening assessment tasks should be realistic and authentic. In other words, assessment tasks should be reflective of real world situations and they should be the ones that learners are likely to encounter¹; Lastly, assessment tasks should measure learners' comprehension of different levels of information in the text. That is, tasks should measure comprehension of both high- and low-level information in the text.

The current findings have another implication for measuring learners' listening comprehension. An inaccurate assessment of learners' listening comprehension might result if their understanding is not measured separately with respect to high- and low-level information, and equal weight is given to all information units in the text. For example, if learners had understood many low-level but only a few high-level information units, an undifferentiated scoring method which only considers the number of information units recalled would have made it appear as though learners comprehended the text better than they actually did. On the other hand, learners who had recalled many high-level but only a few low-level information units might have appeared to be poorer listeners than they actually were. The logic behind this argument is that high-level information should receive more weight, as it is more important to a global comprehension of the text. A lack of differentiation of levels of information might obscure learners' actual listening proficiency when listening comprehension is not separately analyzed with respect to high- and low-level information.

Differentiation of levels of processing is also important in academic learning contexts. A number of studies show that even the advanced L2 learners have difficulty comprehending the main points of lectures (Flowerdew & Miller, 1997; Young, 1994, among others). Such learners were found to pay more attention to decoding the message, word by word, instead of attempting to comprehend the gist of the lecture (Yuan, 1982). In the same vein, teaching practitioners have asserted the significance of summarization skills at a university setting (Cohen, 1994; Sung-Ae Kim, 1998; Mayes, 1990).

Differentiation of levels of comprehension in listening performance has further implications for how we teach English for Academic Purposes (EAP) listening. Brown

¹ In light of this criterion, a summary task seems to be a more ecologically valid assessment task than a recall task since summarization skills are considered important in academic contexts (Cohen, 1994; Fletcher, 1990; Sung-Ae Kim, 1998; Mayes, 1990). Learners are less likely to encounter a situation in which they need to write down everything they can remember from a lecture (i.e., a recall task).

(2001, pp. 260-264) nicely summarizes listening goals for learners of various proficiency levels in EAP contexts: (a) For beginners: getting the gist of a sentence and recognizing a topic, (b) for intermediate learners: analyzing a discourse structure, identifying the topic, evaluating themes and motives, finding major and minor ideas, and making inferences, and (c) for advanced learners: using the lecture introduction to predict its focus, utilizing the lecture transcript to predict the upcoming information, and locating main ideas of a lecture.

VI. CONCLUSION

The goal of this study was to investigate the effects of assessment tasks (i.e., summary and recall tasks) on foreign language learners' listening comprehension. The current findings showed that task types play an important role in L2 spoken text processings. These findings that different tasks measured different levels of listening comprehension have important implications for both testing practitioners and researchers in the comprehension field. For testing practitioners, as a number of researchers asserted, one should note that the learner's comprehension score represents complexity in an assessment task performance and it is affected by the characteristics of the particular task to a large degree (Bachman, 1991; Chalhoub-Deville, 1995; Douglas, 1998; Shohamy, 1997; Yi'an, 1998). Language testing practitioners should be aware of this fact and need to exercise caution in both interpreting the learner's comprehension score and developing assessment tasks.

For researchers in the comprehension field, as Berne (1992) and Wolf (1993) argued, the results of studies which employed different assessment tasks may not be comparable because assessment task types themselves are an influential factor in affecting the outcomes of the studies. The lack of comparability and generalizability of studies in L2 listening comprehension may make it difficult to synthesize the empirical findings into any general statements about comprehension.

Despite the implications of the current study discussed so far, the current study is not without limitations. Care must be taken in generalizing the present findings. This study calls for a more in-depth analysis of L2 learners' listening comprehension performances on a larger scale. The present study compared the two types of production tasks (i.e., recall vs. summary). Future studies are recommended to compare the use of recognition and production tasks, such as a multiple-choice question task versus a recall or a summary task, to measure learners' L2 listening comprehension. Such research would enable us to examine how these tasks measure different levels of listening comprehension. Further research is clearly in order to gain a deeper understanding of the role of assessment task types in L2 oral text processing.

REFERENCES

- Alderson, C. (2000). *Reading assessment*. New York: Cambridge University Press.
- Bachman, L. F. (1991). What does language testing have to offer? *TESOL Quarterly*, 25, 671-704.
- Bentley, S., & Bacon, S. E. (1996). The all new, state-of-the-art ILA definition of listening: Now that we have it, what do we do with it? *Listening Post*, 56, 1-5.
- Berne, J. (1992). *The effects of test type, assessment task, and target language experience on foreign language learners' performance on listening comprehension*. Unpublished doctoral dissertation. University of Illinois, Urbana-Champaign.
- Bernhardt, E. B. (1991). *Reading development in a second language: Theoretical, empirical and classroom perspectives*. Norwood, NJ: Ablex.
- Bernhardt, E. B., & Deville, C. (1991). Testing in foreign language programs and testing programs in foreign language departments: Reflections and recommendations. In R. V. Teschner (Ed.), *Issues in language program direction: Assessing foreign language proficiency of undergraduate* (pp. 43-59). Boston, MA: Heinle & Heinle.
- Brindley, G. (1998). Assessing listening abilities. *Annual Review of Applied Linguistics*, 18, 171-191.
- Buck, G. (1997). The testing of listening in a second language. In C. Clapham & D. Corson (Eds.), *Encyclopedia of language and education* (Vol. 7, pp. 65-74). Dordrecht, Netherlands: Kluwer Academic Publishers.
- Brown, H. (2001). *Teaching by principles: An interactive approach to language pedagogy*. New York: Longman.
- Carrier, K. (1999). *The learner-interlocutor relationship as a factor in second language listening comprehension: A sociolinguistic perspective*. Unpublished doctoral dissertation. University of Pennsylvania, Philadelphia.
- Chalhoub-Deville, M. (1995). A contextualized approach to describing oral language proficiency. *Language Learning*, 45, 251-281.
- Chen Q., & Donin, J. (1997). Discourse processing of first and second language biology texts: Effects of language proficiency and domain-specific knowledge. *Modern Language Journal*, 81, 209-227.
- Cohen, A. (1994). English for academic purposes in Brazil: The use of summary tasks. In H. Clifford & K. Parry (Eds.), *Testing to assessment: English as an international language* (pp. 174-204). London: Longman.
- Douglas, D. (1998). Testing methods in context-based second language research. In L. F. Bachman & A. Cohen (Eds.), *Interfaces between second language acquisition and language testing research* (pp. 141-155). New York: Cambridge University Press.
- Fletcher, M. B. (1990). *Effects of text structure and text difficulty on summary writing*. Unpublished doctoral dissertation. Harvard University, Boston.

- Flowerdew, J., & Miller, J. (1997). The teaching of academic listening comprehension and the question of authenticity. *English for Specific Purposes*, 16, 27-46.
- Flowerdew, J., & Tauroza, S. (1995). The effect of discourse markers on second language lecture comprehension. *Studies in Second Language Acquisition*, 17, 435-458.
- Glenn, E. (1989). Content analysis of fifty definitions of listening. *Journal of the International Listening Association*, 3, 21-31.
- Goldman, S. R., Saul, E. U., & Cote, N. (1995). Paraphrasing, reader, and task effects on discourse comprehension. *Discourse Processes*, 20, 273-305
- Horiba, Y. (2000). Reader control in reading: Effects of language competence, text type, and task. *Discourse Processes*, 29, 223-267.
- James, C. J. (1986a). Listening and learning: Protocols and processes. In B. Snyder (Ed.), *Second language acquisition: Preparing for tomorrow* (pp. 38-45). Lincolnwood, IL: National Textbook.
- James, C. J. (1986b). *From principles to protocols: The contributions of John B. Carroll and Robert Lado to the testing of listening comprehension*. Quiryat Anavim, Israel: Language Testing Symposium. (ERIC Document Reproduction Service No. ED 278 217).
- Jamieson, J., Campbell, J., Norfleet, L., & Berbisada, N. (1993). Reliability of a computerized scoring routine for an open-ended task. *System*, 21, 305-322.
- Kim, Sung-Ae. (1998). An investigation of the relationship between L2 proficiency and summarization skills. *Journal of The Applied Linguistics Association of Korea*, 14, 95-125.
- Lorch, R. F., Klusewitz, M. A., & Lorch, E. P. (1995). Distinctions among reading situations. In R. F. Lorch & E. J. O'Brien (Eds.), *Sources of coherence in reading* (pp. 375-398). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Lynch, T. (1998). Theoretical perspectives on listening. *Annual Review of Applied Linguistics*, 18, 3-19.
- Mayes, A. M. (1990). An analysis of summary protocols of university ESL students. *Applied Linguistics*, 11, 253-271.
- Mendelsohn, D. (1998). Teaching listening. *Annual Review of Applied Linguistics*, 18, 81-101.
- Mills, C. B., Diehl, V. A., Birkmire, D. P., & Mou, L. (1995). Reading procedural texts: Effects of purpose for reading and predictions of reading comprehension models. *Discourse Processes*, 20, 79-107.
- Morley, J. (1995). Academic listening comprehension instruction: Models, principles, and practices. In D. Mendelsohn & J. Rubin (Eds.), *A guide for the teaching of second language listening* (pp. 151-165). San Diego: Dominic Press.
- Riley, G. L., & Lee, J. F. (1996). A comparison of recall and summary protocols as measures of second language reading comprehension. *Language Testing*, 13, 175-187.

- Rinehart, S. D., & Thomas, K. F. (1993). Summarization ability and text recall by novice studiers. *Reading Research and Instruction, 32*, 24-32.
- Roebuck, R. (1998). *Reading and recall in L1 and L2: A sociocultural approach*. Stamford, CT: Ablex Publishing Corporation.
- Rubin, J. (1994). A review of second language listening comprehension research. *Modern Language Journal, 78*, 196-204.
- Scott, M. L., Stansfield, C. W., & Kenyon, D. M. (1996). Examining validity in a performance test: The listening summary translation exam (LSTE)—Spanish version. *Language Testing, 13*, 83-109.
- Shohamy, E. (1997). Testing methods, testing consequences: Are they ethical? Are they fair? *Language Testing, 14*, 340-349.
- Shohamy, E., & Inbar, O. (1991). Validation of listening comprehension test: The effects of text and question type. *Language Testing, 8*, 23-40.
- Thompson, I. (1995). Testing listening comprehension. *AATSEEL Newsletter, 37*(5), 24-31.
- Wolf, D. W. (1993). Comparative assessment tasks used to measure FL reading comprehension. *Modern Language Journal, 77*, 473-489.
- Yi'an, W. (1998). What do tests of listening comprehension test?—A retrospection study of EFL test-takers performing a multiple-choice task. *Language Testing, 15*, 21-44.
- Young, L. (1994). University lectures—macro-structure and micro-features. In J. Flowerdew (Ed.), *Academic listening: Research perspectives* (pp. 159-176). New York: Cambridge University Press.
- Yuan, D. (1982). *Chinese scientists' difficulties in comprehending English science lectures*. Unpublished master's thesis. University of California, Los Angeles.

Applicable level: adult English education

Key words: measurements, recall task, summary task

Euen Hyuk (Sarah) Jung
 Department of English Language and Literature
 Yonsei University
 134 Sinchon-dong Seodaemun-gu Seoul, South Korea
 E-mail: junge@yonsei.ac.kr

Received in May, 2007

Reviewed in June, 2007

Revised version received in August, 2007