Rethinking Assessment of Thai EFL Learners' Speaking Skills

KEMTONG SINWONGSUWAT
Prince of Songkla University, Thailand

Bio Data:
Kemtong Sinwongsuwat has a Ph.D. in English Language and Linguistics from the University of Wisconsin-Madison and has been lecturing at Prince of Songkla University for almost five years after her graduation. She currently teaches both undergraduate and graduate students at the Faculty of Liberal Arts and has a special interest in Conversation Analysis (CA), interactional linguistics, corpus linguistics, and the development of Thai EFL learners' oral communication skills.

Abstract
This paper reassesses the mainstream tasks used for evaluating Thai ELF learners' speaking skills: face-to-face interview and role-play. Based on final and preliminary findings from small-scale classroom research, it examines the capacity of these two tasks to assess the learners' conversation skills in particular and recommends a task that is more oriented towards features of natural conversation; i.e., non-scripted role-play. It is argued that if implemented with an appropriate rubric, this task should enable us to better assess the students' ability to converse in naturally-occurring communicative situations and with practice make it possible to develop students with better proficiency in English conversation. Some implications are also noted for English conversation teaching stemming from the proposed application of the assessment task in Thai EFL classroom contexts.

Keywords: Speaking-skill assessment, face-to-face interview, scripted and non-scripted role-plays, Conversation Analysis (CA), Thai EFL learners, English conversation abilities

Introduction

Nowadays, EFL learners' speaking skills are assessed mainly via the elicitation of talk in such scenarios as in a face-to-face speaking test or in a computer-mediated oral test which is part of a large-scale language proficiency test. Undertaken in the semi-direct test mode in which learners are required to respond to a series of prompts delivered by a computer either online or CD-ROM-based, the latter approach to assessing speaking is gaining popularity as witnessed in its implementation by large assessment organizations such as ETS, Pearson, and Cambridge ESOL (Galaczi, 2010). Despite the benefits of such a test especially in

assessing a large population of test takers, it is however undeniable that for teachers and practitioners in most institutional settings, including classroom contexts, the traditional face-to-face direct test, where learners are required to interact with each other or with the examiner(s), is still more feasible and even necessary.

Internationally, the movement towards communicative language teaching coupled with the publication of such function-oriented proficiency benchmarks for speaking as the 1986 American Council on the Teaching of Foreign Languages (ACTFL) Guidelines has inspired a large body of research into oral proficiency assessment with a communicative measure (Turner, 1998). Over the last couple of decades, there have been observed a growing number of studies investigating various features of oral proficiency interviews, their construct validity, comparability with other test formats, as well as the performance of the parties involved in the test and factors that affect it. Results from these studies have provided crucial empirical evidence especially for developing oral proficiency rating criteria more suitable for particular settings.

On a larger scale of proficiency testing, efforts have mainly been put into developing assessment criteria which are more oriented towards the communicative-competence model of language proficiency and the interactive nature of talk elicited through face-to-face interaction. Questions have been raised especially with regard to the validity and the generalizability of interview-format tests in measuring conversational abilities (Bachman, 1988; Henning, 1992; Lantolf & Frawley, 1988; Lazaraton, 1992; van Lier, 1989; Young & Milanovic, 1992). More recent studies, e.g., Csépes (2010), Galaczi (2010), Xi (2010), took a closer look at the validity issues surrounding the use of computer-based, semi-direct tests to assess speaking proficiency. It was argued that although allowing for a relatively narrow definition of the construct of oral communication ability, these tests are still practical as placement, large-scale survey tests of speaking proficiency.

In a smaller scale of testing such as in school or university settings, there are also attempts to integrate the communicative competence model into speaking assessment. Hayati & Askari (2008), for instance, tried to construct scales more reliable for assessing oral communication abilities, distinguishing communicative from linguistic components. Other studies, e.g. Sook (2003), pointed to teachers' use of non-authentic, uninteractive speaking tasks to assess learners' communication abilities, indicating that generally in practice in EFL classrooms the communicative language teaching and assessing models are yet to be adequately realized, not even to mention the realization of the more recent model of speaking abilities which is oriented more towards the interactional, co-constructed nature of talk (Lazaraton, 1992; Louma, 2004; Young & Milanovic, 1992).

In Thailand, as far as language testing and assessment is concerned, the situation is not so much different; while there are attempts at the national level to modify the English syllabus to accommodate the predominant communicative-competence model, the teaching and testing practices still cannot keep pace with the change (Prapphal, 2008). In fact, students' communicative abilities are still assessed by means of paper-and-pencil multiple-choice tests, particularly in large-scale school and university admission exams. While there are concerns with implementing the national education policy to assure students' overall English proficiency of certain

standard as well as launching programs to deal with teachers' apparently problematic classroom teaching and testing practices, as far as speaking abilities are concerned there is a dearth of empirical studies taking a closer look into what is actually practiced in Thai EFL classrooms, where language teaching and assessing are intertwined. The assessment of oral tasks indeed awaits investigation in future research (Chinda, 2009).

By no means trying to fill in such a gap left for language testing professionals, this paper was written from a teacher, classroom practitioner's perspective in an attempt to examine two major tasks often used in assessing Thai learners' speaking skills; namely, face-to-face interview and role-play. On the basis of the author's classroom observations as well as final and preliminary findings from small-scale classroom research supervised, the paper outlines different interactional features elicited from the two test tasks and proposes role-play of the non-scripted type as an alternative assessment method. It is argued that while proven to be a challenge to most Thai EFL students, non-scripted role-play will help to push them out of their comfort zone to experience more genuine features of naturally-occurring conversation, and, with practice, to enable them to acquire skills necessary for conversing naturally and confidently in real-life communication. Some implications for conversation teaching and assessing are also provided at the end on the basis of the proposed assessment method. It is suggested that we, as teachers and practitioners, consider engaging learners in activities more like naturally-occurring conversation and assessing oral performance with a more appropriate rubric or rating/scoring checklist based on the real nature of genuine conversational interaction.

Speaking Assessment Tasks Generally Used in Thai EFL Classroom Contexts
In fact, there are several speaking-assessment tasks often used by teachers for assessing learners' oral communication skills, for instance, responding orally to question slips, describing and reacting to visual prompts, story-telling, and giving an oral presentation (Chinda, 2009; Khamkhien, 2010; Sook, 2003). However, of all the direct performance-based assessment tasks, face-to-face interview and especially role-play apparently are the most popular choices among teachers for assessing speaking skills of Thai EFL learners since they seem to most approximate natural conversation, the primordial form of oral communication in everyday life.

Interview in particular is considered one of the most popular means of evaluating speaking skills (Underhill, 1998; Weir, 1993). Used in official speaking tests locally and internationally, it is a direct, face-to-face interchange between the learner and the interviewer(s) in which the former's performance is evaluated. In an oral-interview assessment, there is sometimes an assessor present who does not take part in the spoken interaction but listens, watches, and evaluates the learner's abilities. The interview interaction may also be audio/video recorded for further detailed evaluation.

As far as role-plays are concerned, although not explicitly distinguished and examined in the literature, there are two main types often implemented in Thai EFL classrooms: scripted and non-scripted role-plays. In scripted role-plays, learners are asked to work in pairs or small groups. They are given written roles/prompts

related to target scenarios and allowed time to prepare conversation scripts as well as to rehearse the conversations before acting them out in front of the class and/or the teacher. On the other hand, in non-scripted ones, students are assigned to perform the activity on the basis of the class-based prompts given with little or no preparation; they are allowed time neither to prepare nor to rehearse any written scripts in advance.

Generally, in assessing students' speaking performance elicited in the interview and the role-play, scoring rubrics used often include some or all of the following aspects: pronunciation/accent, grammar/structure, vocabulary, content, fluency, and comprehension. It has been argued that rather than conversational competencies, these rubrics are more oriented towards measuring linguistic competencies (Kormos, 1999). Examining conversation proficiency assessment, Ussama & Sinwongsuwat (2012) compared scores obtained from two independent raters' evaluating video-recorded speaking performance of learners with the same low level of English proficiency in interview and two-party peer or role-play interactions with a rubric containing these items. Preliminary findings of the study revealed that the total scores from both test tasks were not significantly different at the 5% level. It was suggested that both interview and two-party peer interactions produced similar results when used in assessing the overall oral English performance of the learners with the traditional rubric.

Differences in Interactional Features of Talk Elicited from Different Assessment Tasks

Despite the similar performance-score results produced by the two assessment activities, i.e. interview and role-play interactions, using the traditional rubric, preliminary close analysis of the learners' talk elicited from the two test tasks using the Conversation Analysis (CA) approach[1] revealed differences in interactional features of the talk elicited. As far as the interactants' organization of conversation is concerned, in the interview interaction, since it was the interviewer who had the right to open and close the conversation as well as to ask questions and introduce new topics, learners did not get to perform any of these actions. They were mostly prompted to answer the questions asked and were not given much opportunity to negotiate for meaning as the interviewer played a dominant role in turn initiation and allocation. In addition, the learners engaged in the interview interaction also had little chance to reformulate trouble turns since most of the turns were somehow repaired by the interviewer. The learners in peer interaction, on the other hand, made their best effort to initiate, collaboratively construct turns and maintain ongoing sequences. Through such interactional features as sequence openings and closings, introducing new topics and topic shift, they were apparently able to exercise the basic rights they normally have in natural conversation in their native language. Additionally, when interacting with their peers, they made attempts to repair problematic turns or shift the topic if the repair attempts failed.

---

[1] The ethnomethodological approach to the analysis of everyday conversation which was originally developed in sociology in the work of Sacks, Schegloff, & Jefferson (1974). CA methods primarily involve fine-grained moment-by-moment analysis of closely-transcribed data of recorded naturally-occurring conversation or talk-in-interaction in general.

As far as turn size is concerned, in the interview assessment examined, since the ability to initiate or extend the topic was controlled by the interviewer, the learner's turns seemed to be smaller and less complex than the interviewer's. In the interview interaction, the interviewer obviously emerged as a more dominating speaker as he/she spoke more, took longer turns, made all the initiation moves, and was the only speaker to expand on his/her own ideas. In contrast, in peer interaction, the quantity of talk was balanced between the two participants with the same level of proficiency. Their rights and duties in the interaction were apparently equally distributed in the turns being constructed.

The talks elicited from the two assessment activities also reportedly differ as far as the overlapping talk is concerned. In the interview interaction, the students proffered only minimal responses such as "yes" or simply a nod. In the case of two-party peer interaction, on the other hand, frequent overlaps mostly involved an extension of the prior topic or a support for the speaker who initiated the topic.

Another salient feature found in the interaction elicited from the two test tasks lies in the difference in repair organization. The transcribed data examined in the study showed that the important type of repair organization occurred in both test tasks was repair for meaning and fluency; grammatical errors were often ignored if they did not obscure the meaning. However, the two assessment tasks apparently differed in the repair strategies used. Collaborative completion of a turn was not found at all in the interview task while a wider range of repair strategies were resorted to in the two-party peer interaction apart from collaborative completion such as asking for repetition, asking for clarification and seeking confirmation.

In fact, the study also showed that the Thai learners in the interview interaction employed more gesturing turns. When confronted with difficult questions asked by the teacher interviewer, they tried to respond to the question first with gestures and later with words even though they mostly failed to produce a complete turn, ending up with a smile and a pause. In peer interaction, without the help of an expert or a high-proficiency speaker, when struggling with certain words, the learner also resorted to gestures to mime the target word. However, in case of failure to finish an ongoing sequence in peer interaction, the learners were found to resort to minimal verbal resources such as acknowledgment tokens, e.g., "yes", fillers, e.g., "uhm" or agreement tokens, e.g. "I think that too" to keep the conversation going.

As preliminarily reported, despite not producing significantly different results when used with the traditional linguistic-oriented rubric, the interview may be more of limited value than peer interaction as far as conversational competencies are concerned. Unlike role-play or peer interaction, it does not seem to allow learners to produce interactional features sufficiently characterizing their conversational competency. Apparently, being an unequal social encounter, it does not seem to create a situation in which natural conversation can occur and learners' conversation skills can appropriately be assessed.

Non-scripted Role-plays: A Better Choice?

Through close analysis of learners' talk, it seems that if we are to assess learners' oral performance, focusing particularly on conversation skills, then role-play may be a more appropriate choice. However, when choosing role-plays, teachers are confronted with two main options: scripted and non-scripted role-plays. Normally, scripted role-plays are a more popular choice among Thai teachers. In this type of role-plays, learners have time to prepare conversation scripts based on given situations and rehearse them for the best performance to be evaluated. Low-proficiency learners, in particular, often find the role-play of this type easier to handle.

Despite the popularity of scripted role-plays in Thai EFL classroom contexts, Chotirat & Sinwongsuwat (2011) argued that non-scripted role-plays may be more effective in helping learners to master relevant features of language use in naturally-occurring conversation. Their investigation of Thai learners' repair in talk obtained from role-plays indicated that both role-play types most elicited self-initiated, self-repair--the type of repair frequently found in natural conversation. However, the practices involved in the repair, which is initiated and carried out by the speaker of the trouble source, were not the same in the two types of role-plays. In scripted role-plays, learners' repair revealed their attempts primarily to adhere to the conversation script that they had rehearsed. They obviously made efforts to recall their own turns, particularly the turns they had cooperatively prepared. Trying to stick to the original script, they performed repair, for instance, when an unfamiliar word was introduced or a misplaced turn was projected. Most of the time when conducting repair in scripted role-plays, learners apparently repeated the units which were deviant from the script, rather than resorting to fillers such as "uh" and "um." Gazing upward, they apparently tried to recall the turns prepared occasionally with some cues from their conversation partners.

In non-scripted role-plays, on the other hand, a good number of self-initiated self-repairs seemed to show students' efforts to finish their turns by collaboratively searching for relevant words and thereby keeping conversation going. When conducting a word search, which is a typical feature of naturally occurring conversation (Kurhila, 2006; Park, 2007), learners often resorted to partial repetition of the problematic unit as well as fillers as turn-holding devices until they came up with appropriate words to bring the ongoing turn to a possible end.

Other-initiated self-repair, i.e., the repair which is initiated by co-participants and carried out by the speaker of the trouble source, also seemed to be organized differently with varying degrees of frequencies in scripted and non-scripted role-plays. Occurring much less in scripted ones, this type of repair was initiated especially when a prior utterance was treated as problematic by the higher-level learner, being accompanied by a request for the speaker of the trouble turn to fix the problem. The repair elicited in this type of role-play therefore seems to allow learners an opportunity to practice making a request, which is a very common action in naturally-occurring social interaction.

It was additionally shown that unlike scripted role-plays, in which problems often got fixed in one sequence, non-scripted ones also allowed learners to fix problems, repeat and reformulate their turns across several sequences, giving them a

better chance to improve their turn construction and conversation organization in real time. Accordingly, it seems that when performing non-scripted role-play activities, learners are provided with better opportunities to deal with problems and to practice relevant features of language use in actual interactions than those in scripted role-plays.

In fact, in a more recent study, Rodpradit & Sinwongsuwat (2012) examined the effectiveness of using scripted and non-scripted role-play activities in enhancing speaking performance of hotel engineering staff. Even though both types of role-play activities were able to help improve the staff members' speaking performance, it was non-scripted role-plays that better contributed to the holistic improvement of their performance according to the post-treatment evaluation done by two native speaker managers in face-to-face on-the-job interaction using the traditional rubric. The study further recommended an investigation of the outcome of using the two role-play types with different groups of learners, especially those with a different level of proficiency.

Although it is still too early to lay claim that non-scripted role-plays are better than scripted role-plays for improving and assessing Thai EFL learners' speaking performance, EFL teachers should seriously consider implementing this type of activity more often in their classroom, taking their learners even further away from their comfort zones. With more practice, Thai learners should be able not only to overcome their fear and speaking anxiety but also to gain enough confidence to break the streotypical barrier of them being shy to speak English.

Implications for Classroom Conversation Assessing and Teaching

With the recommendation of non-scripted role-plays as activities to improve and assess Thai EFL learners' conversation performance come not only alternative scoring rubrics more appropriate for assessing learners' conversation skill but also some conversation-teaching ideas related to the application of such a rubric. When using this task type to assess learners' conversation skill, the rubric should be adjusted to accommodate features of naturally occurring conversation, rather than simply focusing on discrete items such as pronunciation, fluency, grammar, vocabulary, and comprehension as traditionally practiced. Learners' conversation should be perceived as a sequentially organized social event in which several actions are performed via the interactional practices of different linguistic resources on a moment-by-moment conditional-relevance basis to accomplish certain social goals. Reflecting this perception, the rubric should be oriented towards target social actions or interactional functions learners need to be able to perform as well as towards their ability to manage or organize the conversation such that interactional goals can be accomplished.

For instance, to assess learners' ability to make a telephone conversation, such a conversation competency rubric may contain items such as initiating a summon, responding to the summon, displaying recognition or self-identification, greetings, making initial inquires, managing topics (i.e., initiating, maintaining/expanding, and shifting topics), moving to closing, and closing (Luke & Pavlidou, 2002). The assessor should be oriented towards how learners organize both verbal and non-verbal interactional resources (e.g., vocabulary, grammar, stress, intonation, gaze,

gestures, postures, etc.) into sequences of turns of various actions as well as in conversation repair so that things can get done and communicative goals can be reached (see also O'Sullivan, Weir, & Saville, 2002).

Certainly, a great deal of empirical research is still needed in Thai EFL classroom contexts to nail down the strengths and weaknesses of different speaking-assessment tasks implemented with different groups of Thai learners and mostly Thai assessors and to arrive at a practical rubric with different criteria to be used in the assessment of the learners' conversation skill. Nevertheless, it should not be too early to begin in the conversation classroom to make both teachers and especially adult learners aware of the fact that conversation has its structure. And it is not simply a linguistic phenomenon requiring the mastery use of discrete lexical, phonological and grammatical resources but a socially organized activity in which certain things get done through interactions between social participants with these as well as other language resources.

Being the basis of any kind of talk, ordinary conversation has a structure to which talk-participants are oriented in their sequential organization of turns and in their interpretation of one another's social actions in those turns. Made aware of the structure of conversation, learners can themselves take notice of different kinds of sequences into which a conversation is organized and their variations from opening to closing. In fact, the canonical sequence that they should be able to recognize first and foremost is the adjacency pair sequence, an easily noticeable sequence of two-part actions one of which is contingent upon the other, for instance, summon-answer, greeting-greeting, invitation-acceptance/decline, etc. The concept of adjacency-pair sequences should be introduced in relation to the preference organization system. Namely, in response to any first pair-part turn, they have choices in shaping their turns into a preferred or a dispreferred response, each of which has different social implications. Therefore, with both verbal and nonverbal resources available, learners should, for instance, be taught or made aware of how to shape not only a preferred turn to accept a request but also a dispreferred turn to decline it. Additionally, in keeping a conversation going, they should also be aware of the important roles of acknowledgment and assessment tokens, e.g., "yeah", "umm", "right", "good," etc, as well as repair strategies to deal with conversation problems when they arise.

Especially, it is very important for both teachers and learners to recognize that a conversation has order at all points; phenomena such as fillers, stuttering, pauses of various lengths, overlap and laughter, which they might have disregarded as speech perturbations, are in fact meaningfully and purposefully produced. For instance, fillers such as "uh" and "uhm" may serve as turn-holding devices. Broken utterances and stuttering can be used as devices to get attention from addressed recipients or to preface speakers' dispreferred responses, e.g., a decline to a request (see also Barraja-Rohan, 2009).

These concepts of conversation should be introduced to learners through transcribed excerpts of genuine talks obtained from speakers of the target language. With these concepts in mind, it should be easier for learners to take notice of the structural organization of the conversations to which they are exposed both in and outside the classroom, to make sense of social actions embodied in turns, as well as

to predict what is going on in the conversation and how it unfolds. Consequently, with awareness-raising practice, they should be able to acquire norms of particular conversation scenarios as well as their related interactional features. And production practice coupled with awareness-raising activities should enable us to improve our learners' conversation abilities more effectively.

Conclusion

Although much research is needed, it is definitely not too early for us to start questioning the effectiveness of the traditional means of teaching and especially assessing our EFL learners' conversation skills. And this paper should be considered successfully serving its purpose if it has at least been able to get readers to do just that. While it is fairly obvious, when it comes to speaking, that our goal as teachers and practitioners is to develop learners who possess not only linguistic but also conversational competencies, it is not always straightforward how to achieve that goal. However, as suggested in this paper based on classroom observations and research, one possible place for us to start is to reexamine the traditional view we have long held that conversation is mainly a linguistic event and the traditional way to assess it with the traditional rubric and to turn to a more current view of conversation as a highly-ordered, socially-organized event. This is not to say that linguistic competencies as traditionally assessed are irrelevant and to be disregarded. But one may question to what degree they can be assessed as actually realized via performance-based assessment activities such as scripted role-play. And we just need to find ways to balance both linguistic and conversational competencies in both teaching and assessing.

Once we ourselves start to look at conversation seriously and try to understand it in a different way, we should begin to notice what is going on when our students are engaged in conversation in the target language and see clearer the pros and cons especially of the popular teaching and assessing methods currently adopted. Only then, we will be able to find more effective ways of teaching and assessing to help them develop more desirable conversation abilities.

References

Bachman, L. F. (1988). Problems in examining the validity of the ACTFL oral proficiency interview. *Studies in Second Language Acquisition, 10*(2), 149-164.

Barraja-Rohan, A. M. (2009). *Teaching conversation to adult learners of English with conversation analysis and politeness pragmatics.* Retrieved from http://eslandcateaching.wordpress.com/about/

Chinda, B. (2009). *Professional development in language testing and assessment: A case study of supporting change in assessment practice in in-service EFL teachers in Thailand* (Doctoral dissertation). University of Nottingham, UK.

Chotirat, S., & Sinwongsuwat, K. (2011). Effects of scripted and non-scripted role play activities on oral performance: A case study of repair organization in conversation of Thai college students. *Proceedings of the 3rd International Conference on Humanities and Social Sciences.* Faculty of Liberal Arts, Prince of Songkla University, Thailand.

Csépes, I. (2010). Validity issues in face-to-face versus semi-direct tests of speaking. In L. Araújo (Ed.), *Computer-based assessment of foreign language speaking skills* (pp. 52-62). Retrieved from http://publications.jrc.ec.europa.eu/repository/bitstream/111111111/15037/1/lbna24558enc.pdf.

Galaczi, E. D. (2010). Face-to-face and computer-based assessment of speaking: Challenges and opportunities. In L. Araújo (Ed.), *Computer-based assessment of foreign language speaking skills.* Retrieved from http://publications.jrc.ec.europa.eu/repository/bitstream/111111111/15037/1/lbna24558enc.pdf.

Hayati, A. M., & Askari, E. (2008). Testing oral language proficiency of university EFL students. *Asian EFL Journal, 29*, 26-37.

Henning, G. (1992). The ACTFL oral proficiency interview: Validity evidence. *System, 20*(3), 365-372.

Khamkhien, A. (2010). Teaching English speaking and English speaking tests in the Thai context: A reflection from Thai perspective. *English Language Teaching, 3*(1), 184-190.

Kormos, J. (1999). Simulating conversations in oral-proficiency assessment: A conversation analysis of role plays and non-scripted interviews in language exams. *Language Testing, 16*(2), 163-188.

Kurhila, S. (2006). *Second language interaction.* Philadelphia, PA, USA: John Benjamins Publishing Company.

Lantolf, J. P., & Frawley, W. (1988). Proficiency: Understanding the construct. *Studies in Second Language Acquisition, 10*(2), 181-195.

Lazaraton, A. (1992). The structural organization of a language interview: A conversation analytic approach. *System, 20*(3), 373-386.

Luke, K. K., & Pavlidou, T-S. (2002). *Telephone calls: Unity and diversity in conversational structure across languages and cultures.* Philadelphia, PA, USA: John Benjamins Publishing Company.

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

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2002). Design and analysis in task-based language assessment. *Language Testing, 19*(4), 477-496.

O'Sullivan, B., Weir C. J., & Saville, N. (2002). Using observation checklists to validate speaking-test tasks. *Language Testing, 19*(1), 33–56.

Park, I. (2007). Co-construction of word search activities in native and non-native speaker interaction. *Working Papers in TESOL & Applied Linguistics, 7*(2), 1-23.

Prapphal, K. (2008). Issues and trends in language testing and assessment in Thailand. *Language Testing, 25*(1), 127-143.

Rodpradit, S., & Sinwongsuwat, K. (2012). Using scripted and non-scripted role-plays to improve English speaking performance of hotel engineering staff. *APHEIT Journal, 18*(1).

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language, 50*(4), 696-735.

Sook, K. H. (2003). The types of speaking assessment tasks used by Korean Junior Secondary school English teachers. *Asian EFL Journal.* Retrieved from http://www.asian-efl-journal.com/dec_03_gl.pdf.

Turner, J. (1998). Assessing speaking. *Annual Review of Applied Linguistics, 18,* 192-207.

Underhill, N. (1998). *Testing spoken language.* Cambridge: Cambridge University Press.

Ussama, R., & Sinwongsuwat, K. (2012). Conversation proficiency assessment: A comparative study of two-party peer interaction and interview interaction implemented with Thai EFL learners. *Proceedings of the Phuket Rajabhat University National Conference 2012*, Phuket, Thailand.

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

Weir, C. J. (1993). *Understanding and developing language tests.* United Kingdom: Prentice Hall.

Xi, X. (2010). Automated scoring and feedback systems: Where are we and where are we heading? *Language Testing, 27*(3), 291-300.

Young, R., & Milanovic, M. (1992). Discourse variation in oral proficiency interviews. *Studies in Second Language Acquisition, 14*(5), 403-424.