교육학석사학위논문

The Effects of Stimulus Type on Korean EFL College Students' Performance on an Integrated Speaking Task

듣기 자료의 유형이 한국 대학생의 통합형 영어 말하기 시험의 수행에 미치는 효과

2019년 2월

서울대학교 대학원

외국어교육과 영어전공

노 윤 희

The Effects of Stimulus Type on Korean EFL College Students' Performance on an Integrated Speaking Task

by

Yoon Hee Nho

A Thesis Submitted to
the Department of Foreign Language Education
in Partial Fulfillment of the Requirements
for the Degree of Master of Arts in Education

At the
Graduate School of Seoul National University

February 2019

The Effects of Stimulus Type on Korean EFL College Students' Performance on an Integrated Speaking Task

듣기 자료의 유형이 한국 대학생의 통합형 영어 말하기 시험의 수행에 미치는 효과

지도교수 소 영 순

이 논문을 교육학 석사 학위논문으로 제출함

2018년 11월

서울대학교 대학원

외국어교육과 영어전공

노 윤 희

노윤희의 석사학위논문을 인준함

2018년 12월

위 원 장 _____

부위원장 _____

위 원 _____

The Effects of Stimulus Type on Korean EFL College Students' Performance on an Integrated Speaking Task

APPROVED BY THESIS COMMITTEE:

_____

YONG-WON LEE, COMMITTEE CHAIR

_____

YOUNGWOO KIM

_____

YOUNGSOON SO

Abstract

The Effects of Stimulus Type on Korean EFL College Students' Performance on an Integrated Speaking Task

Yoon Hee Nho

English Major, Dept. of Foreign Language Education

The Graduate School of Seoul National University

The primary goal of the present study was to examine the influence of stimuli types on the test-takers' listening-speaking task performance. More specifically, the study also described which aspects of audio and video stimuli facilitated or hindered the test-taking by exploring the participants' examination processes. Eight test-takers of tertiary EFL learners participated in this study. They were each assigned audio- and video-stimulated listening-speaking integrated tasks. The stimuli topics and formats were counterbalanced (via Latin squares) to control for test content and order effects. Afterwards, participants were invited to an individual stimulated recall protocol in which they watched a video of themselves taking tests, and discussed their own test-taking processes. Qualitative analysis of the oral performance indicated that the visual cues could improve their performance at the macro-, micro-, and word levels. Interview data revealed that the facilitating factors and hindering factors were present in both types of stimuli, but the participants tended to relate positive remarks more frequently with video

stimuli than with audio stimuli. These findings support the argument that the video stimuli should be incorporated into listening-speaking test where test-takers listen to a lecture and then summarize. Consequently, the current study sheds light on how practitioners and theoreticians can reflect this trend of listening-speaking task stimuli in their educational and research practices in relation to the assessment of English for academic purposes.

Key Words: video stimulus, audio stimulus, listening-speaking test, summarizing, discourse analysis, stimulated recall interview

Student Number: 2016-21787

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

The current research investigates the test-takers' performance differences on integrated tasks under the conditions of audio stimulus and video stimulus, and explores the participants' stimuli usage. This introductory chapter provides the background, purpose, and outline of the study. Section 1.1 sets the background and explains the purpose of the study. Section 1.2 identifies the research questions, and section 1.3 ends the chapter by outlining the overall organization of the thesis.

1.1 The Motivation of the Study

Many language test developers have been concerned about making their test more *authentic*. As test *authenticity* is defined by "the degree of correspondence between the characteristics of target language use tasks and those of the test task" (Bachman & Palmer, 1996, p. 23), test developers studied ways to reflect real-life language use situation in their tests (Brown & Abeywickrama, 2010; Fulcher, 2010). In the case of academic language test, integrated tasks can enhance the test authenticity since these tasks approximate some of the real academic tasks where students listen or read, and then speak or write (Brown, Iwashita, & McNamara, 2005; Frost, Elder, & Wigglesworth, 2011). In addition to improving the test authenticity, using integrated tasks in academic language tests has another benefit. Integrated tasks can augment the equity of the test as test-takers do not need to rely on their own background knowledge overly to perform on tasks of different topics (Read, 1990).

Despite these strengths, not many, but only a few of the previous studies or existing tests used an integrated task (as mentioned in Batty, 2015; Yu, 2013). One reason could be that integrated tasks were difficult to score. In the integrated task performance, multiple language skills interact and produce one outcome. Therefore, the scorers cannot be sure whether the poor performance on integrated listening-speaking results from the test-takers' listening skill or speaking skill problem (Choi & So, 2018; Cumming, 2013; Yu, 2013). Another reason is that integrated tasks require a threshold level of ability to perform successfully. For example, on a listening-speaking integrated task, test-takers below this threshold level of listening ability could not understand the source material and may not be able to perform integrated speaking task competently. Therefore, the task was only for those test-takers who achieved or surpassed this minimum level of proficiency (Cumming, 2013).

Included among the integrated tasks used in the previous studies and existing academic language tests were listening-speaking integrated tasks. In the test, test-takers listened to a lecture and then spoke to summarize it (Brown et al., 2005; Educational Testing Service, 2017). This type of test tended to reflect the university students' and instructors' opinions: Most of them believed that summarizing skill (Kirkland & Saunders, 1991) and formal presentation skill (Ferris & Tagg, 1996; Ockey, Koyama, Setoguchi, & Sun, 2015) should be mastered by a successful academic English speaker.

The authenticity of this listening-speaking task would have been maximized by providing test-takers with the chance to attend real lectures.

However, because it was practically impossible, especially in the case of large scale tests, video stimuli were suggested as one of the best alternatives to real lectures, but in reality, audio-only stimuli replaced the real lectures (Educational Testing Service, 2017; Shin, 1998).

So far, some researchers have studied the effects of audio and video stimuli on listening test performance (Batty, 2015; Brett, 1997; Coniam, 1997; Londe, 2009; Suvorov, 2008, 2013, 2015; Wagner, 2010, 2013), and only a few others have investigated their effects on reading-listening-writing test performance (Cubilo, 2011; Cubilo & Winke, 2013). None of the studies to date have investigated the effects of stimuli types on the test-takers' listening-speaking test performance.

With these as a background, the present study attempts to investigate the effects of stimuli types on test-takers' listening speaking performance, and their test-taking process to explore the effects of the two stimuli types by implementing qualitative research methods. This qualitative study is expected to benefit test designers, and researchers who seek to have a better understanding of audio and video stimuli used in an integrated task.

1.2 Research Questions

The specific research questions that guide the current study are as follows:

1. How do test-takers' speaking performances in video-stimulated test differ from those in audio-stimulated speaking test?
2. How do test-takers utilize stimuli in an integrated speaking test?

In this study, university students' performances in video- and audio-stimulated listening-speaking tests will be qualitatively analyzed.

1.3 Organization of the Thesis

The present thesis includes six chapters. This chapter, Chapter 1, explains the factors that motivated this study and presents the research questions it aims to address. Chapter 2 introduces the theoretical framework that shaped this study and presents the relevant literature. Chapter 3 describes the research methodology used in this study, outlining the phases of participant recruitment, materials development, data collection and data analysis. Chapter 4 presents the results of the current study, which are subsequently discussed in Chapter 5. Chapter 6 concludes the thesis by recapitulating the major findings and discussing their implications for theory and practice.

Chapter 2. Literature Review

The purpose of this chapter is to establish a theoretical background for the current study by reviewing key concepts and major findings from previous studies regarding the use of integrated tasks in tests and the effects of stimulus type on the test-takers' performance. This chapter is divided into five parts. Section 2.1 reviews the definition of integrated tasks. Section 2.2 introduces video as a new type of stimuli for speaking tests. In Section 2.3 discusses the roles of listening and speaking skills in the listening-speaking tasks. Section 2.4 summarizes previous studies that explored the participants' perceptions of stimuli types. Finally, Section 2.5 identifies some gaps in the literature.

2.1 Integrated Tasks

Integrated language test tasks or *integrated tasks* can be defined as tasks in which the test-takers generate response(s) based on the input that has been provided (Lewkowicz, 1997, as cited in Frost et al., 2011). The opposite type of task would be *independent task* that involves test-takers' single language skill such as listening skill or speaking skill. Using integrated tasks in a test can have at least two benefits because of the fact that they require test-takers to depend on the input to perform. The first benefit of using integrated tasks is that they can enhance the authenticity of the test. In real-life academic context, students usually speak or write based on the reading input (i.e., books) or listening input (i.e., lectures) (Brown et al., 2005; Butler, Eignor, Jones, McNamara, & Suomi,

2000). Integrated tasks, which require test-takers to perform based on the input, or a stimulus, approximate these real world communicative acts better than independent tasks (Fulcher, 2010) and can thus enhance the test authenticity (Brown & Abeywickrama, 2010; Fulcher, 2010).

The second strength of the integrated tasks is that they can promote equity of the test. In most of the language test situation, test-takers bring their own background knowledge to perform. Test-takers who have much prior knowledge about a given topic would perform better, but those who have less relevant background knowledge would have disadvantage in their test performance. Read (1990) argued that integrated tasks can reduce the undesirable effects of the test-takers' prior knowledge because test-takers draw information from the same input in an integrated task. In integrated tasks, therefore, test-takers' language proficiency can be measured with minimum interference of their background knowledge.

Notwithstanding these strengths, however, not many studies to date have dealt with integrated tasks, especially those involving the speaking skill. The reasons could be that integrated tasks have been used relatively recently in language testing (as mentioned in Frost et al., 2011; Yu, 2013), and that the scoring of speaking tasks is complex than scoring other skill tasks (Luoma, 2004).

Nevertheless, one of the widely used listening-speaking tasks in academic English tests would be the oral presentation of a summary (Brown et al., 2005; Educational Testing Service, 2017; Frost et al., 2011). For instance,

TOEFL iBT includes a listening-speaking task in which test-takers listen to a lecture and then produce an oral summary of the lecture. By using this summary presentation task, the validity and reliability of the test can be enhanced. Specifically, the validity of score interpretation is augmented because the test can elicit skills that a student should master in order to successfully complete academic study in an English-medium institution. Needs analytic studies with university students and instructors have shown that summarizing skill (Kirkland & Saunders, 1991) and formal presentation skill (Ferris & Tagg, 1996; Ockey et al., 2015) are necessary for a student studying at an English-medium university. Those who perform well on an oral summary presentation task can be the one who will be good at real-life academic English.

Another reason why the validity of score interpretation is enhanced with summary presentation task is that oral presentation produces a monologue. In a conversation setting, it is complex to measure a test-taker's proficiency than in a monologue setting (Luoma, 2004). Specifically, the conversation setting makes it difficult to examine the test-taker's interactional competence because the test-taker's conversational performance changes according to the interlocutor's proficiency (He & Young, 1998). In other words, test-takers' proficiency could be better measured without the interference of an interlocutor's proficiency in an oral presentation task.

Not only the test validity, but also the reliability of test scores can be improved by using a summary presentation task in a speaking test. Scoring the summary task outcome could enable consistent scoring than independent tasks.

Summarizing does not involve students' personal opinions or background knowledge, but is strictly dependent on the source text. This convergence of the task outcome could simplify the scoring process and could lead to reliable scoring (Brown & Abeywickrama, 2010; Fulcher, 2010). In short, oral presentation of a summary can be used in speaking assessment of academic English with high level of validity of interpreting test results as an indicator of success in real-life academic tasks and possibly high level of reliability of test scoring.

2.2 Video as Stimuli

The authenticity of the test can be even enhanced by using more authentic stimulus in the listening-speaking integrated tasks. In the listening-speaking integrated task of lecture summarizing, test-takers listen to a stimulus lecture and then speak to present their summary. Allowing the test-takers to attend a real lecture in a classroom would be the best option to maximize the authenticity of the task, but it is practically impossible, especially in the case of large-scale standardized tests. In this case, video clips can be better replications of the real lectures than audio recorded versions in that video can show gestures, facial expressions and nonverbal features of lectures such as images and writings on the board. However, none of the previous studies investigating oral summary listening-speaking tasks or any operational test including listening-speaking integrated tasks have used video clips as stimuli (Brown et al., 2005; Educational Testing Service, 2017; Frost et al., 2011).

To date, studies that have compared test-takers' performance under the conditions of audio and video stimuli investigated the relative effects of stimulus on assessing the target construct of "listening" ability. However, results from those studies were inconsistent; while some studies reported that video had facilitating effects on the test-takers' performances (Wagner, 2010, 2013), other studies discovered that there was no significant difference between the audio- and video-stimulated test performances (Batty, 2015; Brett, 1997; Coniam, 2001; Londe, 2009). Suvorov (2008) even found that video stimuli had negative effects on the test-takers' listening performances.

Only a few of the previous studies explored the effects of stimulus type on the test-takers' performance in integrated tasks (Cubilo, 2011; Cubilo & Winke, 2013) but these studies investigated "writing" tasks unlike the present study that is to investigate integrated "speaking" tasks. Cubilo (2011) and Cubilo and Winke (2013) studied the performances on listening-reading-writing integrated tasks. They scored the participants' written outcome and reported that there was no significant difference in the general essay scores, but that the video group significantly did better than the audio group only in "Language Use" rating subcategory.

While Cubilo (2011) and Cubilo and Winke (2013) provided some insight about the effects of stimuli types on test-takers' language performance, they had a limitation of not using *content visuals* as video stimuli. *Content visuals* illustrate the content features of the video (Cubilo & Winke, 2013; Ginther, 2002; Suvorov, 2008, 2015). That is, *content visuals* can (1) replicate

the oral input (e.g., the lecturer in the video writes down on the board what she is orally explaining); (2) illustrate the oral input (e.g., the test-takers can see a picture of what is being described orally); (3) organize information delivered in the stimulus (e.g., the lecturer provides a diagram of a process that she orally described); or even (4) supplement the oral stimulus (e.g., the visual cues provide information that is not present in the oral stimulus) (Levin, 1989, as cited in Bejar, Douglas, Jamieson, Nissan & Turner, 2000). Bejar et al. (2000) explained that the content visuals of type (1) through (3) could facilitate the listening comprehension while type (4) can even complicate the process. Although some texts or graphs were incorporated in content visuals, these visual cues were not considered as "reading" input because only a minimum level of reading skill was involved in the comprehension of these visuals (Bejar et al., 2000).

The opposite of *content visuals* would be *context visuals* which show the physical and situational context in which the speaker is, such as the lecture room environment, or the speaker's age. After comparing content and context visuals, Suvorov (2015) concluded that the participants were more interested in content visuals, and that they found content visuals more informative than context visuals. However, the majority of the previous studies dealing with the video stimuli used context visuals (Cubilo, 2011; Cubilo & Winke, 2013) instead of content visuals (Wagner, 2010). Other studies (Batty, 2015; Brett, 1997; Coniam, 2001; Londe, 2009; Suvorov, 2008; Wagner, 2013) did not even specify the types of visuals they used although the use of one type of visual can change the degree to which the stimulus is informative or interesting. In a listening-speaking

integrated task where test-takers are required to obtain information by listening to a lecture, content visuals should be used to help test-takers understand the input materials instead of context visuals because the former type seems to be more effective in delivering the concepts of the lectures.

2.3 The Roles of Listening and Speaking

Previous studies informed that types of stimulus can affect test-takers' performance in multiple ways. For instance, Cubilo and Winke (2013) investigated the effects of audio and video stimuli on test-takers' listening-reading-writing task performance, and they concluded that the implementation of video stimuli improved only the "Language Use" subcategory of test-takers' writing. However, their study could not clearly show to what extent each of these three skills that are involved in task completion affected test-taker performances in what ways. The reason could be that in integrated tasks, the final outcome is produced as a result of interactions between the involved skills (Choi & So, 2018; Cumming, 2013; Yu, 2013). That is, test-takers' successful response could prove the proficiency of all of the skills involved, but in the case of less than perfect performances, it would be hard to define which of the skills is responsible for such performances.

The listening-speaking integrated task, which is the focus of the current study, requires a threshold level of listening ability for a successful spoken response (Choi & So, 2018). That is to say, in a listening-speaking integrated task, test-takers have to comprehend the listening input in order to speak well.

Those who could not understand the listening input could not perform successfully in the speaking. Therefore, to study the effects of stimulus types on a test-taker's listening-speaking task performance, the researcher should make it clear whether the unsuccessful performance results from the test-taker's problem in his or her listening skill, or speaking skill.

2.4 Test-takers' Perception

In addition to the effects of stimuli type on test-takers' performance on language test, previous studies sought to investigate the test-takers' preference of stimuli type. After the participants completed the language tests, researchers asked them which type between audio stimuli and video stimuli they liked better. The result of this preference questions were again inconsistent (Brett, 1997; Coniam, 2001; Cubilo, 2011; Cubilo & Winke, 2013). A closer look at these previous studies can yield reasons for this disagreement. Coniam (2001) reported that many of his participants felt distracted by the video stimuli. However, as he pointed out himself as a limitation of his study, his participants did not experience both types of stimuli, audio and video stimuli. This could be quite a critical limitation because his participants could have answered just by guessing what it would be like to use the other type of stimuli.

Another reason for the disagreement could be found in the nature of the investigation method. In the former studies, participants were given options, such as "helpful," "distracting," or "neutral," to express their preference of stimuli type (Cubilo, 2011). This method can be time-efficient, but could be ineffective

in that simplifying one's preference into one word can eliminate much information that affected the decision-making processes. A preference is affected by one's unique experiences and only a detailed investigation would reveal what exactly affected his or her inclination. In order to have a deeper understanding of test-takers' stimuli type preferences, a qualitative research method would be needed (Dornyei, 2007; Levy, 2014).

2.5 Gaps in the Literature

This chapter reviews not only literature underpinning the use of integrated tasks in tests, but also the effects of stimuli types on test-takers' language performance. Based on the review of such previous literature, the following two areas have been identified that were not sufficiently investigated. First of all, the review highlights the need to investigate the effects of stimulus types, audio and video, on test-takers' performance in listening-speaking integrated tasks. Second, it seemed necessary to investigate how each type of stimuli affects listening and speaking performances, and test-takers' perception of each type of stimulus. This study is an attempt to fill in these gaps. To this end, the study uses qualitative research methods which could enrich the understanding of the phenomena in concern: the effects of different types of stimuli on test-takers' performance on listening-speaking integrated tasks, and the participants' usage of the stimuli (Dornyei, 2007). Specifically, discourse analysis is done to compare the performance of the participants. In-depth interviews are conducted with each participant applying the stimulated recall protocol to explore their test-taking

processes and perceptions.

Chapter 3. Methodology

This chapter contains the discussion of the methodological approach and research design employed to examine the research questions set forth in Chapter 1. The chapter consists of four sections that detail the participant recruitment and selection process, materials development procedure, and then present data collection and analysis methods.

3.1 Participants

The selection of participants included two steps. First, the preliminary survey was done to recruit 25 potential participants. Then, potential participants' information was studied to select the eight final participants. The following sections detail the participant recruitment process.

3.1.1 Preliminary Screening Survey

Eight Korean tertiary-level learners of English participated in this study. These final participants were recruited and selected based on the responses to a preliminary screening survey that involved questions about individual variables that were expected to impact the process and outcome of audio- and video-stimulated examinations. These variables were general English proficiency levels, and prior knowledge of potential topics of the source texts (refer to Appendix 1 for the translated version of the survey).

The first variable considered was the participants' English proficiency. Because this variable can affect the listening comprehension (Feak & Salehzadeh,

2001), it was manipulated to make the stimuli type difference the only factor that could influence the test-takers' listening comprehension. English proficiency levels were measured based on the participants' TEPS (Test of English Proficiency developed by Seoul National University) scores. To make sure that test-takers comprehend and then produce a summary of the source text, a TEPS score of 800 or higher was required as the threshold level of listening ability (Choi & So, 2013). Participants who had TEPS scores below 800 could fail to understand the stimulus lectures and could not produce a spoken outcome.

The second variable manipulated was the level of participants' prior knowledge about the topic of the source texts. The prior knowledge of the source texts had to be controlled to prevent the possibility that a participant used his/her background knowledge even without listening to the stimulus. To this end, the preliminary questionnaire collected two types of information. The first question asked the participants' major at university. The second question asked the participants to check their level of prior knowledge about the potential topics of the source texts on a five point Likert scale (1 = *know nothing* ↔ 5 = *know very well*). The questionnaire enlisted ten potential topics for the source texts such as the fields of law, art, or engineering, among which two final topics would be selected.

3.1.2 Participant Selection

The pre-test questionnaire outlined in 3.1.1 as well as the consent form were posted online, on a bulletin board of a university's online community, in

order to recruit tertiary-level EFL learners, who were expected to be the most experienced with academic English use in Korean context. As the goal of using this online survey was to recruit a homogeneous group of participants, only the social science majors and students with TEPS scores of 800 or higher were asked to fill in the questionnaire to narrow down the range of the potential participants.

Initially, 25 potential participants expressed their willingness to participate in the research by completing the online questionnaire. From this pool, the most homogeneous eight participants were sampled based on their prior knowledge and English proficiency. Initially, potential participants checked their knowledge level of ten fields of study in the preliminary survey. Participants who responded with either 1 or 2 to the following five-point Likert scale survey questions were considered for final participants: "Check the degree of your knowledge for each of the following fields of study – environmental issues (EI)," and "– Earth science (ES)." Participants' answers for the other eight fields of study, such as law or arts, were too diverse, and disabled any homogenous grouping of participants. Among the participants with little knowledge about EI and ES, those who had the most similar TEPS scores were selected as final participants. The participants' background information is summarized in Table 3.1. The table also shows the group assignment of the participants in the study. Participants in different groups were presented with different stimuli in different orders.

Table 3.1

Participants' Background Information

Participant[a]	Assigned Group	TEPS Scores	Major[b]	Status	Level of Prior Knowledge[c]	
					EI	ES
Andy	1	917	Management	Graduate	1	1
Sophie	1	873	Psychology	Undergraduate	1	1
David	2	931	Economics	Undergraduate	1	1
Tina	2	865	Management & Economics	Undergraduate	1	1
Clara	3	912	Economics	Undergraduate	2	2
Jessica	3	830	Economics	Undergraduate	2	1
Aaron	4	900	Economics	Graduate	1	1
Marion	4	809	Management	Undergraduate	1	1

Note. TEPS = Test of English Proficiency Developed by Seoul National University, EI = Environmental Issues, ES = Earth Science.

[a] All names are pseudonyms.

[b] The participants were asked to specify their minors or double-majors.

[c] These are responses to the five-point Likert scare survey questions "Check the degree of your knowledge for each of the following fields of study – environmental issues (EI)," and "– Earth science (ES)." (1 = *know nothing* ↔ 5 = *know very well*)

3.2 Instruments

3.2.1 Lecture Scripts

After the participants were recruited, the online questionnaire data were taken into account in developing the listening stimuli. Since all of the selected participants checked that they were unfamiliar (they checked either 1 or 2 out of the 5 scales of knowledge levels) with the fields of Earth science and environmental issues, lectures dealing with those two fields were selected to be developed into stimuli. These two lecture scripts were excerpted from commercial TOEFL preparation books (Hackers Language Research Institute, 2015). One lecture was about methods to study ancient weather patterns (Earth science) with about 250 words or 36 idea units, and the other lecture was about ways to reduce the amount of atmospheric CO_2 (environmental issues) with about 300 words, or 46 idea units. An idea unit is a semantic analysis unit (Foster, Tonkyn, & Wigglesworth, 2000), and a more detailed discussion will be provided in subsection 3.4.1.

3.2.2 Listening Stimuli

For the study, two video formatted lecture stimuli and two audio formatted lecture stimuli were needed. To develop a video stimulus out of each lecture script, a native speaker actress was recruited. The lecturer was a Canadian female in her twenties, and had a master's degree in applied linguistics. She did not memorize the script for the sake of the naturalness (refer to Appendix 2 for

the scripts of lectures). To induce authentic eye contact and hand gestures on the part of the lecturer, three listeners were present as audience during the video recording (as suggested in Cubilo, 2011). The lecturer used the board to write important key words such as “leafy plants” or “ancient tree trunks,” or to draw pictures (e.g., growth rings in a cross section of tree trunk). Some realia (e.g., a rock and two types of leaves) was also used to visualize important concepts of the lecture. The lecturer was video recorded from her waist up. A number of hesitation and false starts were present and silences of about 11 to 12 seconds long could be observed during her writings on the board just as in authentic lectures. The camcorder (SONY FDR-AXP55) and the tripod (Sachtler ACE M GS 1002) assured high quality of video stimuli. To avoid the insertion of content-unrelated or unintended noise in the video, and to free the lecturer’s both hands, a wireless microphone (SONY UWP-D11) was used.

For the audio stimuli, the lectures were audio-recorded anew instead of extracting the audio files from the video stimuli. Extracting audio files from the video stimuli could be simpler, but the long silences in the video stimuli that were created during the board writings could fatigue the audio stimuli listeners (Shin, 1998) because they could not see the cause of these long silences. Therefore, the lengths of two types of stimuli on the same topic were not exactly the same. Rather, the video stimuli were longer than their audio stimuli counterparts. The features that made video stimuli longer were the presence of silences, and the slowing down of the speech observed when the lecturer spoke and performed simultaneously (e.g., demonstrating realia). Unlike in the video

stimuli, there were no pauses that exceeded ten seconds in the audio stimuli because the lecturer did not need to write on the board. Still, a number of short pauses that each lasted for about a second were created by hesitations and false starts. The number of these disfluency features was controlled to be about the same with the video-stimulated version of the lecture.

3.2.3 Post-test Interview Questions

In order to investigate the participants' processes and interactions with video stimuli, video-stimulated recall protocol was used in the post-test interview. Stimulated recall is commonly used to study the participants' cognitive processes during the oral language use (Gass, Behney, & Plonsky, 2013; Gruba, 1993, 1997). The video-recording of each participant's test-taking process was played back in the interview to help them remember and report their experience to the researcher in the post-test interview. The interview was a semi-structured one. That is to say, guiding questions were prepared in advance to guide the interview, as shown in Table 3.2. However, because these preplanned questions were only a guide, additional questions were improvised during the interview to delve into points that the interviewer (i.e., researcher) found worth investigating deeper (Magnusson, & Marecek, 2015). The interview was conducted in Korean to facilitate the participants' statements, but it was later translated and presented in English to support the points made in this research.

Table 3.2

Post-test Interview Questions (adapted from Cubilo, 2011)

1.	Which aspect of the video/audio stimuli helped you understand the lecture?
2.	Which aspect of the video/audio stimuli helped you speak?
3.	Which of the lectures did you prefer, the audio-formatted one or the video-formatted one?
4.	Why did you prefer the audio-/video-formatted test?

3.3 Data Collection Procedure

3.3.1 Study Approval

This study obtained the Institutional Review Board (IRB) approval from the university. The policy requires that the researcher obtain consent from every participant after discussing the purpose of the study, its possible risks, benefits, ways to ensure privacy, and also the associated procedures. Therefore, at the beginning of each online survey and examination, the consent forms were collected from the participants.

3.3.2 Test Administration

The current study's data collection procedure is summarized in Table 3.3. The purpose of this study was to compare the effects of two different types of stimuli on the participants' performance on integrated speaking tasks. To avoid research irrelevant variables, stimuli formats and stimuli contents had to be

counterbalanced via Latin square method (Batty, 2015; Cubilo, 2011; Cubilo & Winke, 2013). As shown Table 3.3, the current study involved four different tasks: two video-stimulated tasks each on ancient weather patterns (AWP) and atmospheric CO_2 (ACO_2), and another two audio-stimulated tasks each on the same topics. Table 3.3 also illustrates how these four tasks were assigned to each of the four groups of participants. Two participants were assigned in each group to make sure that data can be collected to address the research questions even if one of them happen to quit in the middle of the study.

Table 3.3

General Outline of the Data Collection

Preliminary screening		**Task 1**	**Task 2**	**Post-test interview (stimulated recall)**
To manipulate English proficiency and prior knowledge	Group 1	Audio AWP	Video ACO_2	To investigate how test-takers utilize each type of stimuli in the test
	Group 2	Video AWP	Audio ACO_2	
	Group 3	Audio ACO_2	Video AWP	
	Group 4	Video ACO_2	Audio AWP	

Note. AWP = Lecture on ancient weather patterns, ACO_2 = Lecture on atmospheric CO_2.

The examination took place in a university classroom. Each participant signed up for a different time slot and used a designated laptop computer during the test (see Appendix 3 for an example of the task prompt). The testing process was video-recorded to be used in the post-test video-stimulated recall interview as explained in the Section 3.2.3. The test-takers' face, their note-taking, and the stimuli lectures playing on the laptop screen were all recorded simultaneously so that the researcher and the participants could see what he or she did in what moment or scene when taking the integrated speaking tasks. Figure 3.1 demonstrates a frame of this video recording.

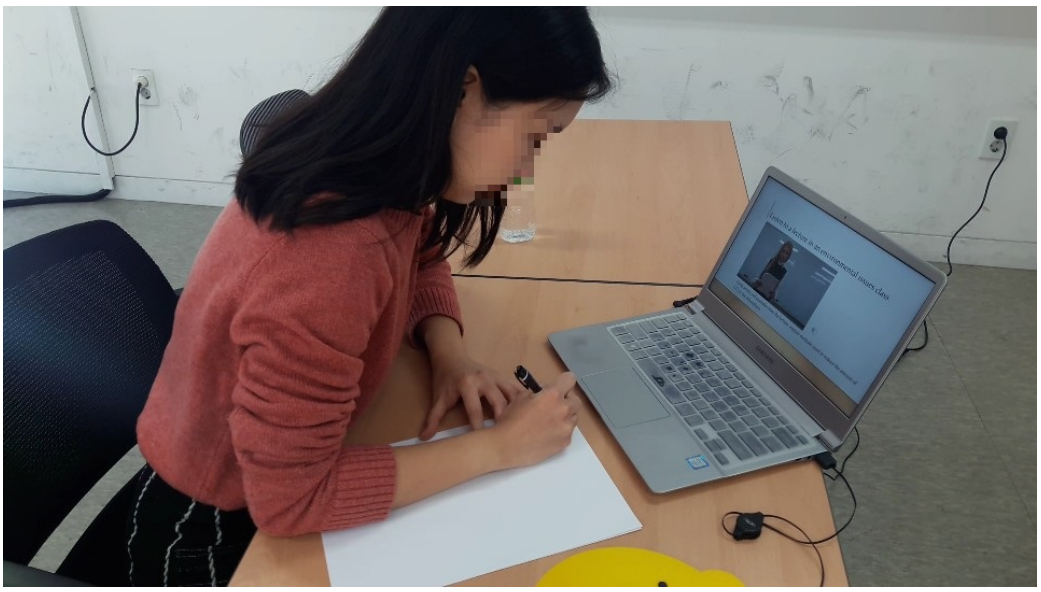


Figure 3.1. Video Recording of the Participants' Test-taking Procedure

After the listening-speaking tests were completed, individual post-test interviews were conducted in the same place. When needed, participants had the

chance to listen to or watch the other format of the stimulus in the post-test interview. For instance, the participant who listened to the audio-stimulated version of the EI lecture in the test could watch the video-stimulated version of the same lecture in the post-test interview when necessary. These interviews were also recorded and then transcribed for qualitative analysis.

3.4 Data Analysis

In this study, two types of data were analyzed. The first type was the summary data that were collected from the participants' listening-speaking tests. The second type was the post-test interview data. The following sections illustrate how those two types of data were analyzed.

3.4.1 Analyzing Summary Data

Given that the focus of Research Question 1 was on how the content of the lecture was summarized by the test-takers rather than on the other linguistic features of performance (i.e., grammar or intonation), the source texts (the lecture scripts) and the participants' summary data were compared semantically by dividing the source lectures and the participants' summary data into idea units. The segmentation of idea units was assisted by an experienced researcher. The two coders discussed their discrepancies in idea unit identification to finally reach a 100% agreement.

This study conducted idea unit analysis because its process was more reliable than the process of proposition analysis which was also widely used in summary discourse analysis (Foster et al., 2000; Frost et al., 2011; Kintsch, 1990;

Kintsch & van Dijk, 1975, 1978, Turner, 1987). However, the original definition of idea unit (Crookes, 1990; Foster et al., 2000) was modified to be applied in the current study. Earlier Kroll (1977, as cited in Crookes, 1990) defined an idea unit as having one of the following forms:

(1) A subject and a verb together with (when present) a (a) direct object, (b) prepositional phrase, (c) adverbial element, or (d) mark of subordination

(2) A full relative clause when the relative pronoun is present

(3) A phrase which occurs in sentence initial position followed by a comma or which are set off from the sentence with commas

(4) A verb whose structure requires or allows a verbal element as object (both verbal elements are counted as one idea unit)

(5) A reduced clause in which a subordinator is followed by a non-finite verb element

(6) A post-nominal *–ing* phrase used as modifier

(7) Other types of elements counted as idea units are (a) absolutes (b) appositives, and (c) verbals.

In the list above, (1) was redundant in itself because (b) a prepositional phrase could also be (c) an adverbial element as in "It takes long time *in the natural environment*." In the case of (2), there was no reason to restrict the case to "when the relative pronoun is present." Both of the two sentences "Leaves [*that*

have jagged edges] are common" and "Leaves [*having jagged edges*] are common" have two idea units regardless of the presence of a relative pronoun. Condition (3) was also redundant, and orthographical rather than grammatical. For instance, the condition (3) overlapped with (1) or (7a) as in the cases of "*When rain strikes rock*, / a chemical reaction happens and rocks absorb the CO_2" (main clause with an adverbial element) or "*Weather permitting*, / we shall meet in the evening" (a main clause with an absolute). In addition, (3) was inconsistent in that it cannot identify the two sentences such as "A chemical reaction happens and rock absorb the CO_2 / *when rain strikes rock*" and "*When rain strikes rock*, / a chemical reaction happens and rock absorbs the CO_2" as having the same number of idea units. Condition (5) was also redundant. A subordinator usually begins with an adverbial element as in "The plant fossils are useful / *when trying to understand prehistoric climate conditions,*" which can also be segmented by conditions (7a) or (4). Finally, conditions (4) and (7c) were conflicting because (4) claimed that a verb and its verbal element were together included in one idea unit while (7c) showed that a verbal element should be separated from its verb.

As explained in the above paragraph, the definition of Kroll's (1977) idea unit presented in Crookes (1990) not only had redundancy and incongruence, but also had some deficiencies. Frost et al. (2011), therefore, raised the need to add a condition segmenting coordinated elements: coordinated verb phrases; coordinated nouns or noun phrases; and coordinated independent adjectives connected to a common verb phrase. For instance, the sentence "We can see the shape of leaf / *and* tree trunk," consists of two ideas, that we can see the shape of leaf, and that

we can see the tree trunk. Without adding the new condition that Frost et al. (2011) suggested, coordinated items would not be segmented into separate idea units.

To add to Frost et al.'s (2011) suggestion, another modification was needed in the current study to differentiate prepositional phrases used as an object complement, as an adverb, or as an adjective. Those used as an object complement or as an adjective needed to go with its verb just as any other complements or adjectives did. Only the ones used as an adverb needed to be separated from their main clause.

The following list of conditions that segment idea units, summarized in Table 3.4, was finally decided to be used in the present study. The list was created by critically reviewing the previous frameworks presented in Crookes (1990) and Frost et al. (2011) and making additional modifications to their frameworks as discussed above. In the example sentences, in Table 3.4, an inserted idea unit is separated from the container idea unit by brackets.

Table 3.4

Modified Conditions that Segment Idea Units (Crookes, 1990; Frost et al., 2011)

(1) A main clause is an idea unit.

(2) A verb whose structure requires or allows a verbal element (e.g., gerund, to infinitive) as object complement is counted with both verbal elements as one idea unit (e.g., One method is *to create artificial trees*.).

(3) A full relative clause that

(a) post modifies a noun (e.g., Leaves [*that have jagged edges*] are common.)

(b) adds information (e.g., He was usually late, [*which always annoyed her*].)

(4) Reduced clauses such as

(a) a participial phrase (e.g., [*Commonly referred to as petrified wood*], these fossils offer a wealth of data.)

(b) a relative clause without a relative pronoun (e.g., There are two methods [*the lecturer explained*].)

(5) An adverbial clause or an adverbial phrase is an idea unit. An adverbial clause is a dependent clause that functions as an adverb. It modifies a verb, adjective or another adverb. There are several types of adverbial phrases and clauses that can indicate time, place, manner, purpose etc.

(e.g., A chemical reaction occurs *when rain strikes rocks*.).

(a) Prepositional phrase used as an adverb, not as an adjective or a complement (e.g., Leaves *with jagged edges*– used as adjective; Put the ball *on the table*. – used as complement; These plants were common *in cooler climates.* – used as adverb).

(6) Appositives (e.g., We can see the tree trunk, [*especially the petrification pattern*], to see the prehistoric climate.)

(7) Coordinated elements such as

(a) Coordinated verb phrase (e.g., The chemical substance *pulls the* CO_2 *out of the air* / and *stores it in the artificial tree*)

(b) Coordinated nouns and noun phrases connected to a common verb phrase (e.g., We can see *the shape of a leaf* / and *tree trunk*.)

(c) Coordinated independent adjectives connected to a common verb phrase (e.g., The city would be *less noisy*, *less polluted*)

Based on the conditions presented above, the stimulus lectures and participants' summaries were segmented into idea units.

Next, the ideas in each of the source lectures were arranged into a structural map, as demonstrated in Table 3.5 and 3.6, respectively. The levels in those maps show the macro- and microstructures of the texts. Concepts from Level

1 to Level 3 can be seen as belonging to the macrolevel of the discourse because they contain generalizations of textual details and important ideas. Concepts in Level 4 could belong to the microlevel that consists of concrete details (Kintsch, 1990; Kintsch & van Dijk, 1975). A summary consisting of Level 1 concepts would be very rough. On the other hand, a summary containing Level 4 contents would rather be a reproduction than a summary. To analyze the quality of the summaries, the idea units of the summaries were compared with its source texts' map for two aspects: which level in the text map the idea unit belonged to; and whether the idea unit in the summary contained (in)correct information.

In the case of Frost et al. (2011), idea units were classified into three categories: (1) correct replication, (2) distortions, and (3) others. The first two categories were each divided into three subcategories. The correct replications were subdivided into "replication," "combination," and "generalization." The incorrect replication was classified into "inaccurate replication," "inaccurate combination," or "inaccurate generalization." In van Dijk (1980), a more detailed classification was presented. He provided six different categories to classify the differences between the original text and the reproduced text: (1) deletion, (2) addition, (3) permutation, (4) substitution, (5) recombination, and (6) level shift. "Deletion" meant that some information was left out in the process of the reproduction, while "addition" meant some supplementation of information. The data were classified as "permutation" when the order of same level information was changed, and as "substitution" when information was substituted for another information. Finally, the category "recombination" was given when there was

information recombined in other units, and "level shift" when the information high up in the hierarchy was downgraded or low-level information was upgraded. In the current study, the categories used in van Dijk (1980) was adopted because this procedure seemed more detailed to understand the cognitive processes that test-takers underwent during the test compared to the Frost et al.'s (2011) procedure.

Table 3.5

The Macro- and Microlevels of the Lecture on Earth Science: Ancient Weather Patterns (AWP)

Level 1			
0[a] Plant fossils provide information about AWP.			
Level 2			
1 Scientists can study the average annual temperature using the shapes of the fossilized remains of leafy plants		**2** Scientists can study the past temperature and precipitation patterns using the growth rings of the petrified wood (ancient tree trunks)	
Level 3			
1.1 The majority of plants had smooth edges = global temperatures likely higher than average.	**1.2** The majority of plants had jagged edges = cooler climate	**2.1** Widely spaced rings = enough precipitation, and benevolent temperature	**2.2** Narrow rings = extended period of draught, and harsh temperature

Level 4			
1.1.1 Plants with smooth edges retain moisture better = advantage in warm climate	**1.2.1** Plants with jagged edges perform photosynthesis better = advantage in cooler climate (short growing season each year)	**2.1.1** Widely spaced rings = quick growth = enough precipitation and benevolent temperature	**2.2.1** Narrow rings = slow growth = the tree experienced an extended period of draught, and harsh temperature

[a] Each cell in the table is numbered to facilitate the discussion of the analysis results, which will be presented in Chapter 4. In this table, the most general idea that introduces the two methods of studying ancient weather patterns was numbered '0' in Level 1, so that in the next levels ideas related to method 1 begin with number '1' and ideas related to method 2, with number '2'.

Table 3.6

The Macro- and Microlevels of the Lecture on an Environmental Issue: Atmospheric CO_2 (ACO_2)

Level 1	
0[a] There are a couple of methods used to reduce the amount of atmospheric CO_2. (can mention the alarming level of atmospheric CO_2)	
Level 2	
1 Reduce the amount of atmospheric CO_2 using artificial (plastic) trees	**2** Reduce the amount of atmospheric CO_2 with enhanced weathering

Level 3			
1.1 Artificial trees reduce atmospheric CO_2 by pulling it out of the air (mechanism)	**1.2** Plastic trees are more efficient (faster) and permanent than natural trees.	**2.1** Weathering occurs when rain strikes rocks on the ground. The rocks absorb CO_2 from the air.	**2.2** Scientists enhanced weathering effect to maximize the chemical reaction
Level 4			
1.1.1 A special type of plastic is made with a chemical compound that pulls CO_2 out of the air. Artificial tree is made out of this special plastic.	**1.2.1** Plastic trees absorb CO_2 a thousand times faster than the natural trees	**2.1.1** However, when this takes place naturally, it takes a very long time to absorb large amounts of CO_2	**2.2.1** Scientists dig up rocks such as olivine. They grind the rocks into powder to maximize surface area and to speed up the reaction. (speed)
	1.2.2 Plastic trees store carbon permanently while natural trees release CO_2 back in the air when they die		**2.2.2** They spread the powder over the land. When rain hits the powder, a large amount of CO_2 is absorbed and stored permanently. (quantity, efficiency)

[a] These numbers can be interpreted in the same way as in Table 3.5.

3.4.2 Analyzing Post-test Interview Data

From the transcription of the post-test interview data, coding was processed. The repeating ideas were organized into "themes" that were next organized into more abstract groupings called "qualitative findings" (Auerbach & Silverstein, 2003). One example process is illustrated in the following:

> *David: The lecturer showed some pictures and she made the contrast very clear.*[1]
>
> *Jessica: The pictures helped me understand the lecture. At first, I didn't get the concept, but with the pictures, I got the overall idea of what she was talking about.*
>
> *Aaron: She visualized what she was explaining so it was easier to conceptualize that part of the lecture in my mind.*

Those repeating ideas in the raw transcription of interview data were organized into a theme, "features of the video stimuli that facilitate the comprehension of lectures." In this way, the grouping of the repeating ideas generated 6 themes in total. These themes were in turn arranged into qualitative

---

[1] In the quotes or excerpts that are presented in this study, the italicized words are English translations of Korean utterances. Words that are not italicized are utterances originally delivered in English.

findings. For instance, features of each stimulus that “help with comprehension” of lectures together with other themes such as “help with production planning,” “allow time to spare” and “help with checking of the accuracy of understanding” were arranged into a qualitative finding, “the facilitating effects of video stimuli in a listening-speaking integrated test.” In total, 4 such final qualitative findings could be identified. In section 4.2, each theme is described to bridge the gap between the participants’ subjective experiences and the current study’s qualitative findings to provide an answer to Research Question 2.

The participant recruitment, data collection, and analysis processes were implemented as described in this chapter. The following chapter presents the results of the study.

Chapter 4. Results

This Chapter is divided into two parts to deal with each of the two research questions addressed: (1) how test-takers' speaking performances differ when provided with stimuli of different format; and (2) how test-takers utilize the stimuli to perform in the listening-speaking tests. In 4.1, the findings about the test-takers' summary data are presented and in 4.2, the post-test interview data are reported.

4.1 Test-takers' Language Performance Results

In total, the test-takers produced 16 summary texts in the listening-speaking integrated tests. Before the idea unit analysis, these summaries' general characteristics were quantitatively compared by using the software Coh-metrix. Tables 4.1 and 4.2 provide the quantitative description of the test-takers' summaries and the source texts on topics AWP and ACO_2 respectively. First, "the number of words" in each text was counted to compare the length of the texts. However, the result of the comparison could not ensure whether the types of stimulus influenced the length of the summaries. The second index, "the number of sentences" in each text was another measure of the text length. "The number of sentences" complemented "the number of words" because the former could inform the approximate amount of information each text had (Foster et al., 2000). The analysis of "the number of sentences", however, could not assure whether the types of stimulus affected the amount of information. Third, "the mean

number of words in a sentence" was calculated for each text since it was considered as a rough estimate of the text's syntactic complexity. The more words a sentence had, the more complex it could be. For the syntactic complexity, it could not be concluded whether audio conditions or video conditions caused the production of syntactically complex summaries. However, it was likely that the summaries were syntactically more complex than the source texts. Finally, the type-token ratio analysis results suggested that the vocabulary used under all conditions tended to be restricted.

Table 4.1

Coh-metrix Analysis Results for AWP Summaries and Source Text

Index	**Audio Summaries**		**Video Summaries**		**Source text**
	Mean	**SD**	**Mean**	**SD**	
Number of words	183.0	50.9	223.0	26.8	245.0
Number of sentences	9.3	3.1	8.0	1.9	14.0
Mean number of words per sentence	20.8	3.6	29.9	9.6	17.5
Type-token ratio	0.5	0.1	0.5	0.0	0.6

Table 4.2

Coh-metrix Analysis Results for ACO_2 Summaries and Source Text

Index	Audio Summaries		Video Summaries		Source text
	Mean	SD	Mean	SD	
Number of words	220.5	33.5	206.3	33.1	319.0
Number of sentences	9.8	1.1	12.5	2.1	20.0
Mean number of words per sentence	22.5	1.2	16.8	3.4	16.0
Type-token ratio	0.5	0.0	0.5	0.0	0.5

After the quantitative characteristics of the texts were overviewed, test-takers' summaries were analyzed at three different levels: macro-, micro- and word levels. The macrolevel analysis dealt with the macrostructural differences between the stimuli lectures and summary responses. Macrolevel ideas of the stimuli lectures were presented from Level 1 to Level 3, as shown in Tables 3.5 and 3.6. Subsequently, more detailed content features of the test-takers' summaries were compared with the microlevel ideas of the stimuli lectures belonged to Level 4 of Tables 3.5 and 3.6. The word level analysis was distinguished from the microlevel one in that it presented an analysis more focused on the test-takers' vocabulary use.

4.1.1. Macrolevel Analysis

The participants' summary discourse analysis revealed that macrolevel errors were present in the audio formatted tests. That is, the participants omitted some macrostructural content in their summary during the audio-stimulated tests. However, as Table 4.3 shows, no macrolevel errors were found in the video formatted tests. Only an addition of a correct inference could be identified in the macrolevel analysis of the video-stimulated tests.

Table 4.3

Macrolevel Discrepancies of the Idea Units between the Stimulus Lectures and the Test-takers' Summaries

Lecture	Video-stimulated Test	Audio-stimulated Test
AWP	▪ Add 2.1[a] (only inferred in the stimulus) (Clara)	▪ Add 2.2 (only inferred in stimulus) (Aaron) ▪ Delete 1.2 - 1.2.1 (Marion)
ACO_2	None	▪ Delete 1.1 - 1.1.1 (David) ▪ Shift levels: present Level 1 as 1.1 - 1.2 - 2.1 instead of 1 - 2 (Jessica) ▪ Delete 2.1, 2.1.1 (Tina)

Note. AWP = Ancient Weather Patterns Lecture, ACO_2= Atmospheric CO_2 Lecture

[a] These numbers refer to the ideas in the cells of Table 3.5 and 3.6. For instance, number 2.1 in the AWP lecture indicate the content in the cell numbered "2.1" in Table 3.5.

All of these macrolevel mistakes would have been avoidable if participants had been provided with some visual supports during the test. One example of a macrolevel error was found in Jessica's audio-stimulated test outcome as underlined in Table 4.3. Jessica thought the lecturer was presenting three methods that were used to reduce the atmospheric CO_2: (1) planting artificial trees, (2) planting natural trees, and (3) using enhanced weathering. She confused the microlevel contents with macrolevel ones. Figure 4.1 shows Jessica's notes where she numbered three methods that she thought were presented by the lecturer.

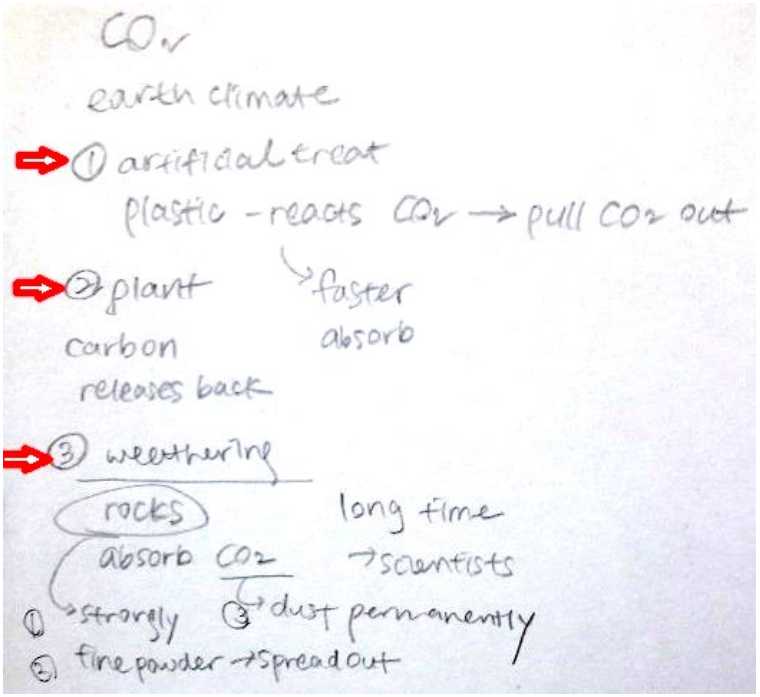


Figure 4.1. Jessica's Note of ACO_2 Lecture

Jessica would not have made such level shift errors if she had been provided with the board works as in the video version where they were mainly used to visualize macrolevel information. This supposition is supported by Jessica's interview data. During the interview session, Jessica watched the video version of her audio stimuli (ACO_2) and gave the following comments:

> *Jessica: If I had watched the video version of this lecture (ACO_2) [in the test], I might not have made such mistakes.*
> *Interviewer: Such mistakes?*
> *Jessica: I thought the lecturer was talking about three* "methods"*, but if I had seen the video, I would have understood the lecturer's points better (...) The lecturer wrote some points on the board, and these points helped me a lot. I think overall, I understood better in the video version (stimulated recall interview).*

Jessica expressed that the writings on the board helped her conceptualize what the lecturer was explaining, and that they helped her visualize which part of the lecture she was listening to. She related the lack of these visualizing effects in the audio stimuli as a potential cause of her misunderstanding.

As mentioned in Table 4.3, Marion demonstrated a macrolevel mistake that was comparable to Jessica's. Table 4.4 shows the differences in the macrostructures of the original lecture (AWP) and its summary produced by Marion. In her audio formatted test (AWP), Marion omitted a macrolevel

concept, information in 1.2 cell in Table 3.5, which was about the relations between a shape of leaves and the temperatures of the area the tree lived. Unlike the microlevel information in 1.2.1, this macrolevel idea was the key information that had to be summarized in the response (Kintsch & van Dijk, 1978; van Dijk, 1980). However, Marion had difficulty understanding this part of the lecture and could not even jot down anything about it in her note. In her interview, she confessed that she could not explain the concept even in her L1, Korean, which proved that she did not understand the concept. If Marion had seen the pictures of leaves as in the video version, she might have comprehended the functions of the leaf and could have possibly mentioned it in her summary.

Table 4.4

Comparison of Macrostructures between the Source Text and Marion's Summary

Original lecture (AWP)	Marion′s summary (AWP)	Analysis of Marion′s summary
0[a]	0	
1	1	
1.1		→ Deletion of 1.1
1.1.1	1.1.1	→ Deleted 1.1, replaced by 1.1.1
1.2		→ Deletion of 1.2
1.2.1		
2	2	
2.1.1	2.2.1	
2.2.1	2.1.1	
2		→ Deletion
2.2		→ Deletion
	0	→ Addition

Note. AWP = Ancient Weather Patterns

[a] These numbers refer to the ideas in the table 3.5. For instance, number 0 in the leftmost column indicate the content in the cell numbered "0" in Table 3.5.

Similar macrolevel errors occurred in David and Tina's outcomes. Table 4.3 indicates that these two participants each omitted a macrolevel structure of the lecture. First, David did not mention 1.1 part of ACO_2 lecture as shown in Table 4.5. Part 1.1 was about the mechanism of artificial trees and had to be mentioned because it belonged to the macrolevel information of the original lecture. Similarly, Tina omitted 2.1 part of the same lecture which was about the mechanism of weathering (for the concepts numbers 1.1 or 2.1 are referring to, see Table 3.6). Considering the fact that a good summary reflects the original texts' macrostructure (Kintsch & van Dijk, 1978; van Dijk, 1980), the omission of these macrolevel ideas corresponded to some defects in the summaries.

Table 4.5

Comparison of Macrostructures between the Source Text and David's Summary

Original lecture (ACO_2) macrosturcture	David′s summary (ACO_2) macrostructure	Analysis of David′s summary
0	0	
	1	→ Addition
	2	→ Addition
1	1	
1.1		→ Deletion of 1.1
1.1.1		

1.2		→ Deletion of 1.2
1.2.1	1.2.1	→ Deleted 1.2 replaced by 1.2.1 and 1.2.2
1.2.2	1.2.2	
2	2	
2.1	2.1	
2.1.1		
2.2		
2.2.1	2.2.1	
2.2.2	2.2.2	
	2.2	
	0	→ Addition
	1	→ Addition
	2	→ Addition

Note. ACO_2= Atmospheric CO_2

[a] These numbers refer to the ideas in the table 3.6. For instance, number 0 in the leftmost column indicate the content in the cell numbered "0" in Table 3.6.

4.1.2 Microlevel Analysis

Microlevel errors are local errors related to micro-level information, Level 4 in Tables 3.5 and 3.6. Although inclusion of any of these Level 4 information is considered not necessary in a good summary, microlevel errors were analyzed if there were any. Test-takers' performance differences at the microlevel could be the proofs of any hindering effects of a stimulus. Unlike the macrostructural errors that appeared only in the audio formatted test results, microlevel errors appeared in both types of tests. Table 4.6 shows the microlevel differences of the idea units between the stimulus lectures and test-takers' summaries. The differences in the table could be microlevel errors, or any positive changes made by the test-takers. Among the items enlisted in Table 4.6, only the errors that seemed to be related to the effects of stimuli types on test-takers' performance were underlined and further discussed in this section.

Table 4.6

Microlevel Discrepancies of Idea Units between the Stimulus Lectures and the Test-takers' Summaries

	Video Stimulus	Audio Stimulus
AWP	▪ Relate irrelevant ideas: "absorb a lot of humidity inside the leaf so it's rainy season" (Jessica) ▪ Add inferred idea: "Scientist use tree rings to study whether the past trees grew in a cooler temperature or higher than average" (Tina) ▪ Permute as 2.2.1 - 2.1.1 (Clara & David)	▪ Relate irrelevant ideas: "growth rings shape the temperature" (Aaron) ▪ Use irrelevant background knowledge: "because it is dark, the shape of leaf can be used to guess moisture" (Marion) ▪ Relate irrelevant ideas: in "we can use shape of leaf to guess moisture," "moisture" should be "temperature" (Marion)
ACO_2	None	▪ Relate irrelevant ideas: provide evidence of efficiency as evidence of permanence (Tina) ▪ Describe "olivine" as artificial rock (Tina) ▪ Relate irrelevant ideas: compare natural trees with artificial trees about "permanence" instead of "speed" (Tina) ▪ Permute as 1.2.2 - 1.2.1 (Jessica)

Note. AWP = Ancient Weather Patterns Lecture, ACO_2= Atmospheric CO_2 Lecture

This study's microlevel investigation, just as macrolevel investigation did, showed that test-takers could perform better under the influence of video stimuli than of audio stimuli.

A microlevel error with some post-test interview data supported the fact that audio stimulus could complicate the use of information that was explicitly mentioned in the lecture. The following quote is an excerpt of the original ACO_2 lecture which Tina listened to during her audio-stimulated test. The first two lines of the following quote presents how the lecturer explained the concept "efficiency."

> "One advantage of using the plastic is that they absorb CO_2 a thousand times faster than real ones. Another plus is that the carbon storage is permanent. As you might know, when real trees die and rot, carbon is released back into the atmosphere. But, obviously, this isn't an issue with plastic trees" (from ACO_2 lecture).

The lecture explained that the efficiency was related to the plastic trees' speed of absorbing CO_2. The natural trees' lifespan was mentioned to explain artificial trees' permanent ability to absorb CO_2, not their efficiency. Excerpt 4.1 demonstrates that Tina explained the concept "efficiency" differently.

> **Excerpt 4.1**: The reason why this author, the lecturer explains the use of artificial trees would be its efficiency. Since authentic trees would,

decade after its death, so by using plastic trees it would help absorbing the CO_2 in a more efficiently, and more helps with longevity (Tina, audio-stimulated test, ACO_2).

Tina explained the plastic trees' efficiency with the evidence of their permanence. The original lecture's information about the speed of absorption disappeared in her summary. In return, Tina dealt with permanence only shortly, by using the new expression, "longevity." If she had heard the sentence "they absorb CO_2 a thousand times faster than real ones" in the stimulus, she might have been able to provide the correct evidence for efficiency.

Some interview data, however, suggest that this part of the lecture could be somewhat difficult to listen to as the two important concepts, "efficiency" and "permanence", successively appeared with only a short interval in between.

Clara: (After watching the video version of the ACO_2 lecture that she experienced in audio during the test) Now I realized that I could not hear this part [during the audio-stimulated test].
Interviewer: Which part?
Clara: The name of the rock, and the part where she explains how faster artificial trees are (stimulated recall interview).

Clara's interview added support to the fact that the evidence for plastic trees' efficiency was somewhat difficult to catch. Other test-takers' interview

data suggested that such missing of information was quite frequent in the audio-stimulated test because test-takers were too busy listening and taking notes at the same time.

> *Andy: In the audio-stimulated lecture, when I could not clearly understand, I just abandon that part because there is the next information explained. I take note of the new information a little bit, but cannot finish writing because the next information is coming out and so on (stimulated recall interview).*

> *David: In the video version, I can take some time to take note but in the audio, I try to write something, but I have to skip writing because there is not enough time.*
> *Interviewer: Why is note-taking important to you?*
> *David: I write down because I have to remember [those ideas in order to speak] (stimulated recall interview).*

In the post-test interview, Andy and David explained the unavoidable nature of information abandoning. The phenomenon occurred only in the audio-stimulated tests because of the audio stimuli's characteristic of delivering information more successively than the video stimuli.

To compensate for such missing information, some test-takers tended to bring their own background knowledge. Excerpt 4.2 is drawn from Marion's

audio-stimulated AWP lecture summary.

> **Excerpt 4.2:** Um for example, um, we, leaf, ah shape of leaf have ah, it can be used to guess moisture, *No*, can be used to standard the temperature, <u>because it's dark, ah leaves dark</u>, it shows, *No*, can be used to standard, to guess moisture (Marion, audio-stimulated test summary, AWP).

It seems that Marion tried to explain one type of leaves' strength in performing photosynthesis. In the process, she inserted her own background knowledge that was irrelevant to the explanation, "because it is dark." Although the presence of light is involved in the photosynthesis process, it did not have any relevance with the current lecture content. Her interview data suggest that this use of background knowledge was not to compensate for her limited speaking skill, but primarily to make up for her problem in understanding.

> *Marion: If someone asked me to solve a multiple choice question about this lecture maybe I might be able to answer. However, if someone asked me to explain the content [of AWP lecture that I listened to in the audio-stimulated test], I wouldn't be able to explain. I would say to that person that I couldn't understand. Even if the person asks me to explain in Korean I wouldn't be able to. Because I could not understand, I could not speak well either. There were a lot of hesitations, repetitions and it took a*

lot of time for me to speak (stimulated recall interview).

As the above excerpts and quotes illustrated, audio stimuli had some hindering effects on the test-takers' performance, and sometimes even caused the incorporation of test-takers' own background knowledge.

On the other hand, in the video-stimulated test-taking, information abandonment could not be found. The interview data revealed that a feature in the video stimuli prevented this information abandonment.

> *Andy: [In the video stimuli] because there were pauses from time to time, I could rearrange information I heard, I could add some notes that I could not take during the lecturer's speech (stimulated recall interview).*

The interview with Andy suggested that the pauses in the video stimuli could reduce the information abandonment: the pauses could slow down the lecture delivery, and provided time to digest and restructure what was learned from the lecture.

In this section, microlevel analysis results were presented to illustrate the hindering effect of the audio stimulus. In the audio-stimulated tests, test-takers tended to miss information and use their own background knowledge to supplement this missing information, however, in the case of video stimuli, no such error could be found. The reason for this difference could be the result of the stimuli differences. Audio stimuli did not facilitate recognition of lecture

contents because the concepts were introduced successively. Test-takers who were occupied by their note-taking tended to miss a lot of important information. However, in the case of video stimuli, test-takers were provided with sporadic pauses that were made during the lecturers' writings on the board. During these pauses, test-takers tended to finish their note-taking and prepare for further listening.

In conclusion of this microlevel analysis, audio stimuli seem to have an effect that hinders the full utilization of the stimuli. However, video stimuli seem to help participants to fully utilize the stimuli.

4.1.3 Word Level Analysis

A closer analysis informed that there were also some word level differences between audio formatted and video formatted test results. While the microlevel section investigated the sentence level errors, this section focuses more on the word level errors and participants' use of synonyms. Table 4.7 below summarizes the word level differences of idea units between the stimulus lectures and the test-takers' summaries. The differences presented in the table include word level errors (e.g., "misuse") and some positive changes made by the test-takers (e.g., "synonym"). The word level differences that seemed to be caused by the stimuli effects were underlined, and further discussed in this section.

Table 4.7

Word Level Discrepancies of Idea Units between the Stimulus Lectures and the Test-takers' Summaries

	Video Stimulus	Audio Stimulus
AWP	▪ Synonym: "prolonged" for "extended," and "water" for "moisture" (Clara) ▪ Synonym: "zigzag" for "jagged" (Jessica)	▪ Misuse: "we can see the petrification pattern" (Sophie) ▪ Synonym: "zigzag" for "jagged" (Aaron) ▪ Misuse: "examining shape of leaf is efficient in prehistoric conditions" (Aaron) ▪ Misuse: "fossilization is good source to study ancient weather" (Marion)
ACO_2	▪ Circumlocution: "rocks that has a strong reaction," "the rocks," "a rock" for olivine (Andy, Sophie, & Aaron) ▪ Synonym: "emit" for "release back" (Aaron) ▪ Misuse: "prominent" for "permanent" (Sophie)	▪ Circumlocution: "the rocks that is efficient," "certain rock" for olivine (David & Jessica) ▪ Synonym: "authentic" for "natural" (Tina) ▪ Misuse: "decade" for "decay" (Tina)

Note. AWP = Ancient Weather Patterns Lecture, ACO_2= Atmospheric CO_2 Lecture

First, Sophie showed her limited knowledge about the word "petrify" as she said "petrification patterns" to refer to the growth rings patterns of an ancient tree trunk:

Excerpt 4.4: And second of all, we can see the tree trunk, especially the petrification pattern to see the prehistoric climate (Sophie, audio-stimulated test summary, AWP).

If Sophie had seen the board work, she would have known that "petrified wood" meant ancient tree trunks. The board work that was presented during the video version of the same lecture is presented in Figure 4.2. In the picture, "ancient tree trunk" and "petrified wood" were related with an arrow. As there were no cues that could provide additional information about the meaning of the word in the audio formatted lecture, she might have guessed the meaning by herself, resulting in a word level error.

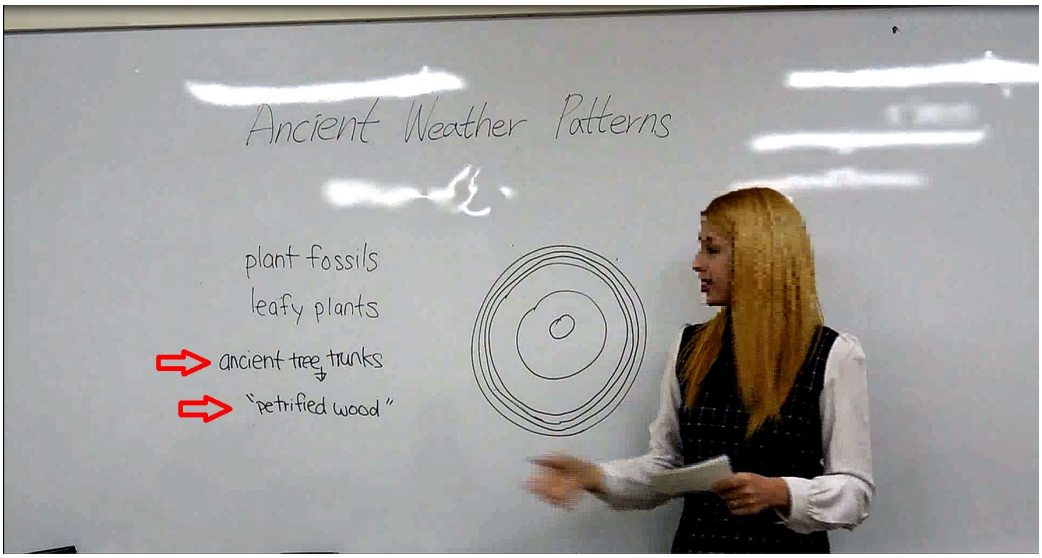


Figure 4.2. Writings on the Board in the AWP Lecture

Another point to be made is about the synonym used to refer to the shape of leaves that have jagged edges. Jessica, in her video-stimulated lecture, saw the image of jagged edges of leaves and referred to it as "zigzag edge leaves" in her summary. The word "jagged" is not commonly used among Korean EFL learners while "zigzag" is widely used even by the low-level English learners. This use of synonym proves that Jessica actually understood the concept of "jagged edges leaf" although she did not use the exact wording of the lecture in her speech. From this incidence, it could be concluded that visual cues facilitated conceptualization and helped test-takers overcome language barrier. Jessica' interview data and her note added credit to this supposition. She mentioned her previous ignorance of the expression "jagged:"

> *Jessica: I think the video version helped me better because the lecturer showed the spiky edges saying* "zigzag" *or something. I didn't know the word she used (...) I don't know if she referred to that form of leaf otherwise, but I would not have conceptualized the form of the leaf if she had not demonstrated it (stimulated recall interview).*

From the two participants' cases, it can be concluded that the presence of visual cues can facilitate the test-takers' understanding of words and their use of synonyms because visual cues enabled the participants' conceptualization of the lecture.

4.2 Post-test Interview Results

After the testing sessions were over, participants were asked interview questions to explore how they utilized each format of stimuli to accomplish the summarizing tasks. The interviews began with some icebreaking questions and then continued with preplanned questions presented in Table 3.2. In response, the interviewees pointed out some characteristics of the test formats and explained how they used them, and how they felt about these characteristics. Interviewees' responses were recorded and then transcribed to be qualitatively analyzed to seek for the answer for Research Question 2. Recurring ideas in participants' interview data could be sorted into six different themes that provided an insight for the qualitative findings of this study. Table 4.8 shows how these repeating ideas were organized into themes, and the themes, into qualitative findings. In the following subsections, each theme is explained to illustrate how test-takers' subjective experiences could provide answers for the current study's concern.

Table 4.8

Post-test Interview Data Arranged into "Repeating Ideas," "Themes," and "Qualitative Findings"

Repeating Ideas	**Themes**	**Qualitative Findings**	
		Video	**Audio**
▪ Introductory sentences ▪ Writings on the board ▪ Pictures and realia ▪ Content-relevant gestures ▪ Pauses	Help with comprehension	Yes	Yes
▪ Introductory sentences ▪ Writings on the board ▪ Pauses	Help with production planning	Yes	Yes
▪ Writings on the board ▪ Pictures and realia ▪ Content-irrelevant gestures	Cause cognitive load and distractions	Yes	Yes
▪ Missing information	Cause information abandoning	**No**	**Yes**
▪ Pauses	Allow time to rest	**Yes**	**No**
▪ Writings on the board ▪ Pictures ▪ Realia	Help with checking of the accuracy of understanding	**Yes**	**No**

4.2.1 Theme 1: Help with Comprehension

Participants gave various comments about the features that facilitated their comprehension of the lecture. Among the four features mentioned, only one was a common feature found in both types of stimulus: introductory sentence of the lectures. Others such as board works, pictures or realia, content-relevant gestures, and pauses were all features belonging to video stimuli only.

Using introductory sentences contributed to the comprehension of both types of stimulated lectures. The introductory sentences in the lectures are quoted in the following:

> "There are a couple of promising methods that can be used to accomplish this goal" (from ACO_2 lecture).
>
> "There are a number of methods used by scientists to study ancient weather patterns" (from AWP lecture).

When hearing them, test-takers could begin preparing for further listening. With the word "methods" in the introductory sentence, for instance, Tina reported that she could know the lecture had a list type frame: she was able to envisage a list of methods coming up. As she proceeded to the lecture listening, she could expect at what point to listen carefully to fill in this list type frame with relevant information. This cognitive process of organizing new information in a conceptual frame could be interpreted as the process of comprehension (van Dijk, 1980). In

this way, the test-takers began their comprehension process by utilizing the introductory sentences (Kintch & van Dijk, 1978).

The second feature that helped test-takers' comprehension, board work, was found only in the video stimuli. Participants explained that the board works facilitated an efficient comprehension.

> *Clara: In the video, the lecturer wrote down some keywords on the board. Those board works helped me (...) The first example is this, the second example is that. I could grasp the important concepts more effectively than in the audio version because the important concepts were given right there [on the board] (stimulated recall interview).*

Just as Clara explained, the writings on the board visualized the important concepts and facilitated test-takers' comprehension of the lecture.

In addition to the board works, the pictures and realia facilitated test-takers' comprehension. In the AWP lecture, there were pictures like jagged edge leaves and smooth edge leaves or even some realia like a rock. In the case of ACO_2 lecture, there were a tree trunk drawn on the board, and a diagram that described a rock absorbing the atmospheric CO_2. David commented that those pictures helped him understand and remember concepts more efficiently. To illustrate how pictures helped him understand better, he provided an example of his own major, economics.

David: When I study my major, economics, it is a lot easier to understand when I see a graph than when I have to read a long description or a complex formula. In the same way, I think pictures helped me conceptualize what the lecturer was saying (stimulated recall interview).

The above interview excerpt shows that David had a positive perception about the video stimulus especially because of the visuals such as pictures or realia.

According to these participants, content-relevant gestures were also facilitating. In the video lecture, the lecturer used diverse gestures. These gestures could be divided into two types: those closely related to the lecture and those irrelevant to the lecture. The first type of gestures included lecturer's hand gestures pointing and emphasizing the edges of two different types of leaves in the AWP lecture. Marion and David commented that this type of gestures helped them focus on the lecture because these gestures were highlighting the important concepts.

David and Marion also described how pauses in the video stimuli facilitated their comprehension. The pauses were created when the lecturer wrote down keywords on the board. They argued that these pauses were signals of the upcoming important information because the lecturer would not write down any trivial words on the board. Therefore, when a pause started, these two participants, as they described themselves in their post-test interview, began paying attention to the lecturer's actions and prepared to take notes.

4.2.2 Theme 2: Help with Production Planning

The post-test interview data suggested that introductory sentences and writings on the board had another function, which was to facilitate production planning. Participants were informed that they were to summarize what they were about to listen to before the lectures began with the task instruction. To manage testing time efficiently, these test-takers tended to plan their production during the listening.

The introductory sentences gave out the frame of the original discourse which facilitated the composition of their own summary. Test-takers all commented in the post-test interview that a good summary should conserve the original text's global structure (Kintsch & van Dijk, 1978), so they focused on the outline of the lecture that was revealed by the introductory sentences. After they found out the structural frame of the source lecture, all they had to do was to fill in the frame with some relevant information. However, as Tina mentioned, the introductory sentences in the current study did not indicate exactly how many methods would be presented, and this vagueness tended to complicate the test-takers' listening experience.

Unlike in the audio-stimulated test, in the case of video-stimulated test, other features of the stimuli facilitated production planning. Writings on the board were another facilitator. Some participants (Andy, David, Sophie, Marion, Sophie) answered that they could figure out the outline of the lecture with the help of the words written on the board. The board works visualized the hierarchy of the

concepts which would have been crucial in selecting the macrolevel information out of the microlevel information (van Dijk, 1980) to include in the summary.

Another unique feature of video stimuli that facilitated the production planning was the presence pauses. Pauses were created in video stimuli as the lecturer wrote down some keywords on the board. After reviewing his own test-taking during the stimulated recall interview, David could see himself whispering words during these pauses. He was trying to reorganize his own summary outline and practice speaking. For David, the pauses were some time to practice speaking.

4.2.3 Theme 3: Cause Cognitive Load and Distractions

Test-takers can have different perceptions about the cognitive load a test imposes on the test-takers because their language proficiency and their background knowledge affect their cognitive process (as mentioned in Kintsch & van Dijk, 1978). Although the amount of information in each lecture, as well as participants' English proficiency and prior knowledge were manipulated to be approximately the same, participants of this study gave different comments about the cognitive load the tests imposed on them. First, Marion felt that the audio stimulus (AWP) was more complex than the video stimulus (ACO_2).

> *Marion: In the audio formatted test, I had to decide whether to take note of a concept or not at the same time I processed the next one I hear. I was overwhelmed by this multitasking (stimulated recall interview).*

Marion knew that she did better in the video formatted test and was favorable to the video stimulus. Marion further explained that in the video-stimulated test, she could see the lecturer writing important keywords on the board which helped her distinguish macrolevel information from the microlevel information. However, in the case of the audio stimuli there was no information source other than the lecture's aural information, and Marion had to decide by herself whether a concept is important, whether to take note, and whether to include it in her summary or not. At the same time, she confessed that she was worried about missing some important concepts while she was occupied by her note-taking. All these thoughts that emerged during the audio-stimulated test made her feel cognitively overwhelmed.

Video stimuli were not without negative comments. As previous literature pointed out (Conaim, 2001; Cubilo, 2011; Cubilo & Winke, 2013), some participants complained about the video stimuli's distracting effects. Clara's statements in the following could reveal why participants would disfavor video stimuli.

> *Clara: [In the video-stimulated lecture,] the lecturer shows a picture, and that was the moment when I felt I had much more information to summarize [than in the audio-stimulated lecture] (...) Later, I realized that the pictures were only supplementary visuals [not providing any additional information] but at that time, I was worried about missing any important ideas that could be hidden in the visuals (stimulated recall interview).*

Clara's interview data informed that some test-takers could be skeptical about having visual cues because they could think that those visuals contain additional information which would complicate their summarizing. Tina had a similar opinion about the video stimuli. To Tina, board works represented cognitive load, too. Tina, who previously defined herself as a participant who preferred audio stimuli better, felt worried when she first saw the lecturer writing something on the board. She felt pressure because she thought the lecturer was adding some new information by writing. More information in the source text would have signaled a more complex summary to be made to her (Bejar et al., 2000). She would have liked it best if she could focus only on the audio stimulus without any visual input.

Andy and Aaron regarded the board works as redundant and therefore, distracting. Andy and Aaron had an impression that the lecturer was giving out redundant information by writing down and by delivering it orally again. What Andy wanted to do was to organize a summary simultaneously as he listed to the lecture. However, he argued that the visual supports inhibited this efficient time management. He could not organize what he was writing but copied the board works in his notes uncritically, without thinking where in his summary it would be included. The next quotes show Andy's and Aaron's opinions about the board works.

> *Andy: I could not understand why the lecturer wrote down easy concepts like* "artificial trees" *(stimulated recall interview).*

Aaron: If someone doesn't know the word "petrify," *the board writing* "petrified wood" *would help. But that was easy, at least for me (stimulated recall interview).*

For these two male participants, only audio input sufficed to understand the lecture.

Just as board writings, gestures also distracted some of the participants. Unlike the facilitating gestures presented previously, the content-irrelevant gestures were said to be distracting. Those negative gestures included the lecturer's personal habits like combing her hair back with her fingers, touching her clothes or face. Andy, David and especially Aaron pointed out that these actions distracted them from the lecture.

4.2.4 Theme 4: Cause Information Abandoning

Although evidence of some incomprehension could be found in both types of tests, a distinctive way of compensating for the incomprehension could be observed in the audio-stimulated tests only: test-takers used an abandoning strategy in the audio-stimulated tests. Participants abandoned the incomprehensible content and did not even jot it down in their note. They could have missed some parts of the lecture and would have tried to compensate for those missed parts (Salehzadeh, 2006) as in any other listening tests. However, having aural information only and no other sources (i.e., reading texts or visual cues) to refer to, it would not have been easy to supplement their deficiencies. As a

result, they had no other option but to abandon the missed part.

However, test-takers' opinions about this abandoning process differed. Andy expressed his relatively favorable stance on this experience.

> *Andy: In the audio version, there is no pause so there is not enough time to note what I hear. Therefore, by the time I had to speak, I had somewhat limited information. I just abandoned that part to make it easier to speak. (...) In the video version, however, I have to speak all of these things I noted. In the video version, it's more difficult to make a summary (stimulated recall interview).*

Andy did not like the video-stimulated test because he felt it had more information to include in the summary. As he was listening, he was concerned about his summary production (speaking) and readily abandoned the missed parts for the sake of his ease of production. However, Aaron or David was more concerned with this abandoning process than organizing his own summary when they were listening to the lecture. These two participants were more worried about the abandoning than Andy was.

> *David: In the audio version, I had no choice. I just had to skip that incomprehensible part. There was no way I could regain that part (stimulated recall interview).*

Aaron: One would want to take some time reorganizing a concept he hears, but in the audio version test, he can't because there is next information coming up right away. Some information processing is kind of left undone (stimulated recall interview).

In video-stimulated tests, the abandoned information could have been complemented to some extent, but in the audio-stimulated test, information abandonment was inevitable because there were no other resources than the lecturer's voice.

4.2.5 Theme 5: Allow Time to Rest

The presence of pauses was an idiosyncratic feature of the video stimuli because they were created when the lecturer was writing on the board. Test-takers all liked these pauses, but the reason they liked the pauses, and how they used them all differed. The roles of pauses were already described twice in subsections 4.2.1 and 4.2.2, but another use of pauses was repeatedly reported in the post-test interviews. For instance, Andy, David and Aaron commented that the pauses meant a physical rest to them. Especially Aaron who described himself as a person who wanted to write down as many things as possible during the lecture, could rest his arm which had been busy with note-taking.

4.2.6 Theme 6: Help with Checking of the Accuracy of Understanding

Just as the previous theme, "Allow time to rest," "Help with checking of the accuracy of understanding" is related solely to video stimuli. Interview data

suggested that the video-stimulated lectures allowed test-takers to check the accuracy of their understanding as they were listening. Tina, Jessica, and Aaron were the participants who recognized this function of the video stimuli. Visual supports such as board works, pictures, or realia helped participants to check their understanding because these video stimulus features approved that the participants were following the lecture well. The following shows excerpts of post-test interview that can support this fact.

> *Aaron: In the video-stimulated test, I understand more easily, and I can even check whether I got it well. I liked that characteristic of the video stimulus (stimulated recall interview).*

> *Jessica: When I listened to the audio version lecture, I couldn't be sure about what I heard. But in the video test, I had the visual materials and also their explanations. When I can see the visualized version of what I only imagined, I can be assured that I'm following well (stimulated recall interview).*

With the visual cues in the video stimuli, like pictures or realia, Aaron and Jessica were continuously checking the accuracy of their understanding during the listening.

As presented in this section, the participants described how different features of each type of stimuli helped or disabled their performance. Both types of

stimuli turned out to have features that facilitated or disabled the performance, but video stimuli seemed to have more features that can boost the test-takers' test performance. A detailed discussion about the results would be presented in the following chapter.

Chapter 5. Discussion

This chapter discusses the results presented in the previous section in further detail. Section 5.1 deals with the effects of stimuli types on the test-takers' oral summaries, and Section 5.2 reveals how test-takers used the different types of stimuli to achieve their goal in the tests.

5.1 The Effects of Stimuli Types on Test-takers' Summary

The first research question of this thesis was "How do test-takers' speaking performances in video-stimulated test differ from those in audio-stimulated speaking test?" This study utilized a qualitative method, idea unit analysis, to analyze the spoken summary produced by each participant. The results of the discourse analysis indicated that summaries formulated under the condition of video stimuli were qualitatively better than those made under the condition of audio stimuli at the macro-, micro- and word levels.

Such findings differed from Cubilo's (2011) and Cubilo and Winke's (2013) results which also investigated the effects of stimulus types on test-takers' integrated task performance. Cubilo (2011) and Cubilo and Winke (2013) scored the test-takers' listening-speaking-writing skills performance based on 5 rating subcategories: "Content," "Organization," "Vocabulary," "Language Use," and "Mechanics." Among these subcategories, "Content" dealt with the relevance and details of information in the test-takers' writing which could correspond to the current study's macro- and microlevel analyses. Also, "Vocabulary" subcatetory

that scored whether the words in participants' writings were correct could correspond to the current study's word level analysis. Unlike the research findings of this study, however, Cubilo (2011) and Cubilo and Winke (2013) found out that the types of stimuli could not cause any significant difference to these "Content" or "Vocabulary" related features of the test-takers' writings.

There could be at least two reasons why the current study results were inconsistent with these two previous studies' results. First, Cubilo's (2011) and Cubilo and Winke's (2013) results could have been affected by the presence of reading input. The two previous studies dealt with listening-reading-writing integrated tasks where test-takers could refer to the reading materials when they had some listening skill problems. The information abandonment would have been less frequent in the listening-reading-writing tasks than in a listening-speaking task condition not because of the stimuli format effects, but possibly because of the presence of reading input. The current study, on the other hand, investigated listening-speaking tasks. Although content visual video stimuli in the current study showed some visual cues, they did not require test-takers' reading comprehension. Specifically, the content visuals used in the current study (1) replicated, (2) illustrated, or (3) organized what was orally delivered (Bejar et al., 2000) so that only a minimum level of test-takers' reading skill could be involved.

Second, the scoring methods were different. In the current study, contents of the test-takers' summaries were compared with the source text contents by idea units. Idea unit was a type of semantic unit that could be reliably identified (Crookes, 1990; Foster et al., 2000; Frost et al., 2011). However, in the previous

two studies, the scoring of the “Content” or “Vocabulary” subcategories was conducted by two scorers who shared their impressions of each summary writing. The relatively less systematic and less reliable scoring in these previous studies could have missed some important differences of content and vocabulary features between the audio and video-stimulated test summaries.

Although the findings were different, the current study and the two previous studies commonly supported the use of video stimulus in skills integrated tests for academic purposes. First, the findings of the macrolevel analysis proved that the video stimuli facilitated one of the most crucial processes in summarizing: the process of selecting macrolevel information out of microlevel information. While no error was found in video stimulated tests, several errors could be identified in audio-stimulated tests.

Second, the microlevel analysis showed that video stimuli could prevent the use of test-takers’ own background knowledge mainly due to the presence of pauses. Accordingly, some test developers might argue that similar positive effects would be possible in an audio-only task, by adding longer pauses in an audio stimulus in places where visual supports were provided in a video stimulus. However, long pauses in the audio stimuli might not function in the same way as they did in the video stimuli. The first reason is that in the audio stimuli the test-takers cannot see the cause of the pause and therefore cannot estimate when the speech is going to resume (i.e., after the lecturer finishes writing a phrase), so might not be able to use the allotted time efficiently. In other words, what allowed video stimuli to prevent the use of test-takers’ own background

knowledge, pauses, were a unique feature of the video stimuli.

Finally, the word level analysis showed that the video stimuli could supplement a constraint of the listening-speaking integrated task. One of the listening-speaking integrated task's constraints was that it was appropriate only for students who had achieved or surpassed the threshold level of listening ability: students who did not have adequate listening proficiency could not perform in the listening-speaking task successfully (Cumming, 2013; Yu, 2013). However, Jessica's use of a synonym (i.e., "zigzag" for "jagged") in the word level analysis demonstrated that with the visual cues, the test-takers could overcome their low listening proficiency and move on to speak.

Unlike the previous studies, the current study could demonstrate that there were stimuli format effects on test-takers' listening-speaking integrated test performance. The three-level discourse analysis findings above suggest video stimuli's strengths that audio stimuli cannot provide. For this reason, in case of the academic English test where students are asked to listen to the lecture and speak to summarize, video stimuli should replace audio stimuli.

5.2 The Test-takers' Usage of Stimuli and Their Perception

To answer the second research question, "How do test-takers utilize stimuli in an integrated speaking test?" individual post-test interviews were conducted using the stimulated recall protocol. The analysis revealed that participants had some shared ideas and reflections about utilizing each type of stimuli. The test-takers' common experiences were arranged into six themes, as

shown in Table 4.7, to better illustrate how each feature of the stimuli facilitated or hindered the test-taking.

The analysis of participants' post-test interview data showed that both audio and video stimuli had beneficial as well as negative aspects. Both stimuli facilitated comprehension of the lectures and production planning. As for the negative features, both of the stimuli had distracting or cognitively overburdening features. However, as the two subsections, 4.2.5 and 4.2.6, describe, the participants tended to make more positive remarks to video stimuli than to audio stimuli. For instance, video stimuli allowed some time for them to rest and helped them with the checking of the accuracy of the understanding. On the other hand, audio stimuli made test-takers abandon information when they could not understand the source lectures. With the audio stimuli consisting of audio-only input that immediately vanishes, the abandoning would have been inevitable in case of incomprehension. However, video stimuli were not related to this abandoning of information. In the video stimuli, the missing information possibly had the chance to be supplemented by the visuals such as board works or pictures.

These results that showed the beneficial effects of video stimuli contradicted Coniam's (2001) findings which reported that participants felt distracted by the video stimuli. However, Coniam's findings could not be generalized to the other test-takers' cases because his participants commented on audio and video stimuli without experiencing them both.

The current study's findings were rather in line with Cubilo's (2011) and Cubilo and Winke's (2013) which found that test-takers' tended to prefer video

stimuli to audio stimuli. However, it should be noted that in their study, the participants' responses were collected by a post-test questionnaire which provided only three options to choose from to describe video stimuli: "helpful," "distracting," or "neutral." Because the test-takers' responses were restricted to one of these three options, and the investigation did not cover the perceptions of audio stimuli, the complex test-taking experiences that affected the test-takers' perceptions could not have been reported in their study. In the current study, on the other hand, the participants' experience of different stimulus formats could have been delved into in a more detailed manner with the use of stimulated recall interview.

Chapter 6. Conclusion

This chapter concludes the thesis by summarizing the major findings of the research. The pedagogical and theoretical implications is then followed by the research limitations.

6.1 Major Findings

The current study achieved its two goals: (1) to describe the test-takers' performance differences in the listening-speaking integrated test under the contexts of audio stimuli and video stimuli; and (2) to explore the test-takers' use of audio and video stimuli in the listening-speaking tests. For the first research question, it was concluded that the test-takers perform better at the macro-, micro-, and word levels when there were visual supports. This result underscored the importance of visual cues in the academic EFL assessment contexts where test-takers were required to listen to a lecture and then asked to speak to summarize it.

As for the second research question, the test-takers' usage of each type of stimuli was described with six themes to investigate how features of each stimulus affected the participants' test-taking process. The findings demonstrated that audio and video stimuli had some common strengths and weaknesses. However, test-takers tended to relate video stimuli with more positive aspects, such as "allowing time to rest" or "help with checking of the accuracy of understanding" during the listening compared to the aspects related to audio stimuli. Audio stimuli were related to the negative features such as "cause information abandoning." In other

words, video stimuli had more beneficial effects on the test-takers' performances and gained more positive feedback.

6.2 Implications

The current study investigated (1) the effects of stimuli types on the test-takers' listening-speaking performance, and (2) test-takers' perceptions of each type of stimuli. The findings of this study have several implications for researchers and test developers.

First, test developers should consider ways to enhance test authenticity when designing language tests. Authenticity is an important quality because it can justify the use of the test (Bachman & Palmer, 1996). As an effort to enhance test authenticity, this study suggested the use of integrated tasks and video stimuli in the case of academic English speaking test where test-takers are required to listen to a lecture and summarize. As another way to investigate the validity and the authenticity of this type of tests, researchers proposed the application of eye tracking technology (Batty, 2015; Gass et al., 2013; Suvorov, 2013, 2015). With this method, the researchers and test designers could learn which features of the real-life lectures test-takers are paying attention to, and how test-takers interact with, the real lectures. The information obtained from the suggested study, when applied to the newly developed video stimuli, would make the test reflect the target language use domain better.

In addition, the current study has a theoretical implication for the construct definition of speaking for academic purposes, which is to include the

ability to handle aural input to speak. Cumming (2013) argued for the need to redefine the construct of academic "writing" construct. He explained that the ability to integrate relevant information of the input materials to their own writings was what most of the real-life students are required to do in academic settings. Just as in the academic "writing" construct's case, academic "speaking" construct should also expand. Considering the real-life academic settings where students are required to incorporate source materials to their speaking, the academic "speaking" construct should include the ability to integrate source materials, and should acknowledge the interdependence of speaking performance on listening (and possibly reading) performance.

6.3 Limitations and Suggestions

This study is not free from limitations. First, the size of the participant sample was relatively small mainly because of the qualitative nature of the inquiry. Consequently, the findings could not extend to a wider population with the same degree of certainty that a quantitative study could. Second, the lengths of video stimuli tended to be longer than their audio counterparts. Having such differences in length could be natural considering the fact that video stimuli contain silences that are created during the writings on the board. However, to make sure the two types of stimuli are not affected by unintended differences in lecture delivery, the audio stimuli could be produced by extracting audio files from the video stimuli and then removing the long silences found in these audio files. Third, the reliability of the interview data coding in Section 4.2 could have been higher if the

coding had been done by more than one coder. As the idea unit analysis in Section 4.1, allowing two or more coders to review the data could have increased the likelihood of reliable results. Finally, the current study mainly focused on the effects of stimuli types on the content features of the test-takers' summaries. However, criteria that can describe a good oral summary presentation in an academic context could be more diverse. Brown et al. (2005), for example, scored the test-takers' summaries based on four major categories: "Linguistic resources," "Phonology," "Fluency," and "Content." To further investigate the effects of video stimuli on test-takers' spoken summary of a lecture, a more diverse language features of the test-takers' performance could be observed.

References

Auerbach, C. F., & Silverstein, L. B. (2003). *Qualitative data: an introduction to coding and analysis*. New York: New York University Press.

Bachman, L., & Palmer, A. (1996). *Language testing in practice*. Oxford: Oxford University Press.

Batty, A. O. (2015). A comparison of video- and audio-mediated listening tests with many-facet Rasch modeling and differential distractor functioning. *Language Testing, 32*(1), 3-20.

Brett, P. (1997). A comparative study of the effects of the use of multimedia on listening comprehension. *System, 25*(1), 39-53.

Brown, A., Iwashita, N., & McNamara, T. (2005). *An examination of rater orientation and test-taker performance on English-for-Academic-Purposes speaking tasks*. (TOEFL® Research Report TOEFL-MS-29; ETS Research Report No. RR-05-05). Princeton, NJ: Educational Testing Service.

Brown, H. D., & Abeywickrama, P. (2010). *Language assessment: principles and classroom practicies*. White Plains, NY: Pearson Longman.

Butler, F., Eignor, D., Jones, S., McNamara, T., & Suomi, B. (2000). TOEFL 2000 speaking framework: A working paper. (TOEFL® Monograph Series, Report No. 20). Princeton, NJ: Educational Testing Service.

Choi, I., & So, Y. (2018). A measurement model for listen-speak tasks. In G. Ockey, & E. Wagner (Eds.), *Assessing L2 Listening: Moving towards authenticity*. (pp. 1-24). Philadelphia: John Benjamins Publishing

Company.

Coniam, D. (2001). The use of audio or video comprehension as an assessment instrument in the certification of English language teachers: A case study. *System, 29*(1), 1-14.

Crookes, G. (1990). The utterance, and other basic units for language discourse analysis. *Applied Linguistics, 11*(2), 183-199.

Cubilo, J. (2011). *Video mediated listening passages: Their effects on integrated writing task performance and note-taking practices* (Unpublished master's thesis). Michigan State University, Ann Arbor.

Cubilo, J., & Winke, P. (2013). Redefining the L2 listening construct within an integrated writing task: Considering the impacts of visual-cue interpretation and note-taking. *Language Assessment Quarterly, 10*(4), 371-397.

Cumming, A. (2013). Assessing integrated writing tasks for academic purposes: Promises and perils. *Language Assessment Quarterly, 10*(1), 1-8.

Dornyei, Z. (2007). *Research methods in applied linguistics*. Oxford: Oxford University Press.

Educational Testing Service. (2017). *TOEFL speaking work sample* [PDF document]. Retrieved from https://www.ets.org/s/careers/pdf/toefl-ibt-speaking-2017-ada.pdf

Feak, C. B., & Salehzadeh, J. (2001). Challenges and issues in developing an EAP video listening placement assessment: A view from one program. *English for Specific Purposes, 20*(1), 477-493.

Ferris, D., & Tagg, T. (1996). Academic oral communication needs of EAP learners: What subject-matter instructors actually require, *TESOL Quarterly, 30*(1), 31-58.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics, 21*(3), 354-375.

Frost, F., Elder, C., & Wigglesworth, G. (2011). Investigating the validity of an integrated listening-speaking task: A discourse-based analysis of test-takers' oral performances. *Language Testing, 29*(3), 345-369.

Fulcher, G. (2010). *Practical language testing*. New York, NY: Routledge.

Gass, S., Behney, J., & Plonsky, L. (2013). *Second language acquisition: An introductory course.* Taylor & Francis.

Ginther, A. (2002). Context and content visuals and performance on listening comprehension stimuli. *Language Testing, 19*(2), 133-167.

Gruba, P. (1993). A comparison study of audio and video in language testing. *JALT Journal, 15*, 85-88.

Gruba, P. (1997). The role of video media in listening assessment. *System, 25*(3), 335-345.

Hackers Language Research Institute. (2016). *Hackers TOEFL actual test speaking*. Seoul: Hackers Language Institute Press.

He, A. W., & Young, R. (1998). Language proficiency interviews: A discourse approach. In R. Young, & A. W. He (Eds.), *Talking and testing: Discourse approaches to the assessment of oral proficiency*. (pp. 1-24). Philadelphia: John Benjamins Publishing Company.

Kintsch, E. (1990). Macroprocesses and microprocesses in the development of summarization skill. *Cognition and Instruction, 7*(3), 161-195.

Kintsch, W., & van Dijk, T. A. (1975). Comment on se rappelle et on résume des histoires. *Langages, 40*, 98-116.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review,85*(5), 363-394.

Kirkland, M. R., & Saunders, M. A. P. (1991). Maximizing student performance in summary writing: Managing cognitive load. *TESOL Quarterly, 25*(1), 105-121.

Levy, P. (2014). Introduction. In P. Levy (Ed.), *The Oxford handbook of qualitative research* (pp. 1-13). Oxford: Oxford University Press.

Londe, Z. C. (2009). The effects of video media in English as a second language listening comprehension tests. *Issues in Applied Linguistics, 17*(1), 41-50.

Luoma, S. (2004). *Assessing speaking*. Cambridge: Cambridge University Press.

Magnusson, E., & Marecek, J. (2015). *Doing interview-based qualitative research: A learner's guide*. Cambridge: Cambridge University Press.

Ockey, G., Koyama, D., Setoguchi, E., & Sun, A. (2015). The extent to which TOEFL iBT speaking scores are associated with performance on oral language tasks and oral ability components for Japanese university students. *Language Testing, 32*(1). 39-62.

Read, J. (1990). Providing relevant content in an EAP writing test. *English for Specific Purposes, 9*, 109-121.

Salehzadeh, J. (2006). *Academic listening strategies: A guide to understanding lectures*. Ann Arbor: University of Michigan Press.

Shin, D. (1998). Using videotaped lectures for testing academic listening proficiency. *International Journal of Listening, 12*(1), 57-80.

Suvorov, R. (2008). *Context visuals in L2 listening test: The effectiveness of photographs and video vs. audio-only format* (Unpublished master's thesis). Iowa State University, Ames.

Suvorov, R. (2013). *Interacting with visuals in L2 listening tests: An eye-tracking study* (Unpublished doctoral dissertation). Iowa State University, Ames.

Suvorov, R. (2015). The use of eye tracking in research on video-based second language (L2) listening assessment: A comparison of context videos and content videos. *Language Testing, 32*(4), 463-483.

Turner, A. A. (1987). *The propositional analysis system* (Tech. Rep. No. 87-2). Boulder: University of Colorado, Institute of Cognitive Science.

van Dijk, T. A. (1980). *Macrostructures: An interdisciplinary study of global structures in discourse, interaction, and cognition*. Hillsdale, US: Lawrence Erlbaum Associates.

Wagner, E. (2010). The effects of the use of video texts on ESL listening test-taker performance. *Language Testing, 27*(4), 493-513.

Wagner, E. (2013). An investigation of how the channel of input and access to test questions affect L2 listening test performance. *Language Assessment Quarterly, 10(*2), 178-195.

Yu, G. (2013). From integrative to integrated language assessment: Are we there yet? *Language Assessment Quarterly, 10*(1), 110-114.

Appendices

Appendix 1: Translated Version of Preliminary Screening Survey

1. English Proficiency

(1) What is your TEPS score? (ex: "851")

_____

2. Background Knowledge and Fields of Interest

(1) What is your majoring area? (Please specify. ex: "OO University Department of Education, English Language Education Major”")

_____

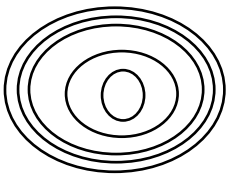
(2) Check the degree of your knowledge for each of the following fields of study. (1 = *know nothing* ↔ 5 = *know very well*)

	1	2	3	4	5
Law					
Management					
Psychology					
Economics					
Earth Science					
Biology					
Environmental Issues					
Art					
Engineering					
Criminal Justice					

Appendix 2: Video Stimuli Scripts

1. Earth Science: Ancient Weather Patterns (Hackers Language Research Institute, 2016)

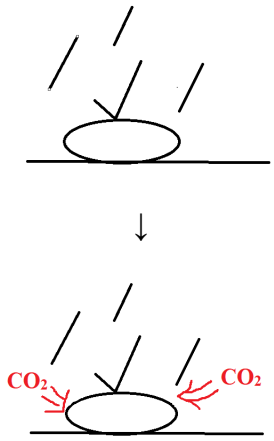
Script	Board works/Gestures
(Before the recording starts)	"Ancient Weather Patterns"
There are a number of methods used by scientists to study **ancient weather patterns**[a]. Today, I want to focus on how plant fossils provide valuable information about past climates. (pause to write on the board)	"Plant Fossils"
Now, the fossilized remains of leafy plants (pause to write on the board) are especially useful when trying to understand prehistoric climate conditions.	"Leafy plants"
This is because the **shape of a leaf** is determined by the, um, average annual temperature of the environment in which the plant lives. If the majority of plants from a specific time period had leaves with **smooth edges** (realia), global temperatures were likely **higher** than average. Um, plants with this type of leaf retain moisture better, which is an obvious advantage in warm climates.	(Show a leaf with smooth edges)
In contrast, plants that produce leaves with **jagged edges** (realia) are more common in **cooler climates** because	(Show a leaf with jagged edges)

they can perform photosynthesis with greater efficiency. This is important if the plant has to cope with a short growing season each year.	
Researchers also study pieces of ancient **tree trunks** to learn about past climates. (pause to write on the board) Commonly referred to as **petrified wood**, these fossils offer a wealth of data.	"Tree trunks (petrified wood)"
Most trees have growth rings (pause to draw on the board). **Widely** spaced rings (pointing to the drawing) mean that the tree grew **quickly** each year, while **narrow** ones (pointing to the drawing) indicate periods of slow growth. Scientists can use this information to determine how temperature and, um, precipitation patterns varied over the course of the tree's life. Um, for example, a series of very narrow rings (pointing to the drawing) may indicate that the area in which the tree lived experienced an extended period of draught.	(Draw an image of a tree trunk) 

[a] Bold letters indicate words to emphasize orally.

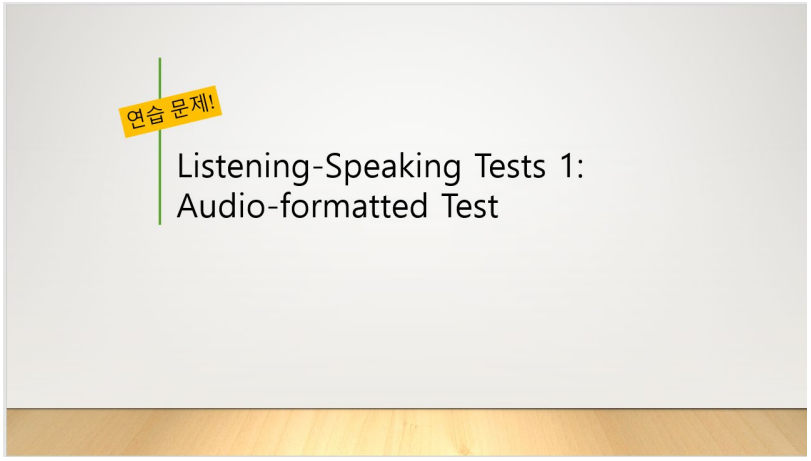
2. Environmental Issues: Reducing the Amount of Atmospheric CO_2 (Hackers Language Research Institute, 2016)

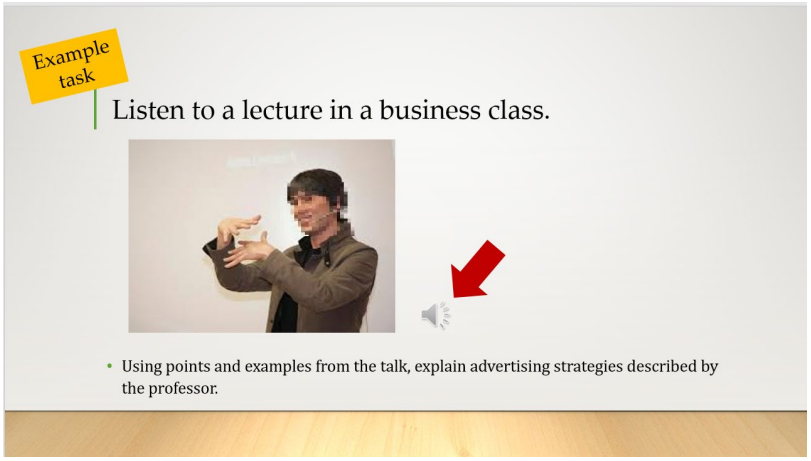
Script	Board works/gestures
(Before the recording starts)	"Reducing the amount of atmospheric CO2"
So the level of CO2 in the atmosphere has been rising steadily for decades. And because of the alarming effect this is having on the Earth's climate, it's imperative that we find a way to **reduce**[a] the amount of atmospheric CO2. Fortunately, there are a couple of promising methods that can be used to accomplish this goal.	
One possibility is to create **artificial trees** that absorb CO2 (pause to write on the board). Here's how it works. A special type of **plastic** is made with a chemical compound that, uh, reacts with CO2. And the result is that the plastic is basically able to **pull** CO2 out of the air. Now you might ask, why not just plant **natural trees** instead? Well, one advantage of using the plastic to build artificial trees is that they absorb CO2, um, a **thousand times faster** than real ones. Another plus is that the carbon storage is **permanent**. As you might know, when real trees die and rot,	"Artificial trees" (Use gestures to explain)

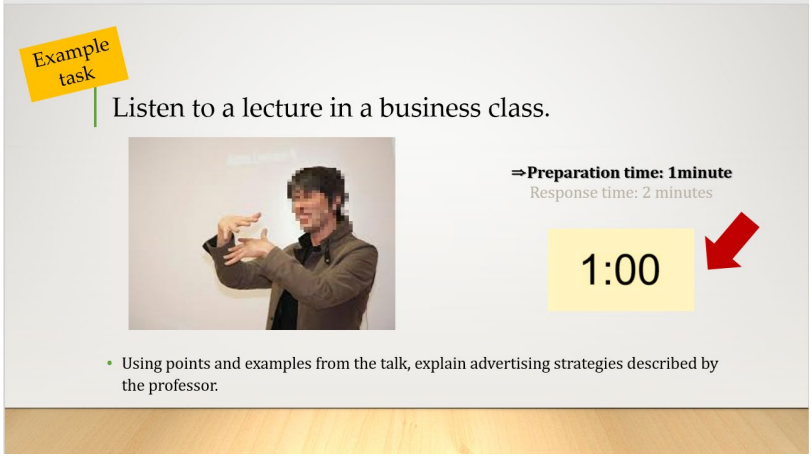
carbon is **released** back into the atmosphere. But, obviously, this isn't an issue with plastic trees.	
Another method is called **enhanced weathering** (writing on the board). (drawing on the board) To give you some background, uh, when rain **strikes** rocks on the ground, a chemical reaction occurs whereby the rocks **absorb** CO_2 from the air (add drawing on the board). This process is called **weathering** and when it occurs naturally, it takes a very long time to absorb large amounts of CO_2.	"Enhanced weathering" 
Fortunately, scientists have devised a way to enhance the process so that CO2 levels can be lowered much more quickly. First they dig up rocks, such as olivine (realia), that react very strongly with CO2. They then grind (gesture) these rocks into a fine powder in order to maximize surface area .. this speeds up the reaction. The powder is spread out over the land, and when rain hits it, a large amount of CO2 is removed from the atmosphere. Best of all, the CO2 can be stored in the dust permanently.	(Show a rock)

[a] Bold letters indicate words to emphasize orally.

Appendix 3: An Example of Task Prompt (for Group A)







Example task

Listen to a lecture in a business class.

Preparation time: 1minute
⇒Response time: 2 minutes

2:00

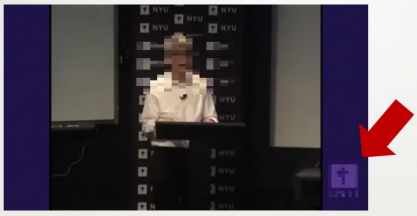
- Using points and examples from the talk, explain advertising strategies described by the professor.

연습 문제!

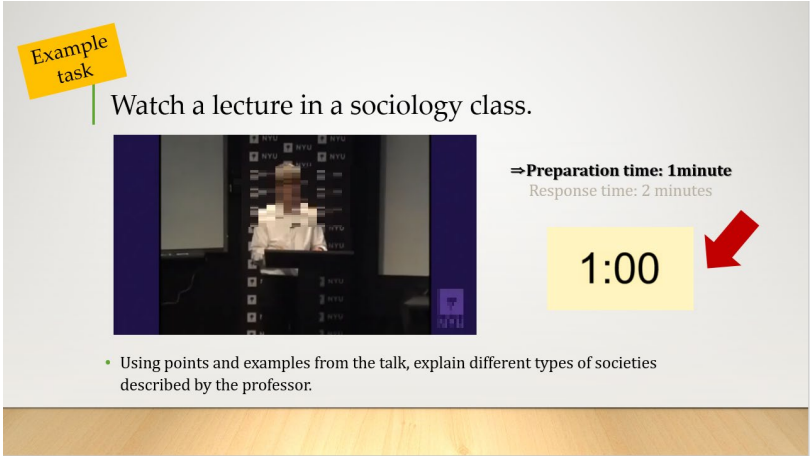
Listening-Speaking Tests 2:
Video-formatted Test

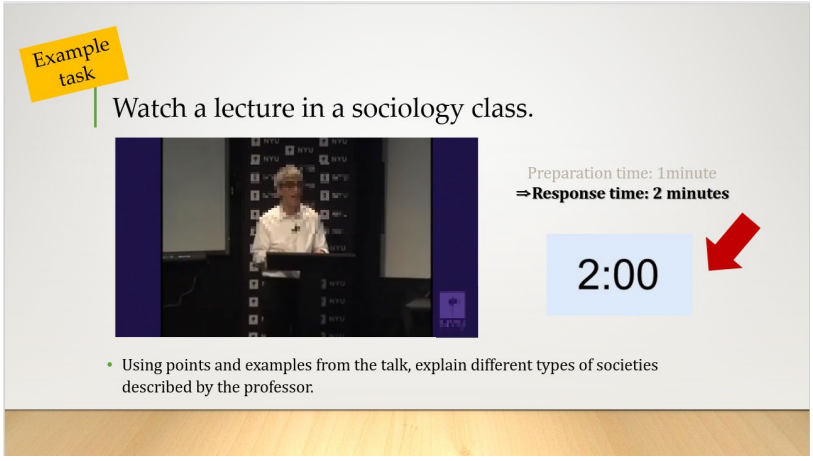
Example task

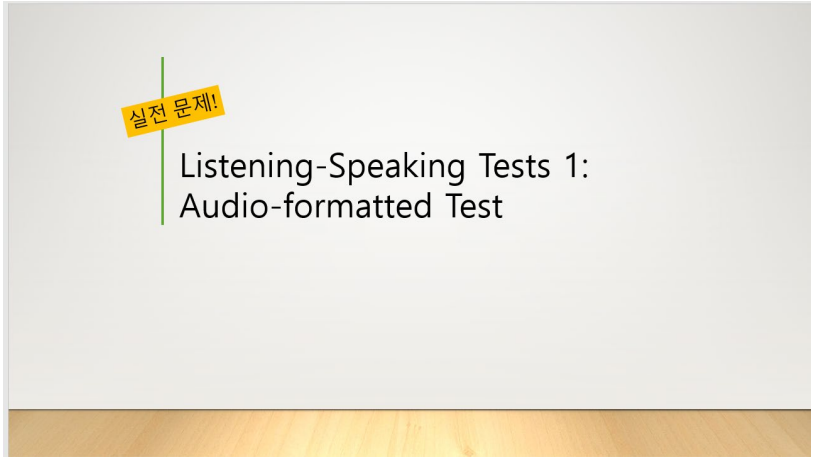
Watch a lecture in a sociology class.



- Using points and examples from the talk, explain different types of societies described by the professor.


Example task
Watch a lecture in a sociology class.
⇒Preparation time: 1minute
Response time: 2 minutes
1:00
Using points and examples from the talk, explain different types of societies described by the professor.


Example task
Watch a lecture in a sociology class.
Preparation time: 1minute
⇒Response time: 2 minutes
2:00
Using points and examples from the talk, explain different types of societies described by the professor.


실전 문제!
Listening-Speaking Tests 1:
Audio-formatted Test

Listen to a lecture in an earth science class.

- Using points and examples from the lecture, explain methods used to study ancient weather patterns.

Listen to a lecture in an earth science class.

⇒Preparation time: 1minute
Response time: 2 minutes

1:00

- Using points and examples from the lecture, explain methods used to study ancient weather patterns.

Listen to a lecture in an earth science class.

Preparation time: 1minute
⇒Response time: 2 minutes

2:00

- Using points and examples from the lecture, explain methods used to study ancient weather patterns.

실전 문제!

Listening-Speaking Tests 2: Video-formatted Test

Watch a lecture in an environmental issues class.

- Using points and examples from the lecture, explain methods used to reduce the amount of CO_2 in the atmosphere.

Watch a lecture in an environmental issues class.

⇒Preparation time: 1minute
Response time: 2 minutes

1:00

- Using points and examples from the lecture, explain methods used to reduce the amount of CO_2 in the atmosphere.

Watch a lecture in an environmental issues class.

Preparation time: 1minute
⇒Response time: 2 minutes

2:00

- Using points and examples from the lecture, explain methods used to reduce the amount of CO_2 in the atmosphere.

국 문 초 록

듣기 자료의 유형이 한국 대학생의 통합형 영어 말하기 시험의 수행에 미치는 효과

노윤희
외국어교육과 영어전공
서울대학교 대학원

본 연구는 듣고 말하는 통합시험에서 두 가지 듣기 자료의 유형이 피험자들의 수행에 어떠한 영향을 미치는지와, 피험자들이 각 듣기 자료를 어떤 방식으로 활용하는지를 알아보고자 하였다. 사전 설문에서 영어 능숙도와 배경지식이 가장 비슷한 8명의 대학(원)생 최종 참여자를 선발하였으며, 이들은 오디오 자료가 제공되는 듣고 말하는 시험과 영상 자료가 제공되는 듣고 말하는 시험을 각각 한 개씩 수행하였다. 피험자들이 접하는 듣기 자료 유형(오디오인지 영상인지)의 순서와 내용(강의 내용) 순서는 라틴 스퀘어 방식에 따라 통제되었다. 과제 수행 후, 각 피험자는 자신의 수험 장면이 담긴 영상을 보며 연구자와 수험 과정에 대한 자극회상기법 인터뷰를 하였다. 피험자들의 구두 요약문을 질적으로 분석한 결과, 시각 정보가 제공되는 조건에서의 요약문이 그렇지 않을 때의 요약문보다 거시, 미시, 그리고 어휘 단계에서 더 나은 것으로 나타났다. 또한, 피험자들의 수험 과정을 탐구한 결과, 과제 수행에 도움이 되는 요소와 방해가 되는 요소는 오디오와 영상 자료 모두에서 찾아볼 수 있었지만, 영상 자료의 긍정적인 기능이 더 많이 묘사된

것을 알 수 있었다. 이러한 결과는 오디오 자료보다는 영상 자료가 학술 언어를 측정하는 듣고 말하기 시험의 듣기 자료로서 쓰여야 한다는 주장을 뒷받침한다. 본 연구의 결말부에는 시험의 진정성 신장과 학술적인 말하기 능력의 정의에 대한 제언을 제시한다.

주요어: 영상 자료, 오디오 자료, 듣기-말하기 평가, 요약하기, 담화분석, 자극회상기법

학 번: 2016-21787