영어어문교육 14권 1호 2008년 봄

The Need for an Alternative Approach to Oral Testing

Mi-jin Joo
(University of London)

Joo, Mi-jin. (2008). The need for an alternative approach to oral testing. *English Language & Literature Teaching, 14*(1), 1-20.

The purpose of this study was to introduce a new type of oral test, a Computerized Oral Test (COT) and to investigate current oral testing methods and teachers' and students' perceptions of the current test methods and the COT in a Korean university classroom. The results of the study showed that only a few teachers were using a Face-To-Face Interview (FTFI) format, mainly due to its impracticality, even though both teachers and students perceived that the FTFI was the most appropriate test format for assessing and helping to improve speaking skills. Both groups felt that they needed a more effective test method for accurate assessment and improvement of speaking skills. The perceptions of the COT were generally favorable although more teachers showed preference for the FTFI over the COT. I in turn proposed the COT format as an alternative oral testing method to the FTFI in the Korean university classroom.

[computerized oral test/face-to-face interview]

I. INTRODUCTION

A Face-To-Face Interview (FTFI) is the most traditional and established of the oral test methods and is what comes to most people's minds when an oral exam is mentioned. The FTFI involves the learner in communicative interaction, which takes place under pressure of time and therefore necessitates rapid language processing and production and which is essential for the improvement of learners' language proficiency (Nam, 2007). It enjoys a high degree of face validity among both teachers and students due to its appearance of an authentic communicative event. The interview format has been considered to be an authentic test form in terms of validity.

However, Clark (1979) acknowledges that the FTFI format does not fully reflect real-life conversational situations, arguing that "in the interview situation, the test taker is certainly aware that he or she is talking to a language assessor and not to a waiter, taxi driver, or

personal friend" (p. 38). Besides, it is difficult to have true interaction between the interviewer and the interviewee and elicit the language in real life conversation because the interviewer normally asks questions and controls the conversation, and the interviewee plays the subservient role answering the questions and responding to prompts initiated by the interviewer (Stansfield, 1991). van Lier (1989) supports this view noting that the FTFI format aims to elicit language rather than to make successful conversation. van Lier (1989) distinguishes the essential features of the FTFI format from the conversation: "the interviewer has a plan and conducts and controls the interview largely according to that plan," while a conversation is characterized by "face-to-face interaction, unplannedness (locally assembled), unpredictability of sequence and outcome, potentially equal distribution of rights and duties in talk, and manifestation of features of reactive and mutual contingency" (pp. 495-496).

In spite of the lack of authenticity of the FTFI format, the FTFI has been considered to be a more authentic test form than any other types of oral tests, and the FTFI format has been preferred and recommended to measure students' speaking skills (Clark, 1979; Underhill, 1987; Van Lier, 1989). Clark (1979) asserts that "the face-to-face interview appears to possess the greatest degree of validity as a measure of global speaking proficiency" and is clearly better than any other forms of tests (p. 38).

There are, however, some constraints and problems for implementing the FTFI, especially if it is undertaken on a large scale. The first constraint of the FTFI format is its impracticality. Practical difficulties can arise in the administration of oral tests such as cost, availability of suitable rooms to carry out the test, and provision of a sufficient number of trained examiners. For reasons of security it is necessary to provide more than one set of test materials and time and resources are needed for standardization of both materials and raters.

Secondly, it would seem fairly obvious that in a situation which uses only one person as both the interlocutor and the rater, the scores will be far less reliable than if two people are involved. Splitting attention between controlling the interview and marking it was found by Chambers and Richards (1993) to be extremely difficult, especially if detailed marking scales were being used. However, the teacher generally functions simultaneously as the interlocutor and the rater because involving two people for oral assessment is not feasible in the normal classroom situation.

The clarity and detail of the marking scales with which raters are provided are an important element in establishing the reliability of test scores. However, Alderson (1991) points out that raters may produce different ratings because they are unreliable, because they misinterpret the criteria or because they use their own internalized interpretations of the words and descriptions.

These sources of difficulty and error can be minimized to a large extent by

standardization training of interlocutors and raters. The training helps them not only to assess their students more reliably by understanding and becoming familiar with the rating criteria and procedure, but also to learn how to interact with test takers consistently. The impact of interlocutors on test taker performance has been recognized by some researchers (e.g., Fulcher, 2003; McNamara, 2000; O'Loughlin, 2001). Skillful interlocutors will avoid giving advantages or disadvantages for some over others because of their different personalities and backgrounds. However, the training of the interlocutors and the raters will also be very costly.

Finally, if interlocutors and raters become tired or stressed due to the large number of students they have to test, or to the fact that they are running behind schedule, this may also affect the reliability of their marking and unfavorably affect students who have their interviews later in the day. In some of the efforts to reduce the constraints of implementing the FTFI, semi-direct oral tests which are designed to approximate as closely as possible the linguistic content and manner of operation of the FTFI have been developed. One of them is a Computerized Oral Test (COT) which I would like to introduce in this study. I believe that the COT, which allow teachers to administer an oral test in a shorter time and with less effort, may be useful as an alternative test method to the FTFI format.

Students' attitudes and belief about second languages influence their learning behavior and motivational patterns and in turn affect their second language performance (Kim, 2006). Likewise, students' attitudes to and perceptions of their tests may affect not only their learning behavior and motivation to learn but also test performance. For example, if test takers perceive that the test tasks they have been asked to do are not reasonable, or feel that they are engaged in pointless exercises, this may affect the way they perform and hence the scores they obtain. As a result, it may attenuate the test's validity.

Therefore, in this study I attempted first to introduce the COT and then, in order to examine the current classroom situation at Korean universities and teachers' and students' attitudes with regards to oral assessment, I investigated the current oral testing methods in the Korean university classroom, along with teachers' and students' perceptions of the oral testing methods, as well as their feelings about the COT format. The following questions were investigated: 1) What oral testing methods do teachers use in their class?, 2) How do teachers and students feel and think about their tests?, and 3) How do teachers and students perceive and think about the COT?

II. LITERATURE REVIEW

For the study, the Oral Testing Software (OTS) (Brigham Young University, http://creativeworks.byu.edu/HRC), which is designed to test speaking proficiency by

Brigham Young University, was used. The test is self-administered but not marked by computer as are other general computer-mediated oral tests. Before introducing the COT more fully, I briefly explored some other types of semi-direct oral tests which had been or would be used: the Computerized Oral Proficiency Interview (COPI), the computer-assisted Basic English Skills Test (BEST), and the Test of English as a Foreign Language Academic Speaking Test (TAST).

1. Other Semi-direct Oral Tests

A number of semi-direct tests of foreign language speaking ability have been developed to reduce the difficulties of implementing a FTFI. One of the most widely known tests is the Computerized Oral Proficiency Interview (COPI). To improve the Simulated Oral Proficiency Interview (SOPI), which is a tape-based speaking test, CAL carried out a 2-year study on the development of a computer-assisted oral test, the Computerized Oral Proficiency Interview (COPI). It is a multi-media computer-administered oral proficiency test and an adaptation of the SOPI. The COPI depends on taped and written directions to elicit language from test takers. The COPI, however, gives test takers more control of various aspects of the testing situation and increases raters' efficiency in scoring the test. It allows the test taker a choice in the following aspects (Center for Applied Linguistics, http://www.cal.org/project/copi.html).

'Thinking and response time'- the COPI provides preparation and response time, but the test takers still have the choice to use up all the time allotted or to click on a button when they are finished preparing or are ready to respond. 'Speaking functions and topics'- test takers are provided choices throughout the program, and an algorithm makes certain that test takers get each speaking function (e.g., narrating in the past) and topic (e.g., food) only once. 'Level of task difficulty'- test takers' self-assessment scores and the choices of levels during the practice task decide their level of difficulty for the first task. However, after the first task, the program alternates from giving test takers choices or not. During the times that test takers are not given a choice, the algorithm pushes test takers to higher level tasks to ensure that they are provided with an opportunity to reach their ceilings. 'Language of the directions'- while lower-level Spanish speakers have English directions, higher level Spanish speakers are provided a choice to get the directions in English or Spanish. However, a later version of the COPI plans to include both English and Spanish directions for lower-level speakers.

Malone (2000) and Kenyon & Malabonga (2001) argue that the COPI can be seen as an advanced oral testing instrument of the SOPI, but the COPI has not been operationalized to testing speaking proficiency because it was developed principally as a research study.

Two of other important computer-mediated oral tests which may be introduced here are:

the computer-assisted Basic English Skills Test (BEST) (Center for Applied Linguistics, http://www.cal.org/BEST/compbest.html) and the Test of English as a Foreign Language Academic Speaking Test (TAST) (Educational Testing Service, http://www.ets.org).

CAL has been awarded a contract by the U.S. Department of Education to produce an updated, computer-assisted oral assessment based on the BEST oral interview. The BEST was developed during the early 1980s by CAL as a means of assessing the English language proficiency of immigrants and refugees who were entering the United States at that time. The BEST contains two parts, an oral interview section and a literacy section. The oral interview section requires a FTFI administering individually, which takes approximately 15 minutes per test taker. The section consists of simulated real-life listening comprehension and speaking tasks, such as telling time, asking for directions, and following directions. The literacy skills section is administered in one hour. The section includes two parts: reading tasks (e.g., dates on a calendar or labels on food and clothing) and writing tasks (e.g., addressing an envelope or writing a rent check). At present, hundreds of adult ESL programs across the United States use the BEST for assessment purposes.

The new computer-assisted BEST is designed to allow test takers to show what they can do in English in a shorter time frame. It is still administered as a FTFI, but the test administrator enters scores into a computer. The computer then selects the next test question to be administered, and continues to adapt the difficulty level of the questions according to the scores entered for each question. This computer-assisted oral test has been available on CD-Rom since September 30, 2002, along with test administration instructions and test administrator training materials.

The TAST was developed to measure test takers' ability to communicate orally in an academic environment. In September 2005, when the new TOEFL (TOEFL, Internet-Based Test) was scheduled to launch, the TAST was included as the speaking section. A revision of the TAST is currently available as a stand-alone test used by individuals and institutions to practice for the new TOEFL. The TAST is delivered via the Internet at secure, official test centers, and test takers cannot take the official test at home on the Internet. Test takers have to speak into a microphone, and responses are recorded and scored by trained raters.

2. COT

Like the COPI and the TAST, the COT is a self-administered computerized oral test. It employs computer technology to allow test takers and computer to work together to produce ratable speech samples. The Oral Testing Software-Enhanced (OTS-E) consists of five separate modules (Larson & Smith, 2001):

1) Question Creation Module – In this module, teachers can create test items. It allows teachers to determine the type of item response elicitation prompt (e.g., text, pre-digitised sound, graphic/picture, video, or combination) required for a given item; decide how much time is allowed for preparing (or thinking) to answer each item; and determine how much time is allowed for test takers to respond to each item. 2) Task List Editor – A program allows teachers to decide which tasks will be included on the test and in what order and how many questions will be given related to each task. 3) Test Installation Module – In this module, teachers can decide in which directory the test will be located (either on an individual computer or on a network server). 4) Test administration module – This module is the 'administration engine' that administers the test and records test takers' responses. It allows teachers to decide in which directory the test takers' responses will be saved (either on an individual computer or a network server); involves a sound check to ensure that the recording of the computer is working right; carries out test takers' IDs check; gives test items to test takers; and ends the test, filling the test takers' responses in the designated directory for subsequent assessment. 5) Test assessment module – In this module, a list of test takers and individual responses are shown. The module allows teachers to take assessment notes in a notepad while listening and assessing each test taker's oral responses; have immediate access to any response of any test takers who took the test (no fast forwarding or rewinding trying to locate the beginning of a given response); be able to take notes on-line about individual test takers' responses and assign a grade to that individual; and print or save the assessment notes regarding test takers' performance for a later time.

In the COT, all contents and methods have to be authored and selected by teachers. While taking the test, test takers hear test directions with the accompanying text and pictures in English and/or in a target language. All responses are to be spoken into the microphone and the responses are recorded either on the hard drive or a removable zip disk on the computer. Some of the advantages of the COT are stated by Larson & Smith (2001) and on its web site (http://www.calico.org):

First, many people can be assessed at the same time by utilizing computers and digital equipment like other semi-direct oral tests (e.g., the COPI and the TAST). This would be a major advantage, allowing teachers with large classrooms to implement an oral test. Secondly, the COT can increase teachers' efficiency in scoring the test. The scoring program allows raters to hear the test takers' responses for each task and listen to test takers in any order. They would then be able to listen to test takers' responses in whole or in part, as many times as they wanted to for a particular task, and likewise go back to previously rated tasks. The program also allows raters to write notes to test takers so that, aside from providing a global rating, raters are able to give overall comments and task-specific feedback to each test taker. These activities would allow teachers to give more reliable

scores to test takers. Thirdly, a variety of response elicitation prompts are possible (e.g., text, sound, graphics, video, or a combination). Lastly, it enables tests and results to be easily and economically stored in and retrieved from the hard drive or a removable zip disk on the computer.

Because of the advantages above, the COT may be an improvement over the SOPI format in terms of its efficiency in scoring test taker performance, storing and retrieving tests and results, and a variety of response elicitation prompts.

Together with the significant advantages which may be associated with the COT, there are some limitations stated in Larson & Smith (2001) and on its web site. The limitations stated here are also the limitations of other computer-mediated oral tests.

First, response elicitation information is restricted to the area of the computer screen. Secondly, for some test takers, taking a test on the computer may cause additional anxiety as compared with a live interview. Lastly, multimedia-capable computers with headsets or speakers and microphones are required to administer the tests, unlike the case of the live interview, in which such a computer facility is not needed.

In spite of its numerous functions stated above, the COT does not use very high level technology compared with more modern testing programmes such as the COPI, because it does not allow test takers to choose either speaking topics or the difficulty level of the tasks presented to them. However, the COT shares features with other computer-mediated oral tests despite its simple technology, in that it is possible to self-administer; for many people to be assessed at the same time; to use a variety of response elicitation prompts (e.g., text, sound, pictures or video); and that the responses are recorded either on the hard drive or a removable zip disk on the computer. Therefore, I believe the present study will be an aid for the researchers who want further study on other computer-mediated oral tests.

III. RESEARCH DESIGN

In order to investigate teachers' and students' perceptions of oral testing methods, questionnaires were conducted. For practical purposes, it was decided to pilot the items of the questionnaire, choosing five students and two teachers to serve as respondents. The results of the pilot study showed that there was an immediate need for some changes. Following the respondents' suggestions, some of the items were modified. The design of the questionnaire and the data collection process are stated below.

1. Questionnaire

The aim of the questionnaire was to measure teachers' and students' perceptions of oral

test methods in their classes (see Appendices 1 and 2). The questionnaire contained four sections: 1) information about class sizes and test methods, 2) perceptions of current test methods, 3) perceptions of the COT, and 4) demographic information.

In Section 2 and 3, 5-point Likert scales were used. The 5-point Likert scales provide a statement that reflects a particular attitude or opinion. They have the advantage of convenience that makes the questionnaire suitable for the study. This convenience lies in obtaining, in a simple manner, the differences among respondents. To obtain a picture of the characteristics of teachers' and students' perceptions and to compare the differences between the two groups this method was appropriate.

2. Procedure

The setting for the study was at three Korean universities. Research was conducted from April to June 2005. Participants were randomly selected from the universities. 24 teachers and 171 students, all who had been involved in English conversation classes, completed the questionnaire.

After the pilot study some of the items on the questionnaire regarding perceptions of test methods were revised, and the final forms were used for the data collection. Before asking the participants to fill out the questionnaire, I showed the COT program and gave them a detailed explanation about its administration, operation, and the types of tasks that could be constructed using the COT program.

Most questionnaires were directly distributed to the teachers. To collect data from the students, some of the teachers were asked to distribute the questionnaires to their students. The time taken for filling out the form was approximately ten minutes. In order to encourage the teachers and the students to express their opinions as honestly as possible, an introductory letter conveying the aim of the questionnaires, and assuring the confidentiality of the information obtained through the questionnaires, was appended to each questionnaire sheet.

3. Data Analysis

For data analyses, SPSS version 13.0 was used to conduct descriptive statistics and chi-square tests. Significant p values from the tests were further investigated by calculating effect sizes using Cramer's V, which is considered as the best measure of association of the chi-square test. Cramer's V is described as ranging from .1 (small effect) through .3 (medium effect) to .5 (large effect) (AcaStat Software,http://www.acastat.com/Statbook/chisqassoc.htm).

IV. RESULTS AND DISCUSSION

1. Current oral testing methods

As shown in Table 1, most teachers (70.9%, n= 17) had over 31 students in a class and 41.7% (n=10) of them used 'a group oral interview (and a group project or a pencil and paper test)' as their test method and only 8.3% (n=2) of them employed the FTFI, even though 58.3% (n=14) of the teachers thought the FTFI was the most appropriate test format for assessing accurately and encouraging improvement of their speaking skills. Almost half of the students (47.95%, n=82) also perceived that the FTFI would encourage them to improve their speaking skills most (see Table 2). However, a fairly high percentage of the students (42.11%, n=72) felt that they improved their speaking, or other skills including speaking, primarily through preparing for their tests. As for the rest, it was listening (32.1%, n=55), reading (11.69%, n=20), grammar (5.4%, n=9), vocabulary (4.6%, n=8), and writing (4.1%, n=7).

From the result of the cross-tabulation analysis of class sizes and testing methods ($\chi 2$=65.17, p=.00, V=.82), it was found that 84.6% (n=11) of the teachers who had between 31 and 40 students gave 'a pencil and paper test' or 'a group oral interview and a project (or a pencil and paper test)' to the students as their mid or final test. 100% (n=4) of the teachers with a larger class size (over 41) used 'a pencil and paper test (and a group project or an oral presentation)' as their exam. Most students who took a pencil and paper test or a combined test, excluding the FTFI and the group oral interview (e.g., a pencil and paper test and an oral presentation or a group project), answered that listening, vocabulary, and/or reading skills were improved most through preparing for the test.

TABLE 1

Class Sizes and Oral Testing Methods

Class sizes	%	N	Test methods	%	N
			FTFIs	8.3	2
10-20	8.3	2	Group oral interviews	4.2	1
			Pencil and paper tests	8.3	2
21-30	20.8	5	Group projects	8.3	2
			Role plays	8.3	2
31-40	54.2	13	Group oral interviews and group projects	25	6
			Group oral interviews and pencil and paper	16.7	4
41-50	4.2	1	tests		
			Pencil and paper tests and group projects	4.2	1
More than			Pencil and paper tests and oral presentations	12.5	3
51	12.5	3	Pencil and paper tests and role plays	4.2	1
Total	100	24	Total	100	24

TABLE 2

The Best Oral Testing Method Perceived by Teachers and Students

	Teachers		Students	
Test methods	%	N	%	N
FTFIs	58.3	14	47.95	82
Group oral interviews	25	6	25.73	44
Pencil and papers	0	0	2.9	5
Group projects	8.3	2	14.0	24
Role plays	8.3	2	5.8	10
Others	0	0	1.2	2
Missing	0	0	2.3	4
Total	100	24	100	171

In the open-ended question of the questionnaire, 'If you didn't use the test method you thought the most appropriate, what was the reason?', some of the teachers similarly answered that they did not use the test they thought the most appropriate because of its impracticality.

> "I think the FTFI is a good test format for the students, but I use a pencil and paper test because it is easy and convenient to mark and grade lots of students."

In brief, there were only a few teachers (n=2) using a FTFI format even though they perceived it was the most appropriate test method to assess and encourage their students' speaking skills to improve. They seemed to utilize 'a group oral interview' as an alternative test method to the FTFI. However, the teachers having a larger class (over 41) used only 'a pencil and paper test (and a group project or an oral presentation)', which saved time and effort.

From this evidence, the impracticality of the FTFI format in the Korean university classroom is demonstrated again in this study. As mentioned earlier, with FTFI, the individual administration and the marking of the individuals' performance will take a great deal of time. In addition, the duration of the test may affect its successful operation in other ways (e.g., the effect of fatigue on both teachers and students). Teachers are always faced with limited resources for developing, administering, and marking tests. A valid and reliable test will be of little use if it does not prove to be a practical one. Weir (1990) also points out the impracticality of the FTFI format:

> "Tests such as a FTFI are difficult and time consuming to construct, require more resources to administer, demand careful training and standardization of examiners and are more complex to mark and report results on. The cost of such tests in large scale testing operations may severely restrict their use" (p. 35).

As an alternative test method to the FTFI format, many teachers (45.9%, n=11) seem to be employing the group oral interview test format wherein students are given an information gap task, which they have to complete by obtaining missing information from each other. The interaction is set up by the teacher who then withdraws from the situation, allowing the negotiation of meaning to take place only between the students. Skehan (1991) sees the group oral interview as a promising technique for the future, stating that using groups is not dissimilar from a real-life language use setting and proposing that it can avoid the asymmetric power relations involved in tester-tested interactions.

There are some potential problems with the technique, such as the issue of how to involve a reticent student in the interaction if the rest of the group fails to do so after the task is set up. Especially in a setting where there is a fairly marked difference in proficiency, it is conceivable that some learners may produce such a minimal sample of language that it is insufficient for assessment, or scores very low, but this does not necessarily mean that their true underlying language ability has been accurately assessed, or that they are incapable of performing communicatively. It would be important to be aware of how personality traits and the group dynamic might affect performance in this test format. The COT format which I would like to consider in this study may also decrease the potential problems of the group oral interview, while taking some of the advantages of the FTFI format.

2. Perceptions of Current Oral Tests

In Table 3, it seems that the responses between the teachers and the students were not very different. 41.6% (n=10) of the teachers and 42.4% (n=73) of the students (strongly) agreed that their test was appropriate for assessing and improving speaking skills while 41.7% (n=10) of the teachers and 23.2% (n=40) of the students (strongly) disagreed with the statement. From a cross-tabulation analysis, it was found that 80% (n=8) of the teachers who (strongly) disagreed with the statement were using 'a pencil and paper test (and a presentation, a group project, or a role play)' ($\chi 2$=49.29, p=.03, V=.72), but the students' perceptions was not significantly related with the test methods ($\chi 2$=37.90, p=.10, V=.24).

TABLE 3

I thought the test was the most appropriate for assessing and improving speaking skills

Participants	SA %(n)	A %(n)	N %(n)	D %(n)	SD %(n)	Missing
Teachers	8.3 (2)	33.3 (8)	16.7 (4)	25.0 (6)	16.7 (4)	0(0)
Students	1.7 (3)	40.7 (70)	32.6 (56)	17.4 (30)	5.8 (10)	1.7(3)

SA=Strongly Agree, A=Agree, N=Neutral, D=Disagree, SD=Strongly Disagree

In Table 3, it can be noticed that quite a high percentage of the students (32.6%, n=56) answered that they neither agreed nor disagreed with the idea. Some of the students appeared to be passive and not to care much about the test methods. This was confirmed by the responses to the open question, 'How do you feel and think about the current test method?'

> "I don't care much about the test method. I am satisfied with the test as long as my teacher gives me a good grade."

> "I have never thought about the test itself seriously. The teacher gives me a test and I just take it."

54.2% (n=13) of the teachers (strongly) agreed that they chose the test because it was more practical while only 29.1% (n=7) (strongly) disagreed with the statement. 58.3% (n=14) of the teachers thought they needed a more effective test method for administering to and assessing the students, and finally to encourage them to improve their speaking skills. 29.2% (n=7) of teachers (strongly) disagreed with it and were using 'a pencil and paper test (and an oral presentation or a group oral interview)' ($\chi 2=52.15$, $p=.01$, $V=.74$). 53.3% (n=88) of students also felt that their test should have changed to better assess and encourage them to improve their speaking skills more properly (33.1% neutral; 13.6% (strongly) disagree).

From the results above, it seems that both teachers' and students' attitudes to their tests do not seem to be particularly positive or negative, neither here nor there. Over half of the teachers chose their test method due to its practicality, but still thought they needed a more effective and efficient test method. Over half of the students also perceived that the test method should change to assess them more properly and to encourage improvement of their speaking skills.

The format of a test may affect test takers' perceptions and attitudes, and it may relate not only to their motivation to learn but also to their test performance, as argued earlier. Hence it is necessary to increase positive attitudes and to reduce negative perceptions of the test. The students with favorable perceptions of their test will perform to the best of their ability. Then teachers and institutions or education authorities will be able to avoid unintended effects on test performance, thereby improving test validity. Further, they will promote beneficial wash-back effects on learning by increasing the students' motivation.

3. Perceptions of the COT

Table 4 shows again that the teachers' perceptions were not very different from those of

the students. Both groups did not show any negative attitudes to the COT. 75% (n=18) of the teachers and 53.2% (n=91) of the students (strongly) agreed that they were likely to use the COT as their exam and only 8.3% (n=2) of the teachers and 22.8% (n=39) of the students (strongly) disagreed with it. Furthermore, only 0% of the teachers and 24% (n=41) of the students (strongly) disagreed that the COT would help to improve speaking skills, although 58.3% (n=14) of the teachers felt neutral. The teachers, however, tended to like the COT less than the students did. Only 29.2% (n=7) of the teachers preferred the COT while 44.4% (n= 76) of the students preferred the COT. A relatively high percentage of the teachers (41.7%, n=10), however, answered that they preferred the FTFI method.

TABLE 4

Perceptions of the COT

	I am likely to use the COT as a mid or final exam.					
	SA	A	N	D	SD	Missing
Participants	%(n)	%(n)	%(n)	%(n)	%(n)	%(n)
Teachers	0(0)	75(18)	16.7(4)	8.3(2)	0(0)	0(0)
Students	15.2(26)	38(65)	22.8(39)	16.4(28)	6.4(11)	1.2(2)
	I prefer the COT to the FTFI.					
	SA	A	N	D	SD	Missing
Participants	%(n)	%(n)	%(n)	%(n)	%(n)	%(n)
Teachers	0(0)	29.2(7)	29.2(7)	41.7(10)	0(0)	0(0)
Students	14.6(25)	29.8(51)	17.5(30)	26.9(46)	9.4(16)	1.8(3)
	The COT will help to improve speaking skills.					
	SA	A	N	D	SD	Missing
Participants	%(n)	%(n)	%(n)	%(n)	%(n)	%(n)
Teachers	0(0)	41.7(10)	58.3(14)	0(0)	0(0)	0(0)
Students	13.5(23)	32.7(56)	28.7(49)	20.5(35)	3.5(6)	1.2(2)

SA=Strongly Agree, A=Agree, N=Neutral, D=Disagree, SD=Strongly Disagree

One of the reasons for the preference of the FTFI format might be that they have the fear of 'doing old things in new ways' as stated in Fulcher (1999, p. 296). The tendency was for them to like doing things in ways they are accustomed to. Cultivating familiarity with the COT through exposure may increase favorable attitudes. Another factor seems to be the teachers thought that the COT would not be practical. One of the teachers commented:

> "I prefer the FTFI format because I will probably need more time to construct test items and it looks complex, but in the FTFI, I can directly interview with the students without any preparation."

It was also confirmed by the answers to the statement asking about the practicality of the COT. Over half of the teachers (62.5%, n=15) neither agreed nor disagreed that the COT would be more practical than the FTFI, and only 29.2% (n=7) agreed with the statement. One of the main reasons for their neutral attitudes was found in responses to the open

question. One of the teachers remarked that she could not answer clearly because she has not experienced the COT:

> "I can not answer clearly because I have never experienced the COT. After experiencing the test format, I would probably be able to respond to the questions more obviously."

In short, the findings indicate that both teachers and students generally perceived the COT positively. However, 41.7% of the teachers preferred the FTFI. Primarily the reasons seem to be that the COT format was not familiar to the teachers and they perceived that the COT would not be very practical compared to the FTFI format. They thought that the COT might be more complex and thus take more time to construct and administer. Some of the teachers and the students showed neutral attitudes because they have not ever actually used it in their classroom and could not completely understand the procedure of the COT and what it really involved.

V. CONCLUSION

This study first attempted to introduce a new type of oral test, which would be practical to administer. Teachers' and students' perceptions of their current oral testing methods with comparison to the COT were then investigated.

It has been shown that most teachers (70.9%) had over 31 students in a class and both the teachers (58.3%) and the students (47.95%) perceived the FTFI as the most appropriate test format for assessing and encouraging students to improve their speaking skills, but only a few teachers (8.3%, n=2) had been using that format due to its impracticality. Practicality was an important factor in choosing their tests. Both the teachers and the students showed neither positive nor negative attitudes towards their current tests, but both groups felt that they needed a more effective test method to more accurately assess, and encourage improvement of, speaking skills. Further, they generally showed favorable attitudes about the COT, although more teachers preferred the FTFI.

In this university context, where computer rooms with numerous multimedia-capable computers including high quality sound cards and speakers are set up, I would like to propose the COT format, which is carried out in the computer room and which is innovative in its capacity to combine several different elicitation techniques in the same test. It allows a large number of students to take the same test at the same time, making it possible to include more items in the test without much effort. The more various items equate greater content and therefore greater construct validity. In addition, it does not need

an interlocutor, which may influence the quality of the test performance. The use of identical questions in the COT will make it easier for untrained teachers to arrive at a decision on the score more easily and quickly. It will also permit the systematic presentation of highly focused and unambiguous stimuli in a way that is not generally feasible in the FTFI setting (Clark, 1979; Clark, 1988). Further, it may be more useful as an achievement test than the FTFI, which is likely to measure what the students have learned in their classrooms-including rather discrete aspects of speaking performance such as vocabulary items and syntactic patterns (Clark, 1988). With all the benefits of the COT, it may lead to better assessment by providing a better balance of reliability, validity and practicality in the assessment of communicative speaking skills.

As with all test types, there are some disadvantages about the COT. The principal concern is a question of authenticity, since there is no human interaction and response elicitation information is restricted to the area of the computer screen. There are many variables, which need to be taken into account when using the COT format such that further research should be conducted to investigate the COT:

First of all, equivalence of direct and semi-direct tests is a most interesting focus for researchers because if it can be established that the tests measure different speaking abilities, then the interchangeability of direct and semi-direct tests is in doubt. Thus, study of the validity of the COT is necessary.

Second, in this study, both teachers and students have not had the opportunity to use the COT in their classroom because of time restriction. If there had been an actual practice of use of the COT in their classroom for a certain period, findings regarding their perceptions of the COT might have been more convincing. Therefore, more close investigation of the perceived advantages and disadvantages of carrying out the COT, and the difficulties in rating performance on the test, through the actual use of the test for a certain period of time in the Korean university context, will help teachers to select the most appropriate test in a given context and for a given purpose.

REFERENCES

Alderson, C. (1991). Dis-sporting life. Response to Alastair Pollitt's paper: giving students a sporting change. In J. C. Alderson, B. North, British Council & International Association of Teachers of English as a Foreign Language. *Language testing in the 1990s: The communicative legacy*. London: Modern English Publications in association with the British Council.

Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation*. Cambridge: Cambridge UP.

Brown, J. D. (1997). Computers in language testing: present research and some future directions. *Language Learning and Technology, 1*(1), 44-59.

Chambers, F., & Richards, B. (1993). Oral assessment: the views of language teachers. *Language Learning Journal, 7*, 22-27.

Clark, J. L. D., & Swinton, S. S. (1979). *An exploration of speaking proficiency measures in the TOEFL context.* Report 4 (No. ETS-RR-79-8).

Clark, J. L. D. (1979). Direct versus semi-direct tests of speaking proficiency. In E. J. Briere & F. B. Hinofotis (Eds.), *New concepts in language testing: some recent studies* (pp. 35-49). Washington, DC: TESOL.

Clark, J. L. D. (1988). Validation of a tape-mediated ACTFL/ILR scale-based test of Chinese speaking proficiency. *Language Testing, 5*, 187-198.

Clark, J. L. D., & Li, Y. C. (1986). *Development, validation, and dissemination of a proficiency-based test of speaking ability in Chinese and an associated assessment model for other less commonly taught languages.* Washington, DC: Center for Applied Linguistics.

Douglas, D. (2000). *Assessing languages for specific purposes*. Cambridge: Cambridge UP.

Fulcher, G. (2000). Computers in language testing. In P. Brett & G. Motteram (Eds.), *A special interest in computers* (pp. 93-107). Manchester: IATEFL Publications.

Kenyon, D. M., & Malabonga, V. (2001). Comparing examinee attitudes toward computer-assisted and other oral proficiency assessments. *Language Learning and Technology, 5*, 60-83.

Kim, K. J. (2006). The beliefs about language learning of Korean college students and their teachers of English. *English Language & Literature Teaching, 12*(3), 1-24.

Larson, J. W. (1998). An argument for computer adaptive language testing. *Multimedia-Assisted Language Learning, 1*, 9-24.

Larson, J. W., & Smith, K. L. (2001). *User's manual for the oral testing software-enhanced, windows edition 1.0.* Brigham Young University.

Luecht, R. M., Champlain, A. D., & Nungester, R. J. (1998). Maintaining content validity in computerized adaptive testing. *Advances in Health Sciences Education, 3*, 29-41.

Malone, M. (2000). *Simulated oral proficiency interviews: recent developments. ERIC Digest* (No. EDO-FL-00-14): Office of Educational Research and Improvement (ED), Washington, DC.

McNamara, T. F. (2000). *Language testing*. Oxford: Oxford UP.

Nam, J. M. (2007). A case study of two Korean students' classroom interactions in a graduate course in America. *English Language & Literature Teaching, 13*(3), 57-75.

O'Loughlin, K. J., & University of Cambridge. Local Examinations Syndicate. (2001). *The equivalence of direct and semi-direct speaking tests*. Cambridge: Cambridge UP.

Roever, C. (2001). Web-based language testing. *Language Learning and Technology, 5*, 84-94.

Shohamy, E., Gordon, C., Kenyon, D. M., & Stansfield, C. W. (1989). The development and validation of a semi-direct test for assessing oral proficiency in Hebrew. *Bulletin of Hebrew Higher Education, 4*, 4-9.

Skehan, P. (1991). Progress in language testing: the 1990s. In J. C. Alderson, B. North, British Council & International Association of Teachers of English as a Foreign Language. *Language testing in the 1990s: the communicative legacy*. London: Modern English Publications in association with the British Council.

Stansfield, C. W. (1991). A comparative analysis of simulated and direct oral proficiency interviews. In S. Anivan (Ed.), *Current developments in language testing*. Singapore: Regional Language Center.

Stansfield, C. W., & Kenyon, D. (1996). *Simulated oral proficiency interviews: an update. ERIC Digest* (No. EDO-FL-96-06): Office of Educational Research and Improvement (ED), Washington, DC.

Stansfield, C., Kenyon, D., Paive, R., Doyle, F., Ulsh, I., & Cowles, M. (1990). The development and validation of the Portuguese speaking test. *Hispania, 73*, 641-651.

van Lier, L. (1989). Reeling, writhing, drawling, sketching and fainting in coils: oral proficiency interviews as conversation. *TESOL Quarterly, 23*(3), 489-508.

Weir, C. J. (1990). *Communicative language testing*. New York: Prentice Hall.

APPENDIX 1

Teachers' perceptions of testing methods in an English conversation class

1. In your English conversation class, how many students do you generally have in each class?
   ①10-20 ②21-30 ③31-40 ④41-50 ⑤more than 50
2. How do you generally assess students' speaking ability in your English conversation class? (Tick all that apply)
   ① A FTFI ② A group oral interview ③ A group project ④A pencil and paper test
   ⑤ A role play ⑥ Other method (please give details)_____

To what extent do you agree or disagree with each of following statements? For responding to items, use the following 5-point scale for each characteristic named below.

① Strongly agree
② Agree
③ Neither agree nor disagree

④ Disagree
⑤ Strongly disagree

3. I chose the current test method because I thought that it was the most appropriate for assessing and encouraging students to improve their speaking skills.
① ② ③ ④ ⑤
4. I chose the current test method because it was practical and easy to conduct compared to other test methods.
① ② ③ ④ ⑤
5. We should look for more practical test methods to assess accurately and encourage students to improve their speaking skills.
① ② ③ ④ ⑤
6. Do you think which testing method is the most appropriate to assess and encourage students to improve their speaking skills?
① A FTFI ② A group oral interview ③ A group project ④ A pencil and paper test ⑤ A role play ⑥ Other method (please give details)_____
7. If you didn't use the testing method you ticked above, what was the reason? _____
8. If I can use the COT in school, I am likely to use it as a mid or final exam.
① ② ③ ④ ⑤
9. If I can use the COT in school, I will prefer the COT to the FTFI.
① ② ③ ④ ⑤
10. I think the COT will be more practical than the FTFI.
① ② ③ ④ ⑤
11. I think the COT will help to improve students' speaking skills.
① ② ③ ④ ⑤
12. Your gender Male _____ Female _____
13. Nationality _____
14. What is your age group? ① 21-30 ② 31-40 ③ 41-50 ④ over 50
15. How long have you been teaching? About _____ years
16. What is the highest level of qualification you have completed?
① Bachelor's degree ② Master's degree ③ Doctor's degree
17. What is your major? _____
Please add any further comments you may have in the space below or back of this sheet.

APPENDIX 2

Students' perceptions of testing methods in an English conversation class

1. How your speaking skills were assessed by the teacher (e.g., a mid or final exam)? (Tick all that apply)
① A FTFI ② A group oral interview ③ A group project ④A pencil and paper test
⑤ A role play ⑥ Other method (please give details)_____
2. How do you feel and think about the current test method(s)? _____
3. Which skills did you particularly improve through preparing for your test?
① speaking ② listening ③ reading ④ writing ⑤ vocabulary ⑥ grammar

To what extent do you agree or disagree with each of the following statements? For responding to items, use the following 5-point scale for each characteristic named below.
① Strongly agree
② Agree
③ Neither agree nor disagree
④ Disagree
⑤ Strongly disagree

4. The current test method was appropriate for assessing and encouraging me to improve my speaking skills.
① ② ③ ④ ⑤
5. The current test method should change to assess me more accurately and encourage improvement of my speaking skills.
① ② ③ ④ ⑤
6. Which test method is the most appropriate to assess and encourage you to improve your speaking skills?
① A FTFI ② A group oral interview ③ A group project ④ A pencil and paper test
⑤ A role play ⑥ Other method (please give details)_____
7. I am likely to use it as a mid or final exam.
① ② ③ ④ ⑤
8. I will prefer the COT to the FTFI.
① ② ③ ④ ⑤
9. I think the COT will help to improve my speaking skills.
① ② ③ ④ ⑤
10. Your gender. Male _____ Female _____
11. Which grade are you in? ① Freshman ② Sophomore ③ Junior ④ Senior
12. Your major. _____

Please add any further comments you may have in the space below or back of this sheet.

Examples in: English
Applicable Languages: English
Applicable Levels: College or University

Mi-jin Joo
Institute of Education, University of London
20 Bedford Way, LONDON WC1H 0AL
Tel: 010-6657-7965
E-mail: ing1115@hanmail.net

Received in Jan., 2008
Reviewed in Feb., 2008
Revised version received in Mar., 2008