The Effects of Different Types of Genres and Tasks on College Students' English Listening Comprehension

Sookhyun Kim
(Ewha Womans University)

Kim, Sookhyun. (2010). The effects of different types of genres and tasks on college students' English listening comprehension. *English Language & Literature Teaching, 16*(1), 61-80.

This study investigates the effects of different types of genre and tasks on college students' English listening comprehension. 104 college students were assigned to Argumentative (AG), Causal and Evaluation (CE), and Narrative (NA) group. They report their listening comprehension through gap-filling (GF), summary (ST), and multiple-choice comprehension task (MC). Results showed that different genre groups were significantly different on overall tasks. Moreover, results from each group also presented that different mean scores on each task indicated a statistical significance. Proficiency levels, higher and lower level group, showed a significant effect for the task on listening comprehension. The results of the statistical analysis, using One-way ANOVA indicated that genre significantly affects Korean college students' listening comprehension, and implied that listening comprehension sores for each genre were also substantially different. Different genre groups resented significant different in overall listening comprehension tasks.

[L2 listening comprehension/genre difference/task difference]

I. INTRODUCTION

Listening comprehension is the most important of the language communication (listening, reading, speaking and writing) skills. Devine (1978) concluded that listening accounts for 50% of the four (writing, speaking, reading, and listening) types. Feyten (1991) stated that listening formed more than 45% of our total communication compared with speaking with 30%. Morley (1991) emphasized listening as an important ability for communication: it is impossible to communicate without understanding what has been

heard. Even though many researchers and teachers remarked on listening to learning a second language, listening comprehension is difficult for foreign language learners to develop and for teachers to assess. Previous researches related to EFL/ESL listening has investigated the following: listening processing, listening variables, listening strategies and listening instruction (Jeon & Park, 2001; Lee, 2009; Nam & Seong, 2009): factors that affect listening comprehension has been one of the most important aspect to study. Listening variables that affect listening ability include linguistic ability, background knowledge, listening strategy and, speed, among others. Park (2001) investigated linguistic ability, background knowledge and test type as listening variables. Choi (1994) selected content schema, speech modification, listening proficiency, and topic familiarity as the most important factors that affect listening comprehension. Several researchers have investigated the factors that affect listening comprehension, but previous research has generally been too broad and the results from this research have consequently been inconsistent.

In order to identify factors that affect listening comprehension, the present study focuses on several different types of texts and tests. Kobayashi (2002) adapted four different types of text: Association, Description, Causation, and Problem-Solution from Meyer's category on reading comprehension. Kobayashi (2002) also focused on different formats of a test to assess different aspects of reading comprehension. Even though participants read the same text, their reading comprehension score could be different due to the test formats. Kobayashi's results prove that different test formats measure different aspects of reading comprehension and different types of texts show different comprehension levels. Moreover, certain test formats had not shown differences from the different text organization indicating the test format was not affected by the text organization; however the other test format showed significant differences among text organizations.

Each genre has a different listening objective, information organization, speaker focus, content and rhetorical structure, and so on; consequently, each genre requires a different degree of comprehension. Carrell (1984) mentioned that the genre has a specific rhetorical organization; this means that one genre may be more difficult to comprehend or remember than another. The features of each genre: which are purpose of listening, information organization, speaker focus, content and rhetorical structure, etc. cause better understanding or interrupting of listening comprehension. In addition, it is possible to show differences in the proficiency levels. The other variable that affects listening comprehension is task based.

The purpose of my study is to investigate whether the different types of genres affect listening comprehension. The features of each genre (listening, information organization, speaker focus, content and rhetorical structure, and so on), are evaluated with a view to determining whether they cause better comprehension or interrupt with listening

comprehension. To identify the relationship between genres and listening comprehension, and genres and task based activities, variety of genre and different types of task will be chosen for this study. Moreover, the level of English listening ability is included in this study.. It seems that listening proficiency level can be another variable on listening comprehension. The study attempts to answer the following research questions:

1. What are the differences in listening comprehension according to the different types of genres?
2. What are the differences in listening comprehension by three different task types?
3. How do students perform on listening comprehension at different levels of their language proficiency?

II. LITERATURE REVIEW

Listening comprehension has been studied for a long time, but the meaning of listening comprehension is still difficult to define. Wolvin and Coakley (1985) defined listening comprehension as the recognition of sound and giving attention to finding out the meaning of the sound. Buck (2001) described listening comprehension as an active process to construct meaning. He concluded that listening comprehension is influenced by various variables. Another definition of listening proposed by Rost (2002) is that listening is a conscious activity under the direct control of the listener. Listening comprehension shares a large number of features with language comprehension, and many characteristics of listening comprehension overlap with those of reading comprehension. Buck (1992) and Bae and Bachman (1998) proved statistically that listening and reading comprehension did not totally overlap one-another; however, there was some overlap among certain aspects. Moreover, through their studies, Freedle and Kostin (1994, 1999) found that the listening sub-skill and reading sub-skills were similar.

1. Genre Difference in Listening Comprehension

The genre has been an important factor that previous research considered (Chaudron & Richards, 1986; Olsen & Huckin, 1990; Dunkel, Henning, & Chaudron, 1993; MacDonald, Badger & White, 2000; Lee, 2006). In foreign language learning, learners need to encounter different genres in order to learn the target language and accept the cultural differences (Christie, 1999). Rost (2002) defined "genre" in terms of what kind of text it is. Rost (2002) clarified narrative, descriptive, comparison and contrast, causal and evaluation, and problem-solution as genres according to the organization of the

information, the listening purpose and the speaker's focus. The organization of the information in the text influenced the difficulty or ease with which each test was understood.

MacDonald, Badger and White (2000) investigated a variety of listening text types based on Hallliday's concepts of field: topic and purpose; tenor: pattern of organization; degree of completeness and mode: patterns of interaction between the speaker and the audience and level of formality. Listening texts were selected from Sociology about population growth, food supply and urbanization, Education from actual lectures recorded in the Education Department on ethnography, Business and Media Studies from the BBC for Further Education College, and Media Studies from actual lectures recorded in the Education Department on ethnography. The levels of difficulty and relevance of materials of listening texts were analyzed and yielded different results. Each listening text provided different characteristics according to the learning situation, the level of formality, and the student's interest.

In their study, Dunkel, Henning and Chaudron (1993) emphasized the components of listening comprehension and the factors that influenced listening comprehension. Their study was designed to propose a framework and model of L2 listening comprehension assessment. According to their model of L2 listening comprehension assessment, different types of listening texts included: classified academic lectures, advertisements, debates, narrative stories, newscasts, speeches, and so on. These types were analyzed by rhetorical function, or by transactional versus interactional language focus, or by levels of formality.

Barry and Lazarte (1998) used different Spanish historical reading texts, which had different topics and different syntactic complexity levels, for English-speaking participants on the recall task. Even the level of syntactic complexity did not produce a significant effect on results, topic difference showed different results. Another study that examined different types of texts dealt with friendly letters, general articles, business letters, and newspaper articles (Allen, Bernhardt, Berry & Demel, 1988). These types of text genres followed different levels of proficiency based definitions of the American Council of Teaching Foreign Languages (ACTFL) Proficiency Guidelines. Kobayashi (2002) took Meyer's model of prose analysis (1985) which used content structure analysis to identify the text. Meyer analyzed the rhetorical relationship and coherence of the text. The Meyer's model presents association, description, causation and problem-solution from loosely-organized to tightly-organized in the degree of interconnectedness of ideas. Kobayashi also investigated association, description, causation and problem-solution, the four types of top-level rhetorical organizations, because the rhetorical relationship is at the highest level in the hierarchy. Raymond (1993) also used Meyer's top level structure and selects description, collection, causation, problem-solution and comparison in her study.

To summarize several previous researches related with different types of text on listening and reading comprehension, the research that I have mentioned in the previous paragraphs investigated the effects of different text types or genre on listening and reading comprehension, and the relationship between listening and reading comprehension and different text types. Some researchers concluded that text types affect learners' listening and reading comprehension, but others showed that there were no statistically significant differences between text types and comprehension.

2. Listening Comprehension Task

A general framework for the aspects and components of listening comprehension assessment contain the purpose of assessment, tasks, response category, scoring method, leveling variables, competence category, characteristics, cognitive operation, and socio-cultural context (Dunkel, Henning, & Chaudron, 1993). Under the tasks components, text, task and difficulty dimension include text type, text element, cognitive operation, response category, item type, sample item and scoring method. Dunkel, Henning, and Chaudron (1993) also categorized listener's response category which is another component that will be addressed in my study and included such responses as: answers questions, defines, describes, follows directions, paraphrases, selects pictures, produces pictures, and translate. These response categories were divided into a production response and a manual selection response as a measurement of listening comprehension. The production response occurred most often under the answers questions, defines, describes, follows directions paraphrases, produces picture, and translates. Answers questions, follows directions, selects pictures, and translates, often led to a manual selection response; whereas defines, describes, paraphrases, and produces pictures seldom generated a selection response. Content word recall and imitation are methods used to assess comprehension indirectly and may be categorized into other responses. Due to the item type, multiple-choice and true-false questions are placed in selection and completion, and short answers and composition are under production.

Shohamy and Inbary (1991) studied response categories and included item types and question types to assess listening comprehension. The study investigated the effects of different item types on the listening comprehension of 150 high school students who were learning English in Israel. Moreover, Park (2000) studied the effects of question representation and question types on listening comprehension. In this study, the researcher examined global and local question types with university students in Korea. According to the results, students within the lower proficiency group had lower listening comprehension results on global questions compared to local question types, but there was no significant difference in listening comprehension of global versus local question types for those in the

higher proficiency group. The results showed that even if scores between global and local items have small differences, there is no significant difference statistically.

The trend of recent research in listening comprehension is to attempt to find the best way to assess listening comprehension (Ockey, 2007). Previous research relating to response category included item type and question type to assess listening and reading comprehension and used different kinds of test items and response categories to evaluate comprehension. The results indicated that comprehension scores may depend on the item and question type and response category. Barry and Lazarte (1995) also concluded that total score; proportional recall for the total score and kenel score; and proportional recall of the essential text propositions are not correlated. Even if a student achieved high scores in total, the student could not achieve high scores in kenel scores. Previous research confirms that to assess listening comprehension, it is impossible to use only one type of test or to produce only one response category. Additionally, Chang and Read (2006) used two different item types to assess listening comprehension, which were multiple-choice formats as a listening comprehension test and an interview with a selected number of students after the listening comprehension test.

Dunkel, Henning and Chaudron (1993) suggested that the measurement of listening comprehension has two categories including a production and a manual selection. Completion, short answers and composition belonged to production while multiple-choice and true-false were the manual selection. Other response categories assessing comprehension are content words recall and imitation. Buck (2001) insists a cloze test is not an appropriate term. Cloze test refers to gap-filling. Even though the purpose of the gap-filling is to fill in the blanks based on what they heard, students fill in the blanks without listening. Henning, Gary, N. and Gary, J. (1983) modified the problems, and propose making the blanks on content words. The gap-filling was renamed a listening recall.

III. METHODOLOGY

1. Participants

The participants are 104 Korean undergraduate students enrolled in a general English class at University in Joen-buk province. The students who have never been to English-speaking countries for studying, participated in three different English classes during the Fall semester of 2008. Originally, 112 students participated in this study, but 8 students did not complete the main experimental session. The scores on the pre-test were analyzed to verify homogeneity, and excluded the scores of the eliminated 8 students. The

students were divided into three groups depending on what kind of class they were taking in the semester. Each experimental group received different types of listening materials. By the reason of opening for whole university students, students' majors ran the gamut form pre-professional to liberal arts. Generally, the targets of a general English class are freshmen and sophomores. The majority of the participants were freshmen and a few of them were sophomores who had not passed a general English class.

To verify homogeneity of each group, a pre-test adopted from TEPS practice book was administered before the main experiment and was a criterion to divide low and high proficiency level within group. A one-way ANOVA performed on pretest scores revealed no significant differences between groups (F = .332, p = .718). The mean scores of each group were 7.94 for the argumentative group, 7.89 for the causal/evaluation group, and 7.97 for the narrative group. Therefore, results in Table 1 confirmed that the three groups were homogenous and were comparable in terms of L2 listening comprehension.

TABLE 1
Descriptive Statistics of TEPS Listening Scores

	SS	df	MS	*F*	*p*
Between subject	3.385	2	1.692	.332	.718
Within subject	515.144	101	5.100		
Total	518.529	103			

* $p < .05$

Based on TEPS scores, higher and lower level groups were divided. The mean scores of three higher-level groups (12.2 for the argument group, 10.8 for the causal/evaluation and 11.6 for the narrative group) were not significant different (F = 2.741, p =.105). The mean scores of the lower-level group ranged from 4.2 to 4.8 and revealed that there was no significant difference (F = .581, p = .575). Considering that the sample number of the higher and lower group was small, the results of descriptive statistics showed that these groups could be comparable.

2. Materials

To measure the students' listening ability, a listening comprehension test was prepared during the beginning of the semester. Before preparing the pre-listening test, the researcher used a survey to verify the familiarity of various English proficiency tests. Considering the participants' listening ability, TOEFL, TOEIC and IELTS were determined to be too difficult to test.

The TEPS practice book was commercially published in 2004. A total of 30 questions were used to identify homogeneity among groups. All listening questions and answers were presented once and it took about 25 minutes to complete the pre-test. The test sheet had only directions, and written scripts related with questions and examples of answer were not provided.

Three listening passages used in the experiment were from different genres: argumentative, causal/evaluation and narrative. The listening passages consisting of argumentative, causal/evaluation, and narrative for the study were selected from Impact Issues. All of the materials were recorded as monologues without visual aids and took less than three minutes. Argumentative passages were discussed amongst the three doctors, but there was not any recorded conversation about their opinions.

All the participants in each genre were given three different tasks to evaluate their performances. Responses to these three types of tasks measured their listening comprehension. Most of the previous studies on listening comprehension employed mainly multiple choice questions type. The present study, however, used three tasks: Gap-filling (GF), Summary (ST), and Multiple-choice comprehension task (MC).

Even though the purpose of the gap-filling task was to fill in the blanks based on what they heard, students filled in the blanks without listening. Henning, Gary, N. and Gary, J. (1983) modified the problems, and proposed making the blanks on content words. While participants were listening, they filled in the blanks by using their vocabulary and grammar knowledge. Eight important content words were replaced by blanks following two rules; (1) to make every tenth word amongst the contents words form the listening passage, and (2) to check that the words that provide a lot of information for understanding the passage are unpredictable. The directions for the summary task (ST) were that students were to write down everything based on what they had heard and remembered from listening. There were two options for summarizing either in Korean or in English. Students may prefer using their native language to writing in English. Regarding students' English ability, writing in English seemed to make students suffered from and interrupted the flow or rhythm of the task. Students were not used to writing in English and had never experienced writing in English.

The third task, creating listening comprehension questions could be easily associated with the understanding of the listening passage. According to Weir (1993) the questions of MC task should be based on the listening passage. The questions which were developed by the researcher were identified and ascertained by a native English speaker from Canada who teaches at a college, and a Korean non-native English speaker who is an English teacher at a middle school. All the questions were written in the native language, in order to reduce the influence of factors such as English reading ability, vocabulary and grammar knowledge on comprehension. With 47 students who were not participants in the main

experiment, the difficulty of the questions in GF and MC were tested for consensus of difficulty level. The statistical results of GF task did not present significant differences ($F = .015$, $p = .985$). The mean score of MC, however, showed differences, and this resulted in the need to modify MC questions. MC questions were revised and confirmed that was not significantly different ($F = .086$, $p = .917$). Moreover, GF and MC have different numbers of questions (10 for GF and 6 for MC) which could be a factor for comprehension. Table 6 shows that those listening passages were not long enough to make 10 questions in MC. An increase in the number of questions poses the risk of producing redundant and/or overlapping questions.

3. Procedures

The main experiment was administrated in the fall semester of 2008 and all experiments including the pre-test were done by the researcher during the participants' regular class hours. The first procedure was to prepare a pre-test that was taken from TEPS. The researcher gave Korean explanation of the pre-test in advance. The students were unacquainted with TEPS even though the directions for each part were on the paper.

After the pre-test, the researcher announced the distribution of the main, higher and lower-level groups and scheduled each group. The students who belonged to the main group were arranged first. The process of the main experiment followed order of main group, lower-level group and higher-level group.

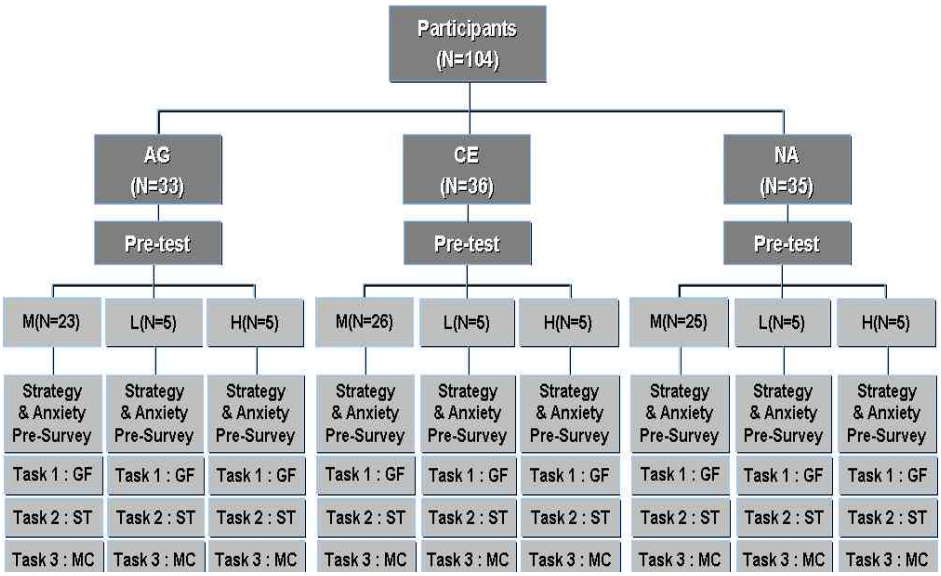
The main experiment took place in the classroom two weeks after the pre-test. After the pre-survey, the main group was instructed to follow the direction. The researcher handed the task sheets for the gap-filling, summary and multiple-choice comprehension to each student. The researcher ensured that students did not look at the task sheets without permission from the researcher. Scarcely had the students received their task sheet when the researcher gave the direction for the first task and asked them to write down their names, student numbers, major and gender in Korean.

The first task that the students fill in the gaps while listening, took less than 10 minutes and the task sheets were returned to the researcher. Before the task, the instructions for each task were addressed by the researcher. The second task was a summary task in which they attempted to write down what they heard and could remember about the listening passage for about 20 minutes. The students were permitted to use both the native language and English. Using both languages promoted to produce more summary without excluding from language ability. It is possible that writing in English is bothering summarization Moreover, all students were accelerated to write sentences rather than single words. After task 2, the researcher also collected all of the task 2 sheet. The third task was the most practiced task among students.

The main group of arguments and the others followed the same steps. The process of the main group took less than 50 minutes to finish. Figure 1 presents the detail description of the overall research design in this study.

FIGURE 1

Description of the Research Design



Notes. N= the number of participants;

H = High listening proficiency; L = Low listening proficiency

AG= Argumentative Group; CE= Cause/Evaluation Group; NA= Narrative Group;

GF= Gap-filling Task; ST= Summary Task; MC= Multiple-choice Task

4. Data Analysis

The listening comprehension task consisted of three parts: 1) gap-filling, 2) summary, and 3) multiple-choice comprehension. When scoring the gap-filling task, the gap that was correctly filled was counted. However, some grammatical errors such as plural nouns, subject-verb agreement, and tense were accepted as alternatives. If the alternatives indicated that students understood the listening, they were given credit in this task.

Each correct answer was worth ten points out of one hundred points in the gap-filling task. The multiple-choice comprehension task included four answer choices, three of which were distracters. The scoring of the multiple-choice comprehension was straightforward and simple in terms of the decision of the right or wrong answer.

The scoring scheme was used to assess students' responses to the summary task. In order to score the summary task, Johnson's (1970) system, based on pausal units or breath

group, was adopted. Chang (2006) also mentioned and adopted Johnson's (1970) system in the study. First, one native speaker who is from Canada took the detailed explanation of the definition of pausal unit which is a criterion in summary task. During listening the three passages: argumentative, causal/evaluation, and narrative, a native speaker divided each passage into pausal units following Johnson' (1970) system.

When he decided on the pausal units, the pausal units were allocated a value in accordance with the value placed on the importance of the information to comprehend the listening passage. Table 2 described each value which is ranked from 1 to 4 points depending on the distribution to semantic weight of the listening passage.

TABLE 2
Explanation of Value

Value	Explanation
4	Information essential to the lecture
3	Important but not necessary
2	Additional or redundant information that helped to reveal the lecture's completeness
1	Trivial or subordinate information, which least affected the lecture contents.

A sample of the summary task was scored twice by two independent raters. One was the researcher and the other rater was an English teacher at a middle school.

Prior to the beginning of scoring, the researcher explained the purpose of the study. Another rater received the description of the pausal unit, value and scoring sheets, and training information or how to score the summary task. Using this rubric, students' summaries were scored. The total point of value that students could get is as follows 131 in argumentative, 163 in causal/evaluation, and 139 in narrative passage. The value points that students scored were converted into a percentage which finally indicated the summary score.

IV. RESULTS AND DISCUSSION

Data from all the listening tasks was aggregated for analysis. The different listening tasks aimed to demonstrate the listening comprehension on the different genres: Argumentative (AG), Causal/valuation (CE), and Narrative (NA). The descriptive statistics of the test results for all the listening comprehension tasks of all participants (N=104) was performed by one-way ANOVA. The Levene Test for the One-Way

ANOVA did not confirm any significant difference (p = .897), which meant the results were robust for identification of the homogeneity and the value in use. The results revealed significant differences among three groups (F = 4.185, p = .018) Therefore, differences on all the listening comprehension tasks found among groups can be attributed to genre differences.

TABLE 3

One-way ANOVA Results of English Listening Task Scores

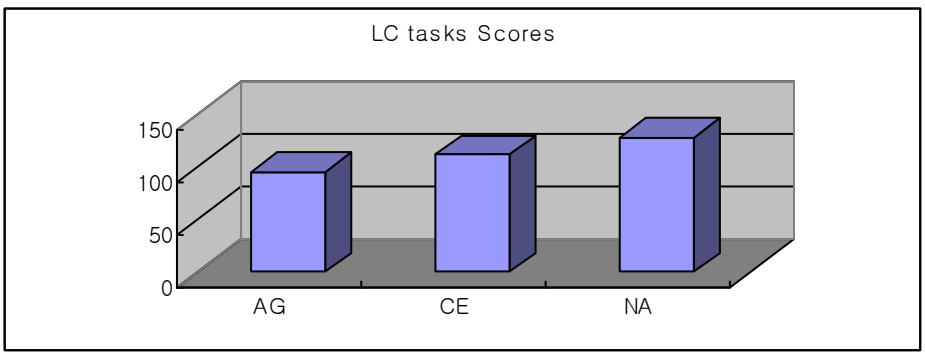
	Source	*df*	SS	MS	*F*	*p*
LC task * genre	between group	2	18,162.139	9,081.070	4.185	.018
	Within group	101	219,145.800	2,169.761		
	Total	103	237308.000			

* $p < .05$

NA group (M = 128, SD = 45.49) scored the highest, then CE group (M = 112, SD = 45.89), and the lowest was by the AG group (M = 96, SD = 48.42). The sum of task scores was 300 each individual task was worth 100 point.

FIGURE 2

Mean Scores for Listening Comprehension Scores by Group



In order to see if there were any differences in terms of task, a One-way ANOVA was submitted for each task. The NA group had the highest mean score, but the detailed task score indicated different results. As such it was necessary to identify the different mean scores of each group by tasks.

Based on the statistical results reported in Table 4, the genre difference in Task 1 showed significantly difference (F = 3.975, p = 022). As shown in Figure 2, Task 1 also had different results compared. The mean score of AG group (M = 29.69), CE group (M =

44.16), and NA group (M = 40. 85) indicated that CE group scored significantly better than the others. The lowest score was by the AG group for Task 1.

TABLE 4
Descriptive Statistics of Task 1, 2 and 3

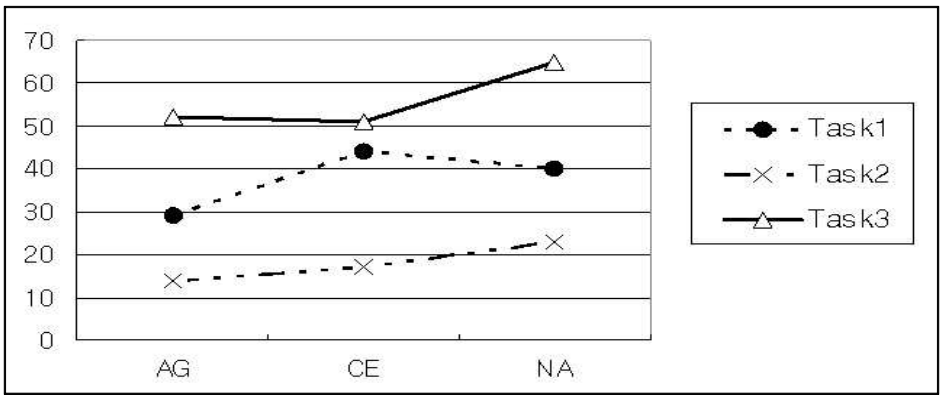
	Source	*df*	SS	MS	*F*	*p*
Task 1 * Genre	between group	2	3907.591	1953.795	3.975	.022
	within group	101	49646.255	491.547		
	total	103	53553.846			
Task 2 * Genre	between groups	2	1364.731	682.366	8.777	.000
	within groups	101	7852.153	77.744		
	total	103	9216.885			
Task 3 * Genre	between groups	2	4010.088	2005.044	3.357	.039
	within groups	101	60318.373	597.212		
	total	103	64328.462			

* $p < .05$

Table 4 also contained the ANOVA result of the second task, the summary. It showed that the genre difference in Task 2 was significantly different (F = 8.77, p = .000). In contrast with the result of Task 1, the result showed the highest score was received by the NA and AG group as the lowest. By and large, the mean score in Task 2 presented a similar shape with the overall result of the mean, but the scores of Task 2 were lower than the overall results.

Moreover, there was a significant difference (F = 3.357, p = .039) in Task 3 (Multiple-choice comprehension Task). The scores of Task 3 suggested that AG and CE groups were nearly the same and AG group was slightly lower than the CE group. NA group indicated the highest mean score as well on Task 2. To obtain a clear picture of the task differences on genre, the scores for each task were submitted to ANOVA. In sum, Figure 3 presented that the results from each task showed differences by genre.

FIGURE 3
Comparison of Tasks' Mean scores by Genre



Task 2 (Summary) and 3 (Multiple-choice comprehension) were performed best by the NA group while AG group scored the lowest in Task 1 and 2. The highest score for Task 1 was received by the CE group, yet the CE group scores the lowest in Task 3. Nevertheless, Task 2 presented a statistically significant difference, and the graph delineated a narrow margin among the three tasks.

The first and second research question considered the effects of genres and tasks on English listening comprehension. Associated with first research question, the results of the statistical analysis, using one-way ANOVA implied that genre significantly affects Korean college students' listening comprehension, and indicated that listening comprehension scores for each genre were also substantially different.

The following results illustrated the effects of the genre difference on listening comprehension focusing on proficiency levels. The Levene Test for the One-Way ANOVA did not reveal a significant difference (p = .409), which meant the statistics results were possible to use and strong to verify the homogeneous group. The results of 5 higher and lower participants from each group indicated significant difference (t = 4.505, p = .000).

TABLE 5
Results of the t-test by Proficiency Level on Task 1, 2, & 3

Proficiency level	Task 1			Task 2			Task 3		
	Mean	*t*	*p*	Mean	*t*	*p*	Mean	*t*	*p*

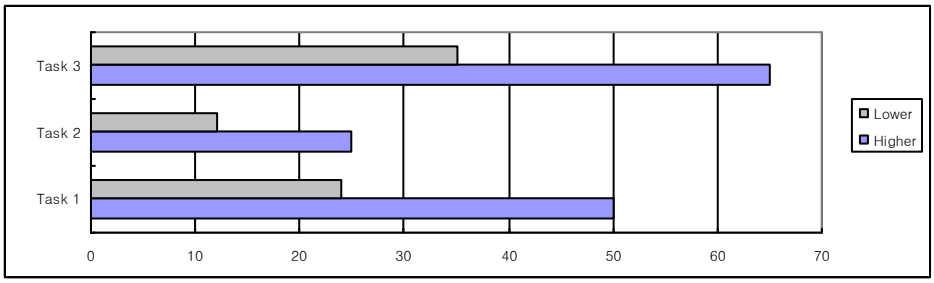
The Effects of Different Types of Genres and Tasks									
Higher	50.66	3.223	.003	25.93	5.595	.000	65.60	3.983	.000
Lower	24.66			12.53			35.66		

* $p < .05$

As shown in Table 5, strong statistical significances were found by proficiency level. The results of the t - test comparing higher and lower group scores revealed that Task 1 (t = 3.223, p = .003), Task 2 (t = 5.595, p = .000), and Task 3 (t = 3.983, p = .000) were significantly different.

FIGURE 4

A Comparison of Higher and Lower Level Group Scores by Task



As summarized in Figure 4, both the higher and lower level group presented the lowest scores in Task 2, and the highest scores in Task 3. A gap between the higher and lower level group illustrated that Task 3 had a wide gap while there was a narrow gap among the three different tasks in Task 2.

TABLE 6

Two-way ANOVA Results of the Higher and Lower Level Groups for Tasks

Source	SS	*df*	MS	*F*	*p*
Genre	46.667	2	23.333	.045	.956
Proficiency	5070.000	1	5070.000	9.750	.005
Genre * Proficiency	1140.000	2	570.00	1.096	.350
Error	12480.000	24	520.00		
Total	61300.000	30			

* $p < .05$

As shown in Table 6, the results of the Two-way ANOVA indicated a significant effect for listening comprehension. There was no main effect for the genre difference on the listening comprehension task (F = .045, *p* = .956). Proficiency, however, showed a significant effect for the tasks on listening comprehension. Results of the ANOVA presented no correlation between genres and proficiency level on the listening comprehension task (F = 1.096, p = .350).

The third research question focused on whether different proficiency levels would affect the listening comprehension in a second language. With regards to the third research question, the results proved that listening comprehension scores between members of the higher and lwer proficiency level groups were significantly different. That is, higher proficiency level learners perform better than their counterparts in the lower proficiency level group, regardless of the different types of genres and tasks administered.

V. CONCLUSION AND IMPLICATION

Based on the results of listening comprehension for the narrative genre which loosely organized structure facilitates understanding of the text, the overall statistical results show that the highest scores were obtained by members of the narrative group and the lowest scores received by participants in argumentative group. Meyer & Freedle (1984) mentioned that organization pattern affects language learning.

However, detailed listening comprehension scores by each task indicated contradictions. Especially figures for GF (Gap-filling), task 1, present a different pattern as compared with the overall results. The highest score for the GF task appeared in the CE group (Causal and Evaluation). This conflicting evidence implies that the feature for assessment tasks is considerable in this study. Chaudren and Richards (1986) used a cloze listening exercise, referred to as Gap-filling in this study, in which learners had to stop to answer the test during listening. To accomplish both ST (Summary) and MC (Multiple-choice comprehension) tasks, learners were required to understand an entire listening text and to remember the whole content. In contrast, GF demands a different information process for task completion. Since listening comprehension is not essential for executing this task, listening tended to focus only on completing the missing words of the activity.

Listening comprehension is difficult to measure because of the mental operations involved in the comprehension process. Listening comprehension can only be observed through indirect and second-hand measurements such as assessment tasks that must reflect the listener's mental activity (Buck, 1997). Current findings in this study validate the theory that assessment task types play a significant role in presenting listening comprehension and promoting listening comprehension process. Different types of tasks

seem to measure different levels of information. In order to assess and gain accurate listening comprehension, it is vital that second language learners gain exposure to and training on multiple assessment tasks which are based on the strengths and weakness of the various tasks (Berne, 1992; Brindley, 1998; Rhodes, Watson, & Barker, 1990; Shohamy, 1997).

Moreover, in foreign language learning, learners need to encounter different genres for promoting the target language learning and accepting the cultural differences (Christie, 1999). According to the different genres, listening comprehension appeared differently on listening comprehension tasks from previous researches. Different test item types show different listening comprehension. When a teacher assesses listening comprehension, a teacher considers what type of text students will hear and what kinds of tests students will be given. Teachers prepare various genres of listening materials and tests type to train their students.

In EFL settings, the findings of the study provide evidence of the three genre groups gained the highest score in multiple-choice comprehension task. This task type is the most familiar type and is used in school as test responses. On the other hand, summary task required ore complex comprehension and cognitive process. By developing different types of information and cognitive processing, teachers need to provide various assessment tasks. Students will become more able to complete complex listening tasks.

In addition, the results can be expected washback effects to develop test items. Some researchers have proved that different test items activate different processing demands (Bern, 1993; Hansen & Jensen, 1994). Buck (2001) asserts for that the test has an influence on classroom teaching called the washback effect. This effect occurs teacher teaches to the test and learners focus their learning on subjects that will be on the test. Training different types of genres and test types on listening can promote learners' skills and strategies.

Considering the limitations of the study, it is necessary to examine a greater variety of genre and task types in future research. Meyer categorizes five types of rhetorical organizations. Each rhetorical organization has different structures and characteristics. Rost (2002), also, mentions different genres that have different information organization, listening purposes, and speaker focus. The different genres have different features and characteristics. The features and characteristics of a genre could cause different effects on listening comprehension. Because of this, a greater variety of genres need to be studied. Moreover, a small number of higher and lower level groups were selected for this study. Thirty students is too small a sample size to generalize differences of listening comprehension on proficiency level.

REFERENCES

Edu chosun. (2004). *Ready for TEPS*. Seoul: Edu Chosun.

Allen, E. D., Bernhardt, E. B., Berry, M. T., & Demel, M. (1988). Comprehension and test genre: Analysis of secondary school foreign language readers. *The Modern Language Journal*, *72*(2), 163-172.

Barry, S., & Lazarte, A. A. (1995). Embedded clause effects on recall: Does high prior knowledge of content domain overcome syntactic complexity in students of Spanish? *The Modern Language Journal*, *79*(4), 491-504.

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.

Brindley, G. (1998). Assessing listening abilities. *Annual Review of Applied Linguistic, 18,* 171-191.

Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press.

Carrell, P. L. (1984). The effects of rhetorical organization on ESL readers. *TESOL Quarterly, 18*(3), 441-469.

Chang, Y-F. (2006). On the use of the immediate recall task as a measure of second language reading comprehension. *Language Testing, 23*(4), 520-543.

Davis, J. N., & Bistodeau, L. (1993). How do L1 and L2 reading differ?: Evidence from think aloud protocols. *The Modern Language Journal*, *77*(4), 459-472.

Devine, T. G. (1978). Listening: What do we know about fifty years of research and theorizing? *Journal of Reading, 21*, 296-304.

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*(2), 180-191.

Feyten, C. (1991). The power of listening ability: An overlooked dimension in language acquisition. *The Modern Language Journa*l, *75*, 173-80.

Hansen, C., & Jensen, C. (1994). Evaluating lecture comprehension. In Flowerdew, J. (Ed.), *Academic listening* (pp. 241-268). Cambridge: Cambridge University Press.

Henning, G., Gary, N., & Gary, J. (1983). Listening recall: A listening comprehension test for low proficiency learners. *System, 11*(3), 287-293.

Lee, J. (2009). The effects of using listening comprehension strategies on TOEIC listening comprehension and moderator model. *English Language & Literature Teaching, 15*(4), 345-364.

Jeon, B., & Park, G. (2001). A study on recent research findings in EFL/ESL listening. *Foreign Languages Education, 8*(1), 293-317.

Kobayashi, M. (2002). Method effects on reading comprehension test performance: Text organization and response format. *Language Testing, 19*(2), 193-220.

MacDonald, M., Badger, R., & White, G. (2000). The real thing?: Authenticity and academic listening. *English for Specific Purposes*, *19*(3), 253-267.

Meyer, B. J. F. (1975). *The organization of prose and its effects on memory*. Amsterdam: North Holland Publishing Co.

Meyer, B. J. F. (1985). Prose analysis: purpose, procedures, and problems: Parts I and Ⅱ. In B. Britton & J. B. Black (Eds.), *Understanding expository text* (pp. 269-304). Hillsdale, NJ: Lawrence Erlbaum.

Meyer, B. J. F., & Freedle, R (1984). Effects of discourse type on recall. *American Educational Research Journal, 16*(1), 121-143.

Morley, J. (1991). Listening comprehension in second/foreign language instruction. In M. Celce-Murcia (Ed.), *Teaching English as a second or foreign language* (pp. 81-106). Boston, MA: Heinle&Heinle Publishers.

Nam, E-H., & Seong, M-H. (2009). The effects of dictation practice in English listening lasses. *English Language & Literature Teaching*, *15*(4), 177-197.

Oh, S. Y. (2001). Two types of input modification and EFL reading comprehension: Simplification versus elaboration. *TESOL Quarterly*, *35*(1), 69-96.

Olsen, L. A., & Huckin. T. N. (1990). Point-driven understanding in engineering comprehension. *ESP Journal, 9*(1), 33-47.

Oxford, R. (1989). Use of language learning strategies: A synthesis of studies with implications for strategy training. *System, 17*, 235-247.

Oxford, R. (1990). *Language learning strategies: What every teacher should know.* New York: Newbury House.

Park, G-P. (2000). The effect of question presentation and question type on listening comprehension. *English Teaching*, *55*(1), 141-156.

Park, G-P. (2001). The effects of linguistic ability, background knowledge, and question types on EFL listening comprehension ability. *English Teaching, 56*(2), 245-263.

Raymond, P. M. (1993). The effects of structure strategy training on the recall of expository prose for university students reading French as a second language. *The Modern Language Journal, 77*(4), 445-458.

Rhodes, S. C., Watson, K. W., & Barker, L. L. (1990). Listening assessment trends and influencing factors in the 1980s. *Journal of the International Listening Association, 4,* 62-82.

Rost, M. (2002). *Teaching and researching listening*. London: Longman.

Saito, Y., Horwitz, E. K., & Garza, T. J. (1999). Foreign Language Reading Anxiety. *The Modern Language Journal, 83*(2), 202-218.

Shohamy, E. (1997). Testing methods, testing consequence: Are they ethical? Are they fair? *Language Testing, 14,* 540-549.

Shohamy, E., & Inbar, O. (1991). Validation of listening comprehension tests: The effect of text and question type. *Language Testing, 8*, 23-40.

Tsui, A., & Fullilove, J. (1998). Bottom-up of top-down processing as a discriminator of L2 listening performance. *Applied Linguistics, 19*, 432-451.

Weir, C. (1993). *Understanding and developing language tests*. New York: Prentice Hall.

Examples in: English
Applicable Language: English
Applicable Levels: Secondary

Sookhyun Kim
Sujeong-town APT 9/304
Dunsan-Dong, Seogu, Daejeon
Tel: 010-2053-2727
Email: sookhyunkim0624@hanmail.net

Received in January, 2010
Reviewed in February, 2010
Revised version received in March, 2010