

# An Investigation of Cognitive Operations on L2 Listening Comprehension Performance: An Exploratory Study

Hui-fang Shang, I-Shou University, Taiwan

## Abstract

The purpose of the paper was to investigate the interactions and differences among components of cognitive operations on EFL listeners' comprehension performance. More specifically, components of the cognitive operations in terms of interpreting main ideas, identifying details, and interpreting implications were the focus of the study to measure 63 EFL sophomores' listening comprehension competence at one university in Taiwan. Three main questions were addressed: (1) What is the most difficult cognitive operation in terms of interpreting main ideas, identifying details, and interpreting implications? (2) For different proficiency listeners, what is their listening performance in terms of those three cognitive operations? (3) How are listeners' perceptions regarding their listening performance consistent with the quantitative research results? By applying a qualitative and quantitative research method, results of the study demonstrate that subjects perform better on items referring to main idea questions than on items referring to inference ones. Detail questions seem to affect the low listeners' performance differently. Implications of the results for the selection of test types and question types on listening comprehension tests are drawn to arrive at highly valid listening comprehension tests.

## Introduction

Listening is a critical element in the competent language performance of adult EFL learners. As language teaching has moved toward comprehension-based approaches, at the same time, many scholars and teachers get involved in constructing a valid listening test to evaluate EFL learners' comprehension performance. Although a growing number of evidence suggests that tests of L2 listening comprehension can demonstrate construct validity under certain conditions (Buck, 1992; Dandonoli & Henning, 1990), the definition of construct validity has been a controversial issue for some time (Buch, 1991). There exists neither research specific to second language listening (Richards, 1983), nor general consensus on the best techniques for assessing that construct (Henning, 1992). However, when attempting to develop a L2 listening comprehension test, it is essentially important to understand the conceptual framework of listening comprehension construct and identify the critical aspects of listening processes, including how listening functions. In the following, the listener comprehension processes which focus on how listener functions to affect listening comprehension performance are presented.

## Literature Review

The definition of construct validity has been a controversial issue for a long time because some scholars consider the various types of validity – content, criterion, and construct – as each constituting a separate type of validity (Bachman, 1990; Messick, 1988). To obtain evidence for construct validity requires continuous testing of hypotheses about a test (Shohamy, 1991). The other controversial issue regarding constructing a valid listening test is the lack of generally agreed upon proficiency guidelines for “listening comprehension” (Witkin, 1990). According to Shohamy (1991), listening comprehension can be characterized “by the need for simultaneous interpretation since in most situations the listener is denied the option of reviewing and reconsidering the information presented. The listener, therefore, must rely on

immediate comprehension, and on the ability to retain the information in the memory for further analysis” (p. 24). Duzer (1997) also proposed several basic processes at work in listening. For example, the listener takes the raw speech and deposits an image of it in the short-term memory; predicts information expected to be included in the message; recalls background information (schemata) to help interpret the message; and determines the information to be held in the long-term memory.

Current approaches view the comprehension process as an interaction between the listener’s background knowledge and text. Two types of listening comprehension processing are occurring: bottom-up processing (deriving meaning of the message based on the incoming language data, from sounds to words, to grammatical relationship, to meaning) and top-down processing (utilizing background knowledge to derive meaning from and interpret the message) (Duzer, 1997). Van Dijk and Kintsch (1983) identified such two types of listening comprehension processing as local strategies (requiring attention to relations and links between local clauses and sentences) and global strategies (focusing on the overall coherence of a text). Shohamy (1991) further divided local strategies into subcategory as trivial strategies (focusing on numerical details, not directly related to the main topic). A number of studies focus on the manner in which L2 listeners deal with text processing: Hildyard and Olson (1982) found that efficient listeners utilize background knowledge as an interactive base for text processing, while low level students relate mostly to local details. They found that efficient listeners are so successful at using their background knowledge to interpret the new text that they do not successfully distinguish between the original text and implications they draw. Wolff’s (1987) study also indicated that data-driven processing is only fragmentary for efficient L2 listeners; instead, they will activate more L1 knowledge in the form of concept-driven schemata. O’Malley, Chamot and Kupper (1989) found out that effective L2 listeners use both top-down and bottom-up strategies to construct meaning, while ineffective listeners just determine the meanings of individual words. Shohamy’s (1991) study also showed that subjects with low proficiency level perform better on items referring to local cues than on items referring to global ones. Jensen and Hansen (1995) pointed out that it doesn’t mean that the low proficiency listeners do not use content schemata for global understanding. However, they may not select appropriate schemata. “Selecting appropriate schemata depends on having a successful and somewhat automatic interaction between the input, linguistic knowledge and world knowledge to construct the larger units of meaning and to comprehend the discourse” (p. 102). On the other hand, the high proficiency listeners can successfully “decode the input, interpret the semantic content and integrate the new information into his or her own knowledge system” (p. 102), in order to comprehend a message.

Another view for listening processes is based on Lund’s (1990) taxonomy: Listeners must possess six listening functions in order to comprehend a message. Those six functions are identification, orientation, main idea comprehension, detail comprehension, full comprehension and replication. Dunkel, Henning and Chaudron (1993) then proposed a tentative model with four similar categories for comprehending a listening text. Those main categories of text meanings are orientation meanings (persons and their relationships and topic of the text), detail meanings (simple lexical meanings and single prepositional meanings in the text), main ideas (the derived principal proposition of a text), and implications (meanings derived from the listener’s application of background knowledge and logic to the text). Dunkel, Henning and Chaudron called those components of text meaning as “cognitive tasks” (p. 186). To measure listening comprehension, they considered “identification and interpretation as defensible, necessary, and sufficient examples of comprehension” and they regarded such measurements as “cognitive operations (p. 186).”

## **Purpose of the Study**

Although a number of theories have demonstrated about how listening functions and how listening comprehension processes, yet there are substantially no experimental researches concerned about the interactions among components of cognitive operations on different proficiency listeners' performance. Therefore, the purpose of the present study was to use Dunkel, Henning and Chaudron's (1993) theory as a model to specify the interactions and differences among components of cognitive operations on EFL listeners' comprehension performance. More specifically, selected components of the cognitive operations in terms of interpreting main ideas, identifying details, and interpreting implications were the focus of the study to measure EFL listeners' comprehension competence. In other words, the purpose of such listening assessment was to evaluate whether EFL listeners with different proficiency levels could distinguish successfully among cognitive tasks on major, literal, and implied meanings for each listening text. To examine listening comprehension competence, three different cognitive operations in terms of detecting major meanings from different grammatical forms or sentence types (interpreting main ideas), as well as distinguishing between literal meanings (identifying details) and implied meanings (interpreting implications) with different proficiency listeners were proposed and presented.

To explore the above issues, the following questions were concerned:

1. What is the most difficult cognitive operation in terms of interpreting main ideas, identifying details, and interpreting implications?
2. For different proficiency listeners, what is their listening performance in terms of those three cognitive operations?
3. How are listeners' perceptions regarding their listening performance consistent with the quantitative research results?

To probe the questions above, subjects' listening comprehension performance with three proficiency levels was calculated by computing a one-way ANOVA analysis to investigate their mean differences among those three cognitive operations. A Post Hoc test was further employed to examine which cognitive operation yielded higher listening performance for different proficiency listeners on a listening comprehension test. Besides, a semi-standardized questionnaire consisting of 15 questions was administered to the EFL listeners to assess their perceptions of the difficulty levels for each cognitive operation.

## **Methodology**

### *Subjects*

The subjects in the study were 63 sophomores at Applied English Department enrolled in an intermediate listening class at I-Shou University in Taiwan. The subjects ranged from 18 to 25 years of age, with a mean age of 19.6. A demographic questionnaire was administered to gather information about the subjects' backgrounds. Results from the questionnaire showed that all of the subjects had experienced formal instruction in English for an average of 6 years by the time they came to study at I-Shou University. 81% of the subjects did various kinds of practices to improve their English listening proficiency in their free time, such as listening to English songs and radio programs, watching western movies and CNN news, etc. However, 15% of them didn't do any practice at all.

Subjects were then divided into three proficiency groups – low, intermediate, and high -- on the basis of their scores on the simulated TOEFL listening test taken from the previous semester. The test results ranged from 6 to 13 out of 13, with the mean of

Table 1 Number of Subjects and Scores for Each Proficiency Group

	Low	Intermediate	High
Scores	6~9	10	11~13
Number	13	20	30
(% of total sample)	(21%)	(32%)	(47%)
Total	63		

10.48 and with the median of 10. Those subjects whose scores ranged from 11 to 13 were labeled “high”, while those whose scores ranged from 6 to 9 were labeled “low”. The subjects with scores of 10 were labeled “intermediate”. The number of subjects for each proficiency group was given in Table 1.

Choosing appropriate listening materials and items for the research purpose is at present largely a subjective process. In this study, a textbook of “The Heinle & Heinle TOEFL Test Assistant: Listening” (Broukal, 1995) was selected to provide listening questions of various situational and conversational contexts. To ensure that subjects were quite familiar with the content and topic, three different listening texts with the similar background of campus events were chosen for the research purpose. Based on Dunkel, Henning, and Chaudron (1993), many leveling variables such as text type, text difficulty, lexical complexity, content imagery, cultural proximity, organization, clarity, rate of speech, topic familiarity, etc. may influence listening comprehension competence. In this study, as a result, text type, content imagery, clarity, rate of speech, topic familiarity were tried to be controlled by choosing the similar listening texts from the same book. The language used in the recording text which was attached with the textbook was implicit and fragmented. There were many repetitions, redundancies, interruptions, pauses, etc.

In that textbook, three different listening tasks containing “paraphrasing,” “listening for details,” and “making inferences” were selected on the basis of the present research purpose to measure listeners’ comprehension competence in terms of using local, trivial, and global strategies in listening comprehension processes. The purpose of using “paraphrasing” was to test listeners’ understanding of the meaning of the main ideas presented. The listeners had to choose the similar meaning of the major ideas with other words or statements. As to “listening for details” section, listeners had to understand the whole conversation to answer detail questions about the speakers and the subject of the conversation. In regard to “making inferences,” listeners had to draw conclusions from the information given by the speakers. The answers to inference questions were not directly stated in the conversation.

### *Procedures*

Subjects were tested simultaneously in the language laboratory. For the measure of listening comprehension, 12 items were constructed based on the components of each listening task. The items for “paraphrasing,” “listening for details,” and “making inferences” were 4, 3, and 5, respectively (see Appendix A). The questions in each task were played twice and the subjects were allowed to take notes while listening and answering the questions. Each listening task lasted for about 4 minutes and consisted of the text type of extended conversations with the constructed item type of multiple choices. The duration of each measure was about 30 minutes. Subjects were asked to answer the questions by paper-and-pen after listening to each passage. In this study, two scoring methods were used as follows: binary (correct/incorrect) of item scores and rating scale (a self-perceived survey regarding the difficulty levels for each task) (see Appendix B). Subjects would get 1 point if they chose the correct answer. After finishing the measure, subjects were asked

to complete a questionnaire that included a self-rating of their listening performance for each listening task.

### *Measurement Instruments and Data Analyses*

Subjects' performance with three different proficiency levels on the three assigned cognitive listening tasks was examined to determine their interactions and mean differences. The three main questions were taken into considerations as followed:

1. What is the most difficult cognitive operation in terms of interpreting main ideas, identifying details, and interpreting implications?
2. For different proficiency listeners, what is their listening performance in terms of those three cognitive operations?
3. How are listeners' perceptions regarding their listening performance consistent with the quantitative research results?

Accordingly, the following hypotheses were formulated in relation to the research questions:

1. The cognitive operation related to interpreting implications is the most difficult question type because those three proficiency listeners, especially the low proficiency listeners, have to use both local and global strategies successfully.
2. The high proficiency listeners can successfully use both bottom-up and top-down processing to determine major, literal, and implied meanings than the intermediate and low proficiency listeners.
3. The intermediate listeners' performance regarding those three cognitive operations will be better than the low listeners'.
4. The low proficiency listeners will perform better on items referring to detail questions than on items referring to main idea and inference questions.
5. Listeners' perceptions will be consistent with the experimental research results.

To explore the hypotheses above, subjects' listening comprehension performance with three proficiency levels (independent variables) was calculated by computing a one-way ANOVA analysis to investigate their mean differences among those three cognitive operations (dependent variables). A Post Hoc test was further employed to examine which cognitive operation yielded higher listening performance for different proficiency listeners. Besides, a self-rating questionnaire (see Appendix B) consisting of 15 questions was administered to the subjects to probe their perceptions of the difficulty level of each cognitive operation. The percentage based on the subjects' answers was calculated by the technique of frequency.

## **Results**

In order to test the first hypothesis, listeners' mean scores regarding the relative difficulty level of the cognitive operations was maintained in each of the three groups. Results from Table 2 reveal that the means of detail (trivial) items for those three proficiency groups constitute the lowest scores. It is assumed that the trivial questions are the most difficult ones for listeners in the low, intermediate, and high groups. This also suggests that each group performs worst on the detail questions. Such finding doesn't support the first hypothesis that the question type of interpreting implications is the most difficult one for those three groups, nor support the fourth hypothesis that the low groups will perform better on the detail questions than on the main idea and inference questions.

Since the trivial (detail) questions seem to be problematic and present inconsistent findings, it was decided to conduct the statistical comparison of means between the local (main idea) and global (inference) questions only. Such findings show that the means for the local questions are higher than the global ones except for the low group. (Table 3 shows the significant differences at the 0.05 level for those three groups in the local and global questions.)

In regard to test the second and third hypotheses, a one-way analysis of variance was computed to examine if there was a significant difference for those three cognitive operations on those three proficiency groups. Results in Table 4 indicate that there is a significant difference at the .001 level for those three operations.

More analyses of the data on those three question types was conducted in order to gain insight as to the listening performance employed by levels of listeners.

Table 2 Classification of Scores Obtained according to Three Question Types for Three Proficiency Groups

	Low Mean (%)	Intermediate Mean (%)	High Mean (%)
Main Idea (local)	67.3	73.8	80.8
Detail (trivial)	51.3	46.7	63.3
Inference (global)	78.5	70	79.3

Table 3 A T-Test Analysis between Local and Global Questions among Three Groups

	Mean	N	SD	S.E. Mean	t	df	Sig.
Low							
Local	67.3	13	.751	.208			
Global	78.5	13	.641	.178	-4.382	12	.001
Intermediate							
Local	73.8	20	.686	.154			
Global	70	20	.946	.212	-2.463	19	.024
High							
Local	80.8	30	.898	.164			
Global	79.3	30	.890	.163	-4.097	29	.000

Table 4 A One-way ANOVA Analysis for Those Three Proficiency Groups

Score	SS	df	MS	F	Sig.
Between Groups	148.265	2	74.132	97.342	.000
Within Groups	141.651	186	.762		
Total	289.915	188			

Findings in Table 5a show that the high group outperforms the intermediate and low groups in those three cognitive tasks because the high group receives the highest mean scores among those three cognitive operations. However, in detail and inference sections, the low group outperforms the intermediate one. Although there is only one significant difference (.032) between the low and intermediate groups in the inference section (see Table 5b), the descriptive findings point to a hierarchical order of difficulty in the detail and inference sections: The intermediate proficiency group obtains lower mean scores which present a greater degree of difficulty than the low group. Such results show that the intermediate group doesn't perform better than the low group in the detail and inference sections.

As to examine the consistency between the qualitative and quantitative research results, further analyses of the data from listeners' self-report (see Table 6a, 6b, and 6c) demonstrate a hierarchical order of difficulty for those three groups: The low group considers those three question types as the most difficult ones, followed by the high group, and then followed by the intermediate group. It is assumed that the low group should obtain the lowest mean scores on those three question types. However,

Table 5a A Post Hoc Test among Low, Intermediate, and High Groups in Three Cognitive Tasks

	Group	N	Mean	SD
Main Idea	Low	13	2.692	.751
	Intermediate	20	2.950	.686
	High	30	3.233	.898
Detail	Low	13	1.539	.776
	Intermediate	20	1.400	.940
	High	30	1.900	.923
Inference	Low	13	3.923	.641
	Intermediate	20	3.500	.946
	High	30	3.967	.890

Table 5b Significance between means in Three Groups among Cognitive Operations

	F	Sig.	
Main Idea	Low/Intermediate	1.115	.299
	Intermediate/High	2.679	.108
	Low/High	.230	.634
Detail	Low/Intermediate	.641	.430
	Intermediate/High	.373	.544
	Low/High	.025	.875
Inference	Low/Intermediate	5.050	.032*
	Intermediate/High	.905	.346
	Low/High	1.432	.238

from Table 5a, it is clear that the low group receives higher mean scores than the intermediate group in the detail and inference sections. As a result, such qualitative finding isn't completely consistent with the research results. Another finding shows that even though the high group obtains the highest mean scores on those three question types (see Table 5a), the high group expresses a greater degree of difficulty, comparing with the intermediate group. It is concluded that such finding isn't consistent with the research result, either.

*Table 6a Self-Report of Difficulty Level for Main Idea Questions*

	Very easy	Easy	OK	Difficult	Very difficult
Low	0	0	8 (66.7%)	3 (25%)	1 (8.3%)
Intermediate	0	0	16 (84.2%)	3 (15.8%)	0
High	2 (4%)	2 (4%)	36 (72%)	10 (20%)	0

*Table 6b Self-Report of Difficulty Level for Detail Questions*

	Very easy	Easy	OK	Difficult	Very difficult
Low	0	0	7 (53.8%)	5 (38.5%)	1 (7.7%)
Intermediate	0	0	12 (66.7%)	6 (33.3%)	0
High	0	1 (2%)	31 (63.3%)	17 (34.7%)	0

*Table 6c Self-Report of Difficulty Level for Inference Questions*

	Very easy	Easy	OK	Difficult	Very difficult
Low	0	2 (15.4%)	6 (46.2%)	4 (30.8%)	1 (7.7%)
Intermediate	0	0	16 (84.2%)	3 (15.8%)	0
High	1 (2%)	3 (6.1%)	32 (65.3%)	13 (26.5%)	0

According to findings from the previous analysis (see Table 2), listeners in the low group receive the highest mean scores in the inference questions, followed by the main idea ones and then followed by the detail questions. The findings show that the detail questions are the most difficult ones, followed by the main idea and then followed by the inference ones. However, based on Table 7a, the low group expresses that the main idea questions are the least difficult ones. Such finding isn't consistent with the previous research result (see Table 2) that the inference questions are the least difficult ones. Another result (see Table 7a, 7b, and 7c) shows that those three groups point out that detail and inference questions are quite difficult in comprehending the listening texts. Such result partly supports the first hypothesis that the cognitive operation of interpreting implications is the most difficult question type.

*Table 7a Frequency of Three Question Types for Low Proficiency Group*

	Very easy	Easy	OK	Difficult	Very difficult
Main Idea	0	0	8 (66.7%)	3 (25%)	1 (8.3%)
Detail	0	0	7 (53.8%)	5 (38.5%)	1 (7.7%)
Inference	0	0	7 (53.8%)	5 (38.5%)	1 (7.7%)

*Table 7b Frequency of Three Question Types for Intermediate Proficiency Group*

	Very easy	Easy	OK	Difficult	Very difficult
Main Idea	0	0	16 (84.2%)	3 (15.8%)	0
Detail	0	0	12 (66.7%)	6 (33.3%)	0
Inference	0	0	12 (66.7%)	6 (33.3%)	0

*Table 7c Frequency of Three Question Types for High Proficiency Group*

	Very easy	Easy	OK	Difficult	Very difficult
Main Idea	2 (4%)	2 (4%)	36 (72%)	10 (20%)	0
Detail	0	1 (2%)	31 (63.3%)	17 (34.7%)	0
Inference	0	1 (2%)	31 (63.3%)	17 (34.7%)	0

## Discussion

The first finding of this study is that each group performs worst on the trivial (detail) questions. Based on listeners' self-reports toward the difficulty and problems of answering the detail questions correctly, most of them express that it's difficult to memorize all the information from the listening texts. The demand to recall details may have caused confusion. For example, in the listening text, the speaker discoursed many course names and different time periods. The listeners predicted that the questions would be asked something about numerical details in the "listening for details" section. As a result, the listeners just concentrated on memorizing those insignificant details, instead of emphasizing on the whole comprehension. Such severe demands on the listeners' memory load may interfere with attending to the more relevant and important tasks in the text.

Findings regarding the comparison of local and global questions show that the subjects, except for the low group, perform better on items referring to the local questions

than on the global ones. Such result seems to suggest that questions classified based on different levels of cognitive processing yield different levels of performance; that is, items which focus on comprehension of local information are more attainable than items which relate to macro information in the text (Shohamy, 1991). These results may be applied to schema theory which distinguishes between local and global strategies for comprehension (Van Dijk & Kintsch, 1983). It seems more difficult to employ the concept to generalize, infer, and synthesize the information from the cognitive processing than to get the main idea information only.

A number of studies indicate that low level listeners relate mostly to local details and they just determine the meanings of individual words, so they perform better on items referring to local cues than on items referring to global ones (Hildyard & Olson, 1982; O'Malley, Chamot & Kupper, 1989; Shohamy, 1991). However, according to the findings of the present study, the low proficiency listeners perform better on the global items than on the local ones. Comparing the self-report from the low listeners, they express that the local items are less difficult than the global ones. It seems that there is an inconsistency from the quantitative and qualitative research results. Regarding the inconsistency, it can be hypothesized that the low listeners guess the answers mostly while taking the listening test since it's a multiple-choice test. Yet in their perceptions, the low listeners still consider that the inference questions are the most difficult question type because it's not easy to draw conclusions from the information given by the speakers. The similar situation of inconsistency also appears to the high listeners. Even though they perform best on those three cognitive operations, the high listeners still have no confidence in answering the inference items correctly, so they feel that the inference items are quite difficult.

Findings regarding the listening performance between the low and intermediate listeners indicate that the low group outperforms the intermediate group in the detail and inference sections. It may be speculated that the listening texts which were used to affect the degree of listening comprehension are relatively difficult to the intermediate group as opposed to the other.

## Implications

The results obtained in this study have direct implications to the construct validity of listening comprehension tests. As was noted in the beginning of the paper, obtaining evidence for construct validity has been a controversial issue and it requires continuous testing of hypotheses about a test (Shohamy, 1991). The hypotheses tested

here relate to the type of cognitive tasks which best tap listening comprehension competence. The results provide information which indicates that different types of cognitive operations result in different performances for different proficiency listeners. Basically, local questions are easier than global ones; although it's clear that other factors, such as familiarity with the topic, background knowledge, grammatical formality and complexity, text density, etc. may affect listeners' performance. Further research should look at the influence of those factors on the listening comprehension performance. Clearly, a valid listening comprehension test should attempt to control those factors and include both local and global question types to test listeners' bottom-up and top-down cognitive processing.

Similarly, to test listening comprehension, a valid listening comprehension test should also consider the test type. From the present study, it's apparent to figure out that a multiple-choice test allows listeners to guess the answers easily; it's difficult to evaluate listeners' real comprehension ability. As a result, using more test types that include blank filling or short answers should certainly reflect listeners' listening comprehension competence.

Furthermore, trivial questions which relate to irrelevant recall of names or numerical data, may distract listeners' attention on the whole comprehension. Such cognitive operation serves no meaningful purpose as evaluating listening comprehension. Therefore, it's recommended to avoid the insignificant numerical details on listening comprehension tests.

In conclusion, evaluating listening comprehension is a challenge. It's a challenge to demand a variety of test types and question types. To result in more construct valid listening comprehension tests, teachers should carefully select a variety of testing instruments, in order to better reflect the trait of L2 listening comprehension.

## References

- Bachman, L. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Broukal, M. (1995). *The Heinle & Heinle TOEFL test assistant: Listening*. Boston, MA: Heinle & Heinle Publishers.
- Buck, G. (1991). The testing of listening comprehension: An introspective study. *Language Testing*, 8(1), 67-91.
- Buck, G. (1992). Listening comprehension construct validity and trait characteristics. *Language Learning*, 42, 313-357.
- Dandonoli, P., & Henning, G. (1990). An investigation of the construct validity of the ACTFL proficiency guidelines and oral interview procedure. *Foreign Language Annals*, 23, 11-22.
- Dunkel, P., Henning, G., & Chaudron, C. (1993). The assessment of an L2 listening comprehension construct: A tentative model for test specification and development. *The Modern Language Journal*, 77, 180-191.
- Duzer, C. V. (1997). Improving ESL learners' listening skills: At the workplace and beyond. Project in Adult Immigrant Education, National Center for ESL Literacy Education. Retrieved October 19, 2003, from <http://www.cal.org/nclle/DIGESTS/LISTENQA.HTM>
- Henning, G. (1992). Analysis of a content-word-recall approach to the testing of listening and reading comprehension. In R. Courchene, J. Glidden, J. St. John, & C. Therien (Eds.), *Comprehension-based second language teaching* (pp. 339-349). Ottawa, Ontario: University of Ottawa.
- Hildyard, A., & Olson, D. (1982). On the comprehension and memory of oral versus written discourse. In D. Tannen (Ed.), *Spoken and written language* (pp. 19-32). Norwood, NJ: Ablex Publishing.
- Jensen, C., & Hansen, C. (1995). The effect of prior knowledge on EAP listening-test performance. *Language Testing*, 12(1), 99-119.
- Lund, R. J. (1990). A taxonomy for teaching second language listening. *Foreign Language Annals*, 23(2), 105-115.
- Messick, S. (1988). Meaning and values in test validation: The science and ethics of assessment. Paper presented at the AERA Convention, New Orleans.
- O'Malley, J. M., Chamot, A. U., & Kupper, L. (1989). Listening comprehension strategies in

- second language acquisition. *Applied Linguistics*, 10, 418-437.
- Richards, J. C. (1983). Listening comprehension: Approach, design, procedure. *TESOL Quarterly*, 17, 219-240.
- Shohamy, E. (1991). Validation of listening comprehension tests: The effect of text and question type. *Language Testing*, 8(1), 23-40. 16
- Van Dijk, T., & Kintsh, W. (1983). *Strategies of discourse comprehension*. N.Y.: Academic Press.
- Witkin, B. (1990). Listening theory and research: The state of the art. *Journal of the International Listening Association*, 4, 7-32.
- Wolff, D. (1987). Some assumptions about second language text comprehension. *Studies in Second Language Acquisition*, 9, 307-326.

## Appendices

### Appendix A Three Question Types with the Tapescript

A: Write down the letter of the sentence that correctly rephrases that section of the conversation.

1. (A) The woman doesn't mind that the cafeteria is crowded.  
(B) The woman wants the student to know that all the new tables are crowded.  
(C) The woman doesn't object to having the new student sit at her table.
2. (A) Neither the man nor the woman knows how to read the computerized class schedule.  
(B) The woman understand the new student's confusion and is willing to help him.  
(C) The man want to show the woman the way around campus, but he can't read his schedule.
3. (A) Although the man got the time of his physics class right, he had trouble with his location.  
(B) The man was late for his world affairs class because he went to the Science Building.  
(C) By the time the man got to the Science Building, he realized he should have been managing his schedule better.
4. (A) The man confused the number of the building with the time of his class.  
(B) The man was confused because he had three classes in the same building.  
(C) The man confused the woman because he was supposed to be in another class in the same building at three o'clock.

B: Write the correct answer about detail questions.

5. What is the man's first problem?  
(A) He can't register for the European Literature class he wanted.  
(B) He has signed up for too many classes.  
(C) All of the classes he has signed up for are not available.  
(D) He has selected two classes being given at the same time.
6. According to the conversation, why wasn't the man's schedule of classes accurate?  
(A) The schedule was printed incorrectly.  
(B) A change had been made that did not appear on the schedule.  
(C) The schedule was from the previous semester.  
(D) Required courses were not included on the schedule.
7. According to the conversation, with which of the following courses is there not a registration problem?  
(A) Physics  
(B) Biology  
(C) Introduction to Psychology  
(D) American Literature

C: Write down the correct answer about inference questions.

8. Where is this conversation probably taking place?  
(A) In a classroom  
(B) In a lecture hall  
(C) In a library  
(D) In a dormitory
9. What is the probable relationship between the two speakers?  
(A) Two classmates  
(B) A counselor and a student

- (C) Two office coworkers  
 (D) A lecture and an attendee
10. What is the woman's attitude toward the man's choice of subject for his history report?  
 (A) Indifferent  
 (B) Positive  
 (C) Negative  
 (D) Confused
11. What does the man imply about most people's idea of the United States?  
 (A) Most people think of the East and West coasts.  
 (B) Most people think of the stories about eh Wild West.  
 (C) Most people think of big Eastern cities or vast Western land tracts.  
 (D) Most people think of the National Sections and Monuments.
12. What can be inferred about when this conversation is taking place?  
 (A) In the morning  
 (B) At midday  
 (C) In the afternoon  
 (D) In the early evening

*Appendix B A Self-perceived Questionnaire*

Sex: Male \_\_\_\_\_ Female \_\_\_\_\_

Age: \_\_\_\_\_

Years of English Learning before Entering ISU: \_\_\_\_\_

Previous Semester's Scores on "Listening & Conversation": \_\_\_\_\_

Do you do any practices to improve your English listening ability in your free time?

Yes \_\_\_\_\_ No \_\_\_\_\_ If yes, what are they?

1. How do you think of using the simulated TOEFL items to evaluate (test) your listening ability?

1                      2                      3                      4                      5  
 very effective      effective      no opinion      ineffective      very ineffective

2. How do you feel the difficulty for "Paraphrasing Section"?

1                      2                      3                      4                      5  
 very easy            easy            OK            difficult      very difficult

3. Do you have any problems in doing this section? Yes \_\_\_\_\_ No \_\_\_\_\_

4. If yes, what are they?

5. How do you feel the difficulty for "Detail Section"?

1                      2                      3                      4                      5  
 very easy            easy            OK            difficult      very difficult

6. Do you have any problems in doing this section? Yes \_\_\_\_\_ No \_\_\_\_\_

7. If yes, what are they?

8. How do you feel the difficulty for "Inference Section"?

1                      2                      3                      4                      5  
 very easy            easy            OK            difficult      very difficult

9. Do you have any problem in doing this section? Yes \_\_\_\_\_ No \_\_\_\_\_

10. If yes, what are they?

11. Overall, which section do you like to do most? Paraphrasing, Detail, or Inference Section

12. Why?

13. Overall, which section do you dislike to do most? Paraphrasing, Detail, or Inference Section

14. Why?

15. Overall, what are your problems in doing the listening comprehension test?