

## Introduction

*Speaking* seems intuitively the most important of all the four language skills since people who know a language are referred to as 'speaker' of that language, as if speaking included all other skills of knowing (Ur, 1996). However, speaking skill is not only a crucial part of the language learning process, but also the one skill which has often been neglected in an EFL classroom. In addition, in foreign language proficiency testing, oral performance is one of the most difficult skills to assess since the reliability of scoring has always been doubted.

Recently, the concept of World Englishes (WEs) has been brought out as the fact that there are many types of English existing around the world, and some researchers pointed out that *standard English* should no longer represent the only model or norm for English language use (Kim, 2005). In addition, based on the fact that many learners who had been taught English language by nonnative speakers in the countries of English as a second/foreign language, Kim (2005) claimed that using the rating criteria based on native speakers' standards to measure EFL learners' oral proficiency was not appropriate for the actual use of English in international context. Therefore, it is important for language educators and test designers to reconsider the purposes of teaching oral skills, the goals of speaking tests, and the standards of assessing ESL/EFL learners' oral proficiency, since it cannot be denied that the natural function of speaking is more for meaningful message delivery than the use of language form.

### Statement of the Problem

Traditionally, the major goal of language teaching has been helping learners to develop communicative competence: to master words, sounds, grammar patterns, and sentence structures of the new language. In Taiwan, using the Grammar-Translation Method as the main teaching instruction in English

classroom had been popular for many decades (Huong, 2004). Since the goals of the Grammar-Translation Method mainly emphasized developing students' language knowledge in reading and writing skills instead of oral skills, "there are few opportunities for expressing original thoughts or personal needs and feelings in English" (Bailey, 2005, p. 17).

The idea of the communicative approach has emerged as a new direction in ESL/EFL teaching and learning since the 1970s. In Taiwan, a great deal of attention has been paid to revise EFL teaching materials and curriculum design, and improve teaching facilities to reach the communicative goals. However, the idea of how teachers should improve in doing student evaluations by promoting the communicative approach has been neglected. Nevertheless, even though Taiwanese students can receive good grades in English courses, it does not guarantee that their English oral proficiency has achieved a certain level of competency. In addition, for language teachers in Taiwan, most English language assessments were conducted by pencil-paper tests without considering the importance of oral production in language learning (Cheng, 2006). The reasons that teachers avoided doing oral tests included the amount of time it took, the large size of student population, lack of training in conducting speaking assessment, lack of effective and efficient instruments (Kim, 2003), students' negative reactions toward oral testing, and the teacher's lack of confidence with his or her own target language fluency (Teng, 2005).

From the EFL learner's perspective, oral language proficiency testing is the most complex and difficult task among all of the assessments of the four language skills. Wang (2003) conducted a survey of Taiwanese college students in freshman English classes and she found out that within the four language skills, speaking ability was the one that the students thought they should improve the most (83.7%). That meant that a large part of those Taiwanese college students thought their speaking skill was deficient. Most of the Taiwanese students lacked speaking

practice in the target language both inside and outside of the English classroom. That caused them to have a lack of confidence and an unwillingness to speak. Moreover, they felt panic when pressured into English oral testing.

### **Purposes of the Study**

Based on the situations discussed above, the role of speaking skills has become more important in language teaching and the response to that need has taken the approach of communicative language teaching (Kim, 2005). Moreover, oral language proficiency assessments have become one of the most central issues in language testing needed to be explored. Unfortunately, few studies had been completed with the focus on foreign language speaking skills in Taiwan (Chen, 2001; Li, 2003; Lin, 1996; Pan, 2002; Wang, 2003), and little research in the field of foreign language oral proficiency assessment could be found. Therefore, the present study focused on investigating Taiwanese College English Teachers' preferences and their perceptions of using the types of speaking tasks in assessing students' oral proficiency. The study aimed to answer the following questions:

1. What are college English teachers' general preferences of using the types of speaking tasks in assessing students' oral proficiency?
2. How do college English teachers' background characteristics—gender, native language, English teaching experience to Taiwanese students, certified or uncertified English teachers, and having training or not having training for rating ESL/EFL speaking assessment—influence their preferences of using types of speaking tasks in assessing students' oral proficiency?

## **II. REVIEW OF LITERATURE**

Unlike the written test, which is usually held in scheduled class time in a regular classroom without preparing any special arrangements, an EFL oral test

usually tends to be more difficult to design, administer, as well as score. In addition, the detailed aim and purpose of the oral test has to be set before the test being designed: to avoid the waste of the recourses as well as an inappropriate and unsuitable test being produced. For instance, a placement test should accurately decide and place the right class for a particular learner; on the other hand, a proficiency test should clearly identify the learner's general level of language ability, according to his or her oral performance (Underhill, 1987).

### **Oral Test Techniques**

There are many different elicitation techniques for testing spoken language. The broad aim of all those techniques is to encourage and motivate learners to actually speak by giving them something to speak about; therefore, the teachers can mark the scores based on learners' production. Some speaking tasks are predictable and controllable of learners' response; however, some tasks are not. For instance, the reading-aloud task produces an entirely predictable response, while the face-to-face interview task is less controlled. According to Underhill's (1987) viewpoint that "the majority of oral tests come with two or more elicitation techniques, with a balance between the more controlled and the less controlled" (p. 44).

#### **Reading-aloud**

The technique of the "reading-aloud" task requires the learner to read aloud to the interviewer (rater), either a passage of text or part of a dialogue in which another learner or the interviewer reads the other part. Speaking tests involving reading-aloud, such as the elementary and intermediate levels of the General English Proficiency Test (GEPT), are generally used when it is desired to assess pronunciation, intonation, word and sentence stress patterns, as distinct from the total speaking skills. However, the ability to read aloud greatly differs from the

ability to converse with another person in an informal and flexible way. The advantages of the reading-aloud task are: 1) greater comparability for the purposes of assessment and greater reliability of scores if the same texts are used for all learners, and there is complete standardization of what each learner says; 2) learners easily understand what is required of them; and 3) simple to administer and quick to score.

Some disadvantages exist when using reading-aloud as a task in speaking assessment, for instance, the technique is not authentic or communicative, and in that the task has limited improvement with learners' oral proficiency. Heaton (1988) suggested that a test more useful in many ways than reading-aloud is the "retelling a story" task. In this type of test, the learner is required to retell a story they have just read. If carefully constructed, the "retelling a story" task can not only assess most of the phonological elements which are otherwise tested by reading-aloud, but also measure other skills, such as reading comprehension and organization as well.

### Question-and-Answer

The "question-and-answer" task contains a series of disconnected questions which usually start with short and simple questions, and gradually increase difficulty with long and complex sentences. An experienced rater usually adjusts the types of the questions (easier or more difficult, or change the way of asking) according to the learner's response, as well as places a learner's proficiency level more or less accurately on the basis of the learner's answers to the questions. However, this task cannot develop into a real conversation, and sometimes the test format is to use the pre-recorded test in a language lab, especially for the large-scale standardized test (for instance, the General English Proficiency Test).

As a deliberate strategy for the question-and-answer task in speaking assessment, Underhill (1987) mentioned that the interviewer should post questions clearly of different types, and the interviewee should give response based on the

types of questions as well, such as: a) yes/no questions; b) opened-ended questions; c) descriptive/narrative questions; d) hypothetical questions; or e) justification of opinion questions.

### Oral Interview

Some teachers think of the “oral interview” task as simply a series of question-and-answer. However, the oral interview is more authentic—it can provide a genuine sense of communication—a direct face-to-face exchange and actual conversation between the interviewer and the learner (Madsen, 1983). In general, an interview is structured: the interviewer sets out to find out certain things about the learner, and intends to elicit answers to certain questions. In addition, the interviewer maintains firm control of the situation. After the learner has finished his answer or comment, the interviewer can decide to make the next move, to develop the further topic, or raise a new question. In other words, unlike the “question and answer” task, the oral interview can be explored in detail, with follow-up questions and promoting, to allow the learner to develop and show his or her proficiency, rather than just responded by a straight answer to a straight question (Underhill, 1987).

Madsen (1983, p. 166) pointed out some advantages of the oral interview:

- It can be one of the most communicative of all language examinations.
- It is remarkably flexible in terms of item types that can be included.
- The scoring tends to be more consistent and simple than the scoring of many guided-technique items.

However, oral interview also has its limitations (Madsen, 1983, p. 166):

- It is rather time consuming, particularly if taped and

scored later.

- It is deceptively easy for it to become a simple question-and-answer session.

From Underhill's (1987) point of view, the oral interview often fails to discriminate effectively at a higher level, and its technique is more suitable for testing learners at the intermediate level or below, where the rating scales can present easily-recognizable learner profiles. However, for the higher level learners, the well-defined rating scales are difficult for the interviewers (raters) to make a clear and accurate discrimination. In addition, Underhill (1987) suggested that the interviewer should maintain a natural speed and style of speaking—avoid using “teacherspeak” or the “peculiar form of over-articulated English” (p. 56) during the process of testing.

### Role-Playing

Role-playing activities can be used successfully to test oral communication ability (Heaton, 1988). In a role-playing task, the learners are asked to take particular fictitious roles and are required to improvise in language and behavior. In a one-to-one test situation, the learner has to converse with the interviewer in a way that is appropriate to the role and the situation given. In the classroom setting, the role-playing task can be used for the test which involves a group of students with different characters assigned for each individual to perform, and they interact to each other for a given situation. Before the test, the learners are usually given a set of instructions which explain in simple language exactly what they are supposed to do in certain situations. Or the instructions may be made more specific, to give the learners more direction and to elicit more comparable language from other learners. In order to elicit the learner's exact language ability and proficiency level, the purpose and procedure should be clearly explained to the learner before the test.

For higher level learners, the critical thinking and decision-making skills also can be included into the role-playing task, the learners use the target language to justify their ideas/opinions and seek agreement from the interviewer or the other members of the group.

### Tasks in the Process of Spoken Language Assessment

Speaking assessment is a complicated process since there are numerous components as potential variables involved and interactive during the process of the test, such as rater, scale criteria, test-taker's performance, task, and scoring (McNamara, 1996). It meant that test score as a product from the rating process is an interaction between the candidate, interlocutor, task, testing performance, scale-criteria, and the rater. Therefore, the candidate's (learner's) oral performance is undoubtedly affected by those potential variables. Clarifying those relationships, Skehan (1998) proposed a programmatic model of oral test performance (Figure 1) showing a well-conceptualized interpretation of contextual interactions. In Skehan's model, the role of the task which is placed obviously right in the middle of the whole speaking test process represents its vital responsibility—to elicit candidate's oral performance for assessment by the rater.



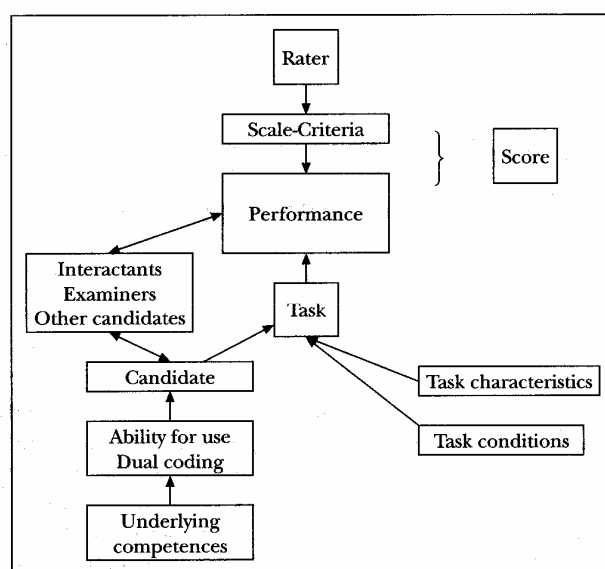


Figure 1. Model of Oral Test Performance (Skehan, 1998, p. 172)

Unlike the large-scale testing situations where the raters have little control over the test tasks, English teachers have more flexibility in employing the tasks and activities in their classrooms to achieve their teaching and evaluating purposes. The tasks in the curriculum planning not only enhance learners' experiential learning as a structured thematic content and activities, but also facilitate the interactions of learner-learner and teacher-learner in the language classroom (Legutke & Thomas, 1991; Candlin, 2001). In language assessments, the tasks are not only facilities to elicit the examinee's potential language ability into their performance, but also reflect the test purposes. Therefore, selecting the most appropriate task for a test should accord with the principles of where and what the test scores will be used for, and what type of information should be provided to the score user (Luoma, 2004).

Bachman and Palmer (1996) defined the task as an activity that “involves individuals in using language for the purpose of achieving a particular goal or objective in a particular situation...[including] both the specific activity and the

situation in which it takes place” (p.44). The task can be administered in a one-to-one setting—with the tester and the test-taker—or in a group or class setting. In either setting, students should feel that they are communicating meaningful content to a real audience. Task-based oral assessment requires candidates to engage in the performance of tasks with the aim of eliciting “authentic” language ability to cope with language demands of the real world situations. Not only did Bachman and Palmer (1996) argue that task characteristics would inevitably impact test scores to some degree, but Wigglesworth (2001) also mentioned that different types of task could influence both quality and quantity of a speaker’s linguistic output. Therefore, some investigations in language testing have addressed the impact on the language of different task characteristics and conditions, such as task variability (Wigglesworth, 2001), face-to-face evaluation versus audio recording (Nambiar & Goon, 1993), individual versus pair testing (O’Sullivan, 2002), the impact of planning time on task performance (Wigglesworth, 1997; Foster & Skehan, 1996; 1997), and task difficulty (Elder, Iwashita & McNamara, 2002; Fulcher & Reiter, 2003).

Robinson et al. (1995) and Wigglesworth (1997) investigated and found out that the types of task with different levels of cognitive demanding do affect learners’ performances. Also, Elder, Iwashita, and McNamara (2002) investigated the impact of performance conditions on task-takers’ perceptions of task difficulty in a test of spoken language, based on Skehan’s cognitive complexity framework. Based on Skehan’s theory, Iwashita & McNamara (2002, p. 249) indicated that task difficulty could be defined by using three different factors:

1. Code complexity: incorporating both linguistic complexity/variety and vocabulary load/variety.
2. Cognitive complexity: involving cognitive processing factors such as information type and organizational structure as well as the familiarity of task topic discourse and genre; and

3. Communicative stress: referring to the logistics of task performance e.g., time pressure, nature of the prompt and number of participants.

The results showed that since there are complex and unstable interactions between different task features and different test-taker attributes (gender, social class, professional experience, and proficiency), the perception of task difficulty seems a multidimensional phenomenon. They concluded that task difficulty could not be accurately estimated after the event on the basis of test-takers' subjective impressions.

Task characteristics and conditions do influence the scores obtained to some degree, as Skehan and Foster (1997) mentioned that:

In oral assessment, close attention needs to be paid not only to the possible variables which can be incorporated or not into the task, but also to the role of interlocutor. This role is central in ensuring that learners obtain similar input across similar tasks. To this end, in developing task-based assessments—particular those for use in the classroom—training needs to include an awareness of the ways in which interlocutors can affect tasks positively and negatively. (p. 206)

Skehan and Foster (1997) also pointed out that it is important to have clear specifications to both assessment task development and the role of interlocutor, in order to make the employed assessment tasks to be more reliable as a goal.

### **III. THE STUDY**

#### **Subject of the Study**

The researcher targeted the college English teachers who have currently taught English courses full-time or part-time at the universities in Kaohsiung, Taiwan, including both native and nonnative English speakers, to participate in this study.

#### **Instrument**

This study aimed to investigate college English teachers' preferences of selecting speaking tasks in the EFL oral proficiency assessment, and how the teachers' characteristic factors influenced their test task selections. The researcher employed a self-design Teacher Survey in this study, including 11 items which regard teachers' personal background information, such as age, native language, academic major, English teaching experience, experience of rating English oral proficiency, and whether they had been trained for rating English language oral proficiency. Those demographic items provided teachers' background information in order to categorize the variables of those teachers.

In this study, individual interviews were also conducted. The researcher further contacted some of the raters who were willing to be interviewed by using telephone, face-to face, or on-line Skype/MSN, in order to provide detailed and richer opinions.

#### **Data Analysis**

Descriptive statistics frequency and percentile were computed and used to describe the demographic information of the participants, as well as to compare those language teachers' preferences of selecting the speaking tasks for oral proficiency assessment.

## IV. RESULTS

The major research interest of this study was to examine and analyze how the English teachers' characteristics as variables influenced their preferences of the test tasks for speaking assessment in the EFL classroom. The results of the study are based on the data gathered from 62 college level English teachers in Kaohsiung, Taiwan. All of them had the experience of teaching the English language to Taiwanese college students.

### **Demographic Data**

A descriptive analysis was first performed on the subjects' responses to Teacher Survey. The items of demographic information regarded the teacher's gender, age, native language, academic major, and level of education. The data regarding the gender of subjects in this study indicated that of the 62 respondents, only fourteen teachers (22.6%) were male, and forty-eight (77.4%) were female. Nearly one-fifth of the teachers were at the age range of 21-30 (19.4%); over half of the teachers were at the age range of 31-40 (56.5%); 14.5% of the teachers were at the age range of 41-50; and 9.6% of the teachers identified their age range as over 50. As for teachers' native language, nine teachers were native English speakers (14.5%), and they were all foreign teachers in Taiwan. The rest of fifty-three teachers were nonnative English speakers (85.5%), and their native language was all Mandarin Chinese.

Teacher Survey also included the items regarding the teachers' English language teaching experience, not specific to college students. According to the data, 17.7% of the teachers had taught English for less than two years; about one-third of the teachers had taught English for 3 to 6 years (35.5%); and nearly half of the teachers indicated that their English teaching experience was more than 7 years (48.4%). As for language teachers' teaching experience to Taiwanese

college students, nearly one-fourth of the teachers had taught less than two years (24.2%); over one-third of the teachers had taught Taiwanese college students English for 3 to 6 years; and 40.3% of the teachers responded that they had taught for more than 7 years.

In terms of the teachers' rating experience for standardized English speaking tests, the data indicated that half of the teachers had little rating experience; there were twelve teachers who had experience less than two years and twelve teachers had experience for 3 to 6 years (38.8%). Seven teachers answered that their rating experiences were more than 7 years (11.3%). As for English teaching certificates, nearly two-thirds of the teachers had English teaching certificates (62.9%). However, regarding language teachers' experience of training for rating ESL/EFL speaking assessment, the data showed that only one-third of the language teachers had been trained—the majority of the teachers (66.1%) had never received any rating training for assessing English speaking skills.

## **Results and Discussion of Research Questions**

Research Question One: What are English teachers' general preferences of using the types of speaking tasks in assessing students' oral proficiency?

Table 1 described the item 11 of Teacher Survey concerning the teachers' responses regarding test tasks which they used in their experiences of assessing students' English oral performance. The responders were allowed to make multiple responses.

As the results showed in Table 1, three of the most popular speaking tasks which had been selected by over half of the language teachers were “face-to-face interview” (62.9% within group), “role-playing” (59.7% within group), and “reading-aloud” (51.6% within group). Besides the three of the most popular tasks, 48.4% of the language teachers had used the “question-and-answer” task, followed by the “pencil-paper test” (46.8% within group) task, and the “pair/group

discussion” and “reporting/presentation” (45.2% within group) tasks.

Table 1

*Frequency of Response for the Overall Test Tasks Used in Speaking Assessment*

Speaking Task	Responses		Percentage of Cases (%)
	Frequency (n)	Percent (%)	
Reading-Aloud	32	13.7%	51.6 %
Pencil-Paper Test	29	12.4%	46.8%
Role-Playing	37	15.8%	59.7%
Face-to-Face Interview	39	16.7%	62.9%
Pair/Group Discussion	28	12.8%	45.2%
Question-and-Answer	30	12.0%	48.4%
Reporting/Presentation	28	4.7%	45.2%
Describing/Telling a Story	11	8.6%	17.7%

However, only eleven out of the sixty-two teachers had ever used “describing/telling a story” as a test task in EFL speaking assessment (17.7% within group).

Research Question Two: How do English teachers' background characteristics—gender, native language, English teaching experience to Taiwanese students, certified or uncertified English teachers, and having training or not having training for rating ESL/EFL speaking assessment—influence their preferences of using types of speaking tasks in assessing students' oral proficiency?

In order to find out whether the teachers' background characteristics related to his or her preferences in selecting test tasks for assessing their students' English oral proficiency, the cross-tabulations would be made to compare the teachers' preferences of test tasks with the characteristics of the language teachers' gender,

native language, teaching experience to Taiwanese college students, their certification status, and having or not having training for rating ESL/EFL speaking assessment.

The sixty-two teachers were divided into two gender groups in Table 2. For male teachers, the “face-to-face interview” task received the highest frequency of usage in assessing speaking (71.4% within group), followed by the “reading-aloud,” “role-playing,” and “question-and- answer” tasks (all three were 57.1% within group).

Table 2

*Frequency of Response for Test Tasks Based on Teachers’ Gender (N = 62)*

Speaking Task	Rater’s Gender			
	Male	% within Group	Female	% within Group
Reading-Aloud	8	57.1%	24	50.0%
Pencil-Paper Test	5	35.7%	24	50.0%
Role-Playing	8	57.1%	29	60.4%
Face-to-Face Interview	10	71.4%	29	60.4%
Pair/Group Discussion	7	50.0%	21	43.8%
Question-and-Answer	8	57.1%	22	45.8%
Reporting/Presentation	4	28.6%	24	50.0%
Describing a Story	1	7.1%	10	16.1%

For female teachers, both the “face-to-face interview” and “role-playing” tasks received the highest frequency (60.4% within group). The data also indicated that female language teachers used the “pencil-paper test” and “reporting/presentation” tasks more than male teachers, while the “question-and-answer” and “pair/group discussion” tasks were used more by male teachers than female teachers. However, the “describing/telling a story” task was



used the least by both female and male teachers: only one male teacher had used this task in assessing speaking skills.

In order to find out how language teachers' native language influenced their preferences in selecting test tasks for assessing oral proficiency, all 62 language teachers were divided into two groups based on their native language: native English speakers and nonnative speakers. As shown in Table 3, the native English speakers preferred the "face-to-face interview" and "question-and-answer" tasks the most (66.7% within group), followed by the "reading-aloud" and "role-playing" tasks (55.6% within group). For nonnative speakers, the "face-to-face interview" task was used the most (62.3% within group), followed by "role-playing" (60.4% within group).

Table 3

*Frequency of Response for Test Tasks Based on Teachers' Native Language*

Speaking Task	Native Language			
	English	% within Native English Speakers	Chinese	% within Nonnative Speakers
Reading-Aloud	5	55.6%	27	50.9%
Pencil-Paper Test	1	11.1%	28	52.8%
Role-Playing	5	55.6%	32	60.4%
Face-to-Face Interview	6	66.7%	33	62.3%
Pair/Group Discussion	5	55.6%	23	43.4%
Question-and-Answer	6	66.7%	24	45.3%
Reporting/Presentation	3	33.3%	25	47.2%
Describing a Story	3	33.3%	8	15.1%

However, while more than half of the nonnative speakers had ever used "pencil-paper test" to test their students (52.8% within group), only one native

English speaker had ever used this task (11.1% with group). It meant that the “pencil-paper test” task was more preferred by nonnative English teachers to employ in testing the EFL students’ oral proficiency, but it was seldom been used by native English teachers. In addition, it was obvious that the native English teachers (33.3% within group) preferred using “describing/telling a story” as a speaking task more than the nonnative teachers (15.1% within group).

Table 4

*Frequency of Response for Test Tasks Based on Teachers’ Teaching Experience to Taiwanese Students (N = 62)*

Speaking Task	Teaching Experience to Taiwanese College Students		
	Less than 2 years (% within group)	3-6 years (%) within group)	More than 7 years (% within group)
Reading-Aloud	7 (46.7%)	13 (59.1%)	12 (48.0%)
Pencil-Paper Test	5 (33.3%)	10 (45.5%)	14 (56.0%)
Role-Playing	8 (53.3%)	15 (68.2%)	14 (56.0%)
Face-to-Face Interview	8 (53.3%)	14 (63.6%)	17 (68.0%)
Pair/Group Discussion	7 (46.7%)	10 (45.5%)	11 (44.0%)
Question-and-Answer	7 (46.7%)	11 (50.0%)	12 (48.0%)
Reporting/Presentation	6 (40.0%)	9 (40.9%)	13 (52.0%)
Describing a Story	4 (26.7%)	2 (9.1%)	5 (20.0%)

Table 4 was developed to understand the relationship between the language teachers’ teaching experience to Taiwanese college students and their preferences of using the tasks for testing English oral proficiency skills. All teachers were divided into three groups based on their teaching experiences: less-experienced (less than 2 years), experienced (3 to 6 years), and well-experienced (more than 7 years).

The results indicated that “role-playing” and “face-to-face interview” were the two most popular tasks for testing oral proficiency in all three groups (53.3% to 68.2% within group). The “reading-aloud” task was more popular in the group of experienced teachers (59.1% within group) than the other two groups, while the “pencil-paper test” task was more popular in the group of well-experienced teachers (56.0% within group). However, “describing/telling a story” was used the least among the eight tasks, especially by experienced teachers (9.1% within group).

Table 5

*Frequency of Response for Test Tasks Based on Certified and Uncertified English Teachers (N = 62)*

Speaking Task	Amount of English Teachers Certified/Uncertified			
	Certified	% within Group	Uncertified	% within Group
Reading-Aloud	19	48.7%	13	56.5%
Pencil-Paper Test	18	46.2%	11	47.8%
Role-Playing	22	56.4%	15	65.2%
Face-to-Face Interview	30	76.9%	9	39.1%
Pair/Group Discussion	17	43.6%	11	47.8%
Question-and-Answer	15	38.5%	15	65.2%
Reporting/Presentation	18	46.2%	10	43.5%
Describing a Story	10	25.6%	1	4.3%

In order to find out whether language teachers' preferences of tasks for speaking tests were influenced by their English teaching certification, all sixty-two teachers were divided into two groups: the teachers who were certified for teaching English language and the teachers who were not.

The results in Table 5 indicated that the majority of the certified English teachers preferred using the “face-to-face interview” as a testing task for assessing EFL oral proficiency (76.9% within group); however, there were only nine uncertified English teachers (39.1% within group) who had ever used this task for testing EFL speaking.

In addition, the “role-playing” and “question-and-answer” tasks were the two most popular test tasks used by those language teachers without any English teaching certification (65.2% within group); however, only about one-third of the certified English teachers had used the “question-and-answer” task (38.5% within group) as the speaking test. Moreover, the results showed that both certified and uncertified English teachers used “describing/telling a story” the least among the eight tasks, especially by uncertified teachers (only one uncertified teacher, 4.3% within group).

Table 6

*Frequency of Response for Test Tasks Based on Teachers' Having or Not Having Training for Rating EFL Speaking Assessment (N = 62)*

Speaking Task	Receiving Rating Training in Speaking Assessment			
	Trained	% within Group	Not Trained	% within Group
Reading-Aloud	8	38.1%	24	58.5%
Pencil-Paper Test	10	47.6%	19	46.3%
Role-Playing	11	52.4%	26	63.4%
Face-to-Face Interview	15	71.4%	24	58.5%
Pair/Group Discussion	11	52.4%	17	41.5%
Question-and-Answer	11	52.4%	19	46.3%
Reporting/Presentation	8	38.1%	20	48.8%
Describing a Story	5	23.8%	6	14.6%

As for investigating whether the language teachers' experience of rating training influenced their selections of tasks for EFL speaking assessment, the sixty-two English teachers were also divided into two groups, based on whether they had been trained for rating ESL/EFL speaking assessment or not.

The results in Table 6 illustrated that the majority of the language teachers who had been trained for rating ESL/EFL speaking preferred the "face-to-face interview" task (71.4% within group), while those teachers who had not been trained preferred the "role-playing" task (63.4% within group).

The results were very similar to Table 5 which compared the certified and uncertified English teachers. Also, for using the "pencil-paper test" task, the percentages of the two groups of teachers in Table 8 (certified and uncertified) and Table 6 (trained and non-trained) were very similar.

With regard to the "reading-aloud" task, the results revealed that the teachers who had not been trained (58.5% within group) for rating ESL/EFL speaking preferred to use the task more than those language teachers who had been trained (38.1% within group). Finally, the "describing/telling a story" task was used the least by both trained teachers (23.8% within group) and non-trained teachers (14.6% within group).

### **Interview Data Results**

The main purpose of the individual interview with the language teachers is to understand better and deeper the reasons why teachers used the particular speaking tasks to assess their students' oral proficiency. There were fifteen raters out of the sixty-two teachers (24.2%) who responded to the researcher's request for a ten-minute interview. The researcher arranged face-to-face interviews, telephone interviews, or on-line interviews with the raters who agreed to participate in an interview. The following section presents the abstracts from the teachers' interview data.

Interview Question: In your English classroom, what types of test tasks do you usually use to assess your students' English language oral proficiency? Why?

"I like to use *presentation* as a speaking test task for my students. Since they are all college students, I believe that *presentation* is a comprehensive task to evaluate their linguistic knowledge, presenting skills, and communicative skills all at once. However, it really takes a lot of time to let everybody finish" (No.42 female nonnative rater with over ten years of teaching experience).

"I prefer using *pair/group discussion* to assess students' speaking skills. First of all, it saves time (I can evaluate two to four students at the same time); second, they will not feel nervous because they are not tested alone; third, they are more willing to talk and speak out; and finally, it can challenge their critical-thinking as well as problem-solving skills. I found my students enjoyed this kind of test" (No.17 female rater, nonnative teacher with three to six years of teaching experience).

"I always use *question-and-answer* to check students' understanding and comprehension, and to test their speaking skills. It will not take much time, and based on their response, you will know if their answers are right or wrong and oral proficiency level right away" (No.1 male native English rater with three to six years of teaching experience).

"The *paper-pencil test* is always the most convenient and objective way to test students: I can control the whole classroom situation. It saves time, and it is fair to everyone. No doubt grammatical rules and vocabulary are also important if you would like to speak the language well" (No.41 female nonnative rater with over ten years of teaching experience).

"*Role-playing* is fun and wonderful to evaluate students' oral performance. Students enjoy this activity: it is not like a test at all. I group them together: they design their own scripts, assign their roles, practice their dialogues, learn the English language through a cooperative way. It is the best way to learn and test the

language” (No.58 female nonnative rater with over ten years of teaching experience).

“I have 58 students in my class, it is impossible for me to conduct face-to-face interviews with each of them...do you know how long that is going to take? I don't think the *pencil-paper test* is an appropriate task to test students' oral skills, but that's the way I can test all students at a time” (No.13 male nonnative rater with over ten years of teaching experience).

“Since some of my students dislike speaking English in class, I made them read aloud in order to have them practice the language directly. Unlike other test tasks, I think *reading-aloud* is not so challenging for students, and it is easier for low-level students to prepare for the tests: they only have to practice more, adjust their pronunciation, tone, and flow of speech” (No. 61 female nonnative rater with less than two years of teaching experience).

## V. SUMMARY, CONCLUSION, AND IMPLICATIONS

As for the teachers' preferences of test tasks for oral language proficiency assessment, the results in this study indicated that “face-to-face interview” was the most popular task (62.9% within group), followed by “role-playing” (59.7% within group), and “reading-aloud” (51.6% within group). Regarding the gender issue, the findings of this study showed that female teachers tended to use the “pencil-paper test” and “reporting/presentation” tasks more than male teachers, while the “question-and-answer” and “pair/group discussion” tasks were used more by male teachers than female teachers. In addition, the “pencil-paper test” task was more preferred by nonnative English teachers to use in testing (52.8% within group), but had seldom been used by native teachers (11.1% within group). As for language teachers' teaching experience, the well-experienced teachers (more than 7 years of teaching experience) used the “pencil-paper test,” “face-to-face interview,” and

“reporting/presentation” tasks more than the experienced (3-6 years of teaching experience) and the less-experienced teachers (less than 2 years of teaching experience). The “reading-aloud,” “role-playing,” and “question-and-answer” tasks were employed the most by the experienced teachers, while the “pair/group discussion” and “describing a story” were more used by the less-experienced teachers than the other two groups.

In terms of certified/uncertified English teachers and trained/non-trained teachers, their preferences were shown similar: the majority of the certified and trained language teachers tended to use the “face-to-face interview” task for assessing students’ oral proficiency, while the majority of the uncertified and non-trained language teachers preferred the “role-playing” task for testing speaking. In addition, the results indicated that the “question-and-answer” task was used by the uncertified language teachers more than the certified teachers, while the “reading-aloud” task was more frequently used by the non-trained teachers than the trained teachers. However, the “describing a story” task had seldom been used among the eight types of tasks for both groups of the teachers.

According to the research data, for most of the teachers, ‘time-consuming’ could be the major concern for testing. According to the raters’ response from the interview, the reason that many teachers preferred the “pencil-paper test” was because it saved time and it was easy to be controlled, especially for a large class size. Around 45% of the teachers had ever used “pair-group discussion” because two to four students could be tested at a time. 59.7% of the teachers had ever employed “role-playing” as a test task for speaking since students could be “grouping together” in testing. However, some teachers realized that “presentation” and “face-to-face interview” came with many advantages in assessing the student’s actual oral performance individually, but those types of tasks did challenge the teachers in certain aspects.

According to Fulcher (2003, p.86), the principal reasons for selecting task



types for a speaking test were:

- Will this task elicit a performance that can be scored?
- Will it be possible to make inferences from the score to the construct the language teachers intend to measure?

Fulcher also mentioned that *task classifications* can provide a direction for test designers to choose the most appropriate tasks for a specific purpose, “giving that tests need to be short enough to be practical and economical, and long enough to be reliable and provide evidence to support valid inferences” (p. 86). Mix of the task types for a speaking test is possible if the purpose of the test is met.

Based on the above issues discussed, to choose the test tasks for assessing EFL students' oral skills, teachers must consider their purposes of the test: to make sure that if the tasks they choose can really elicit oral performance from the students and that is exactly what they target to evaluate about. For instance, critical thinking and problem-solving skills are important for the advanced level language learners, especially college students. To evaluate students' critical thinking and problem-solving skills plus language oral skills, the tasks such as *pair interaction/group discussion* are very useful. On the other hand, for the class with multi-level students or high-anxiety students who do not actively interact with each other, the tasks such as *role-playing* and *dramatic activity* are highly recommended.

In conclusion, according to the research findings of the present study, there were two-thirds of the English teachers who had never been trained in rating EFL oral proficiency. The rating training for oral language proficiency assessment can provide some benefits to the language teachers in certain aspects. The teacher training sections for EFL oral language proficiency assessment will be suggested that it should include: 1) the introduction of assessing communicative language abilities, 2) testing planning and techniques, 3) the introduction of test tasks, rating

scales, and assessment criteria, 4) the illustration of some issues arising in the assessment process, and 5) the practice of scoring samples of various tasks and case discussions. The training sessions aim to not only enhance language teachers' knowledge of testing speaking abilities and understanding the rating process, but also maintain the teacher's judgments to be reliable and consistent. Finally, language teachers should always consider *positive washback*—the benefit that tests offer to learning—tests therefore will be learning devices through which students can receive a diagnosis of areas of strength and weakness, thus having clear study goals.

## References

- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Bailey, K. M. (2005). *Practical English language teaching: Speaking*. New York, NY: McGraw-Hill.
- Bygate, M. (1987). *Speaking*. Oxford: Oxford University Press.
- Candlin, C. N. (2001). Afterword: Taking the curriculum to task. In M. Bygate, Skehan, P. & M. Swain. (Eds.), *Researching pedagogic tasks* (pp. 229-243). England: Pearson Education Limited.
- Cheng, L. (2006, May 10). *New challenge of new TOEFL*. Lihpao Daily. Retrieved November 24, 2006, from <http://publish.lihpao.com/Education/2006/05/10/06d05103/&w>
- Elder, C., Iwashita, N., & McNamara, T. (2002). Estimating the difficulty of oral proficiency tasks: What does the test-taker have to offer? *Language Testing*, 19, 347-368.
- Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition*, 18, 299-323.
- Fulcher, G. (2003). *Testing second language speaking*. Great Britain: Pearson Education Limited.
- Fulcher, G., & Reiter, M. R. (2003). Task difficulty in speaking tests. *Language Testing*, 20, 321-344.
- Heaton, J. B. (1988). *Writing English language tests*. New York: Longman.
- Huong, K. I. (2004). *Students' and teachers' beliefs about communicative language teaching and English grammar instruction in Taiwan*. Unpublished Ed.D thesis, Texas A&M University-Kingsville.

- Kim, H. J. (2005). *World Englishes and language testing: The influence of rater variability in the assessment process of English language oral proficiency*. Unpublished doctoral dissertation, the University of Iowa.
- Kim, H. S. (2003). *The types of speaking assessment tasks used by Korean junior secondary school English teachers*. Retrieved March 15, 2006, from [http://www.asian-efl-journal.com/dec\\_03\\_gl.pdf](http://www.asian-efl-journal.com/dec_03_gl.pdf)
- Iwashita, N., & McNamara, T. F. (2002). Estimating the difficulty of oral proficiency tasks: What does the test-taker have to offer? *Language Testing* 19, 347-368.
- Legutke, M., & Thomas, H. (1991). *Process and experience in the language classroom*. London: Longman.
- Luoma, S. (2004). *Assessing speaking*. Cambridge: Cambridge University Press.
- Madsen, H. S. (1983). *Techniques in testing*. New York: Oxford University Press.
- McNamara, T. F. (1996). *Measuring second language performance*. London: Longman.
- Mehnert, U. (1998). The effects of different lengths of time for planning on second language performance. *Studies in Second Language Acquisition*, 20, 83-108.
- Nambiar, M. K., & Goon, C. (1993). Assessment of oral skills: A comparison of scores obtained through audio recordings to those obtained through face-to-face evaluation. *RELC Journal*, 24, 15-31.
- Robinson, P., Ting, S., & Urwin, J. (1995). Investigating second language task complexity. *RELC Journal*, 25, 35-57.
- O'Loughlin, K. (2002). The impact of gender in oral proficiency testing. *Language Testing*, 19, 169-192.
- Skehan, P. (1998). *A cognitive approach to language learning*. New York: Oxford University Press.

- Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. M. Gass, & C. G. Madden (Eds.), *Input in second language acquisition* (pp. 235-253). Rowley: Newbury House Publishers, Inc.
- Teng, K. H. (2005). *Perceptions of Taiwanese students to English learning as functions of self-efficacy, motivation, learning activities and self-directed learning*. Unpublished doctoral dissertation, the University of Idaho.
- Underhill, N. (1987). *Testing spoken language*. Cambridge: Cambridge University Press.
- Ur, P. (1996). *A course in language teaching: Practice and theory*. Cambridge: Cambridge University Press.
- Wang, Y. C. (2003). *Communication-orientation in Freshman English curriculum: A New response*. In the Papers from the Twelfth Conference on English Teaching and Learning in the Public of China (pp. 88-598). Taipei: Crane Publishing Co., Ltd.
- Wigglesworth, G. (2001). Influence on performance in task-based oral assessment. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic task: second language learning, teaching and testing* (pp. 186-209). Essex, U.K: Longman.

## APPENDIX: TEACHER SURVEY

Please put a mark (√) in the box (□) that applies to you or specify the information about yourself in the other category.

1. Your gender:    ☐ Male            ☐ Female
  
2. Your age:        ☐ 21-30        ☐ 31-40        ☐ 41-50        ☐ Over 50
  
3. Your native language: \_\_\_\_\_
  
4. Your academic major:  
      ☐ Linguistics/English Literature    ☐ TESOL/ ESL Education  
      ☐ Others (specify): \_\_\_\_\_
  
5. Highest level of education:  
      ☐ Bachelor's degree    ☐ Master's degree    ☐ Doctoral degree  
      ☐ Others (specify): \_\_\_\_\_
  
6. How long have you been teaching English?  
      ☐ None    ☐ 0-2 year    ☐ 3-6 years  
      ☐ 7-9 years    ☐ Over 10 Years
  
7. Have you ever taught English to Taiwanese college students? If yes, how long?  
      ☐ None    ☐ 0-2 year    ☐ 3-6 years  
      ☐ 7-9 years    ☐ Over 10 Years

8. Have you had any experience rating non-native speakers' English speaking ability on standardized tests (e.g. placement, selection, exit tests)? If yes, how long?

- ☐ None    ☐ 0-2 year    ☐ 3-6 years    ☐ Over 7 Years

If yes, specify what English speaking tests you have rated before?

9. Do you have an English teaching certificate?

- ☐ Yes    ☐ No

10. Have you ever received training for rating English language oral proficiency?

- ☐ Yes    ☐ No

11. In your English classroom, what types of test tasks are you usually use to assess your students' English language oral proficiency? (more than one answer is allowed)

- ☐ (1) Read aloud  
☐ (2) Pencil-paper test (multiple choice, cloze, short answer, etc.)  
☐ (3) Role-play  
☐ (4) Face-to-face interview  
☐ (5) Pair interaction or group discussion  
☐ (6) Question and answer  
☐ (7) Reporting/presentation  
☐ (8) Describing/telling a story  
☐ (9) Others \_\_\_\_\_

12. I ☐ would NOT like to

☐ would like to participate in individual interview

through ☐ face-to-face

☐ telephone

☐ Skype/MSN chatting.

My contact phone number is: \_\_\_\_\_

E-mail address is: \_\_\_\_\_

Skype/MSN account is: \_\_\_\_\_

(End of the questionnaire. Thanks for your cooperation and kindly help.