Can EFL Interactive Listening Be Validly Assessed?*

CHENG Jingyan
Beijing Information Science & Technology University

Richard B. BALDAUF Jr
University of Queensland, Australia

Abstract

Although passive listening (i.e., hearing language) may occur in some EFL classes through oral reading, oral presentations, or other oral forms, listening instruction is a relatively neglected skill in EFL teaching, and as a consequence, assessment of listening has received relatively limited coverage in the language testing literature. Interactive listening, based on the idea that listening is real-life communication, has been demonstrated to be a dimension of listening worth teaching. However, no successful classroom interactive listening teaching format has yet been reported, let alone assessed. Based on the models designed by previous researchers, an interactive listening teaching format was designed and trailed a university classroom intervention program in Beijing. In order to do this, it was necessary to construct an objective interactive listening assessment which could measure the listener's ability to: (a) negotiate meaning, and (b) to comprehend the speaker's messages (Macaro, 2003; Rost, 2002). To verify its reliability and validity, a standard listening test, the types of negotiation, and self-assessment were also adopted. The results show that compared with traditional listening tests, this interactive listening assessment procedure provided a more comprehensive and objective evaluation of listeners' listening skills.

Key words: interactive listening assessment; final examination; self-assessment

1. Introduction

While research on language acquisition indicates that listening comprehension is central to successful learning (e.g., Brindley, 1998; Ellis et al., 1994; Faerch & Kasper, 1986; Feyten, 1991), assessment of listening ability has received relatively limited coverage in the language testing literature, which may be a consequence of the difficulties associated with listening research (Brindley, 1998; Macaro, 2003). The relatively low profile of listening assessment may reflect the inherent difficulties involved in describing and assessing an invisible cognitive operation. In this regard, a number of reviews of listening comprehension have identified the lack of empirically sound models which could be used to guide testing (e.g., Brindley, 1998: 171).

The concept of interactive listening, which it has been argued is a reflection of the true nature of real-life communication, is now an established part of L2 listening methodology (e.g., Anderson & Lynch, 1988; Rost, 1990; Oprandy, 1994). Unfortunately, to date, no reported empirical and successful classroom teaching format that teachers can easily adopt to implement interactive listening has been proposed, let alone implemented. However, based on the models proposed by Lam and Wong (2000), Murphy (1985) and Field (2002), the authors have designed and trailed an interactive listening format for use with Chinese learners. As a consequence, there was a need to construct an objective and comprehensive procedure to measure learners' listening proficiency, which provides the central focus of this article.

2. Interactive listening

2.1 The tenets of interactive listening

Before examining the possibilities for interactive listening and how its effectiveness might be assessed in a Chinese classroom context, a brief description of the tenets and typical practices of interactive listening are in order. Interactive listening refers to listening in collaborative conversation, in which learners interact with each other or with native speakers. It has been established as a vital means of language development as it requires listeners to display signs of partnership and understanding to fulfil their roles as successful participants in the discourse (Rost, 1990, 2002). Vandergrift (1997) also has argued that interactive listening requires listeners to interact with an interlocutor, requesting clarification or providing feedback in order to ensure successful communication. Since the mid-1980s, interactive listening has attracted growing attention in pedagogic thinking and practice (Bahns, 1995), and Vandergrift (1997) has predicted that given the importance of interactive listening in ordinary social discourse, it could also play a more general and important role in L2 classrooms. Furthermore, Vandergrift (2004) has argued that interactive listening is the technique that listeners engage in most frequently, and some researchers would argue that this is the only dimension of listening worth teaching (e.g., see the debate between Ridgway, 2000a; 2000b and Field, 2000).

With the development of research on the role of interaction between learners and

interlocutors in L2 learning, researchers have focused on a particular variety of interaction often referred to as negotiation. Negotiation between learners and interlocutors takes place during the course of their interaction when either one signals with questions or comments that the other's preceding message has not been successfully conveyed (Pica et al., 1996). A wide variety of question types have been suggested as signals of the negotiation process including: repetition, paraphrase, verification, clarification, elaboration, extension and challenge (Morley, 2001: 81).

Negotiation plays a key role in improving listening comprehension, and is believed to aid second language acquisition (SLA). For example, Krashen's (1982) "comprehensible input hypothesis" stressed the significance of comprehensible input, where the aural reception of language should be just a little beyond the learner's present ability. At the same time, researchers have also shown that negotiation for meaning makes input more comprehensible, and modification is an efficient way to bring about such negotiation (Krashen, 1982; Long, 1985). This suggests that the learning process underlying interactive listening is negotiation, which is the procedure that some researchers have tried to develop for classroom teaching. According to Swain (1985, 2000), one of the potential benefits of interactive listening is compelling the learner to formulate ideas in the target language and in "forcing negotiation", that is compelling the learner to come to understand language which is initially not understood. Following from this principle, a format is proposed for interactive listening which research indicates might be suitable for use in Chinese classroom contexts.

2.2 Interactive listening teaching format

Once having understood the tenets of interactive listening in the context of the Chinese classroom, the question remains how to implement it. Although classroom listening teaching is not the same as ordinary social discourse, test designers can try to imitate aspects of real world communication, in particular, the listening-thinking-negotiating-response procedure. Although there are no empirically sound models to turn to, models designed by Lam and Wong (2000), Murphy (1985) and Field (2002) provide a starting point. Getting inspiration and learning lessons from these three models, we would propose the following format for interactive listening in the Chinese classroom context.

There are three steps required for implementing interactive listening. First, learners are asked to listen for comprehension; second, they discuss the content to fulfil productive activities; and finally, they give oral presentations in pairs. The three steps should be systematically related in classroom teaching. These steps should be preceded by a pre-listening phase to set the context, to provide motivation and to teach important new vocabulary.

<u>Step 1:</u> **Cooperative Listening (Listening)**
• In the first play of the listening material, listening for gist, getting learners to negotiate with each other and comparing their understandings. Different ideas are encouraged.
    • In the second play of the listening material, listening for details. Learners act upon opportunities to paraphrase, ask or answer questions, introduce expansions,

add contextual clarification, verification or repetition. They share their answers and interpretations, and can revise their views after negotiation. Teachers should not provide answers, but encourage learners to argue with each other.

　　• In the third play of the listening material, if learners still are unsure of what they have heard, they interact with the recording directly, interrupting the replay as required for clarification or verification, controlling the pace and discussion. The playback of the recording should not proceed until learners feel secure in their mutual understanding. Teachers can interject and provide the necessary modifications or clarifications when learners become bogged down. Once learners feel secure about their understandings, a final version can be publicized.

　　During the first and second listening, learners have to negotiate with each other to make input comprehensible, where the true meaning of the input is being negotiated and arrived at by listener-listener interaction. If they miss some information, they can resort to their partner for repetition. If they have different ideas from their partner, they can try to recall the text and cite examples to search for correctness. If they are not sure about some information in the recording, they can discuss it with their partner for verification. If they are confused about the text, they can ask for clarification. During the third listening, the learners' control of the pace of the recording allows them to interact with the listening material directly, and with the teacher, acting as a resource, when necessary. These procedures should avoid the difficulties former researchers have encountered. With process and product interwoven, listening becomes a much more interactive activity among teachers, learners, and the recorded material.

Step 2: Communicative Production (Post-listening 1)
During the communicative production stage, the teacher provides learners with an opportunity for meaningful communicative behaviour about relevant topics closely related to the listening materials. Several different types of activities should be provided so that learners can actively select one they are interested in, providing greater motivation and avoiding the chance of imposing stereotypical tasks on learners. These types of activities, as previously mentioned, can be group decisions recorded in the form of statements, choices among alternatives, problem solving, debate, role play, interview, etc. Once an activity is chosen, they work in pairs to generate ideas and opinions.

Step 3: Oral Presentation (Post-Listening 2)
During the oral presentation stage, pairs of learners are invited to address their colleagues and receive feedback. Unlike traditional individual presentations, learners still work in pairs, giving everyone an opportunity to make a public presentation in limited time after a safe and supportive rehearsal. As different pairs may demonstrate different activities, the oral presentations are more interesting and colourful. Pair presentation is very popular with Chinese learners, who are inclined to be shy and inhibited for fear of losing face in public. Working in pairs helps them overcome the anxiety involved in speaking English, and they are able to support each other during the presentation when one of them becomes stuck.

2.3 Interactive listening assessment

The cognitive demands made on the listener are much heavier in interactive listening (Lynch, 2002; Rost, 2002) since the listener must process the input in real time, clarify understanding when comprehension is uncertain, critically evaluate what is understood, and then respond (Vandergrift, 2004). As a consequence, this type of listening is more difficult to test, and is rarely sampled in language tests (Brindley, 1998: 174; Schrafnagl & Cameron, 1988: 88). However, Brindley's (1998: 174) arguement that the key to interactive listening is negotiation between listeners and interlocutors has lead some experts in testing (e.g., Ross & Langille, 1997) to argue that listening tests should attempt to incorporate opportunities for the negotiation of meaning. The inclusion of negotiation in listening tests will make the content more natural and authentic, consequently conducting a separate assessment of listening using an oral interview would be a promising choice (Brindley & Ross, 1997; Ross & Langille, 1997). Brindley (1998) has also argued that different options for incorporating the notion of negotiated interaction in non-participative listening tests need to be explored, and that there is a need to investigate ways to assess interactive listening ability in direct tests of oral interaction such as interviews.

Although the oral interview provides an appropriate setting to test interactive speaking and listening, the notion of construct validity in these tests has been challenged (Rost, 2002). In an oral interview test, the test candidate is placed in the role of the listener and is expected to respond to the interviewer's "prompts", which are usually questions. As a consequence, average oral interview tests are more like a process of elicitation of specific output and compliance with routines than a "normal" interactive conversation (Rost, 2002), although the listener is supposed to be a participant in the communicative interaction. Thus, virtually all work in testing listening comprehension, both research and actual test construction, has been concerned with non-collaborative listening. The ability to perform the listener's role in an interactive, collaborative situation has generally been tested as a speaking skill, usually in an interactive interview type assessment (Buck, 2001). All the problems previously mentioned are due to the difficulty in designing a collaborative listening test. Assessing how well listeners can perform in collaborative interactive situations requires actual communication between two or more participants, and then one or more people to rate the test takers' performance. Setting up such interactive assessments is expensive and very time-consuming (Buck, 2001: 98).

However, the integration of authentic interaction into second language tasks opens a Pandora's Box in terms of test method effects, with a multiplicity of factors potentially affecting listening test performance (Ross & Langille, 1997). However, interactive listening skills, as the reflection of real life communication, are important enough to be included in interactive assessment such as the oral interview (Brindley, 1998; Buck, 2001). Thus, it would be interesting to develop a procedure whereby what was being measured was the listener's ability to (a) negotiate meaning and (b) to arrive at comprehension of the speaker's messages (Macaro, 2003; Rost, 2002).

Taking the previous discussion into consideration, one can see that in any research done to examine this problem, it would be better if negotiation in the oral interview was included and learners' comprehension were measured. But an oral interaction test

alone would not be sufficient. For one thing, in this area of research, interactive listening assessment might work, and so it is not clear whether this test, by itself, is empirically sound. For another, it is also assumed that learners' listening skill, including transactional listening—where they listen to radio or watch TV—has improved after the interactive listening format is implemented. Consequently, there is a need to provide triangulation to examine the validity of this methodology. To do this, three types of assessment were selected for use. The first was a standard listening test, where a pre-test before interactive listening teaching format was given, followed by an end of semester post-test. The second was an interactive listening assessment task designed by the first author. The third was a self-assessment questionnaire, where students were invited to give their opinion about their current listening teaching and to make judgments about their own listening abilities.

3. Designing an interactive listening assessment procedure

The next two sections outline the design of the interactive listening procedure that has been developed and the self-assessment questionnaire used to assess students perceptions of their progress attained through that interactive listening.

3.1 The design of interactive listening assessment

When designing an interactive assessment procedure, one not only had to look at the specific steps involved, but also to consider the types of questions that would be most appropriate and the role of the interviewer in this new procedure.

3.1.1 Interactive listening assessment procedure

Although a significant proportion of listening occurs in interaction, there is no procedure that measures listening skills during oral interaction, with these skills being constantly brushed aside in favour of measuring productive oral skills (Macaro, 2003). Furthermore, although there has been some research on speaking in oral interviews (see, e.g., Brown 2002, McKay, 2007; McNamara, 1996), there appears to be a noticeable lack of specific research on listening assessment in those oral interview situations. This article attempts to address this problem based on previous research findings.

If test tasks could be made more like tasks found in real life, the short answer questions (SAQs) found in an oral interview provide a realistic activity for testing second language listening comprehension (Weir, 1993). However, the performance produced by the candidate needs to be judged based on their appropriate response to questions and appropriate use of clarification questions rather than on linguistic correctness. The procedure for this type of assessment is as follows. First, the candidates listen to a tape recording and take notes concurrently. Then, they are given the questions and hear the text read by the interviewer for the second time. This generally follows what would normally occur when we listen to lectures and take notes. Where students are given the questions in advance, there is a danger that they will process input differently, i.e. they will be listening for specific cues indicating that the answer to the next question is coming up (Weir, 1993). An experiment by Sherman

(1997) also showed that significantly higher scores were achieved with this "sandwich" format. It was also the format which created the least anxiety and was strongly preferred by the test-takers over both the "questions-before" and "questions-after" conditions. After learners read the questions, an interviewer reads the same article at the same speed for the second time. Learners are told they are free to negotiate with the interviewer whenever lack of understanding emerges. They can stop the interviewer to negotiate the meaning whenever they need clarification. When the interviewer has finished reading, it is the students' turn to answer questions, which are like responses in real life interactions. According to Lund (1991), these responses can be oral, non-verbal, or written. In our case, given the time-limits and the need for authenticity, oral responses were chosen. In sum, the procedure for interactive listening assessment is as follows:

First listening: Students simply listen to the whole text, and write notes if necessary. (Questions are then distributed and students are given 2 minutes to look at them.)

Second listening: An interviewer reads the text, and students can negotiate with him/her when no or lack of understanding occurs. (Students answered the questions in sequence)

3.1.2 Factors affecting the objectivity of listening assessment

To maximize the objectivity of this type of interactive listening assessment, two factors emerge from the literature that need to be taken into consideration: 1) the type of questions or tasks, and 2) the role of interviewer.

Types of question or task. Lynch (1995) has divided types of questions or tasks used into two broad categories: divergent and convergent. Studies that used divergent tasks, such as open-ended discussions to investigate interaction, discovered that this type of task does not always encourage negotiation of meaning. On the other hand, convergent tasks seemed to promote more substantial interaction among interlocutors and give them more opportunities to use confirmation checks and clarification requests. Weir (2005) also has argued that if we wish to preserve the authentic nature of the listening materials, we have to focus questions on more global processing skills, which enable us to extract meaning from a spoken text (e.g., gist, main ideas, and speaker attitude/intention). At the same time, if the learner has to answer written questions while listening to continuous discourse, some of the main ideas or important details might be missed while an answer to a previous question is being recorded. But, if the response produced by the learner can be brief and limited, the danger of the writing process interfering with the measurement of listening is restricted. Questions also need to be spaced out appropriately in the text to avoid interference (Weir, 2005). Furthermore, Shohamy and Inbar (1991) found that "local" questions on factual details and the meanings of lexical items were easier to answer than "global" questions on main ideas and inferences. Consequently, to avoid the problems mentioned previously when conducting oral interviews, questions posed by the interviewer should be largely convergent and on specific content, but allowing for a few global questions on the learners' main impression of the article at the end of the listening session.

The role of interviewer. Investigations of interviewer behaviour have identified the interviewer as a significant variable in test administrations (Brown, 2002; Lazaraton, 1996;

Ross, 1992; Ross & Berwick, 1992; Young, 1995; Young & Milanovic, 1992). Wigglesworth (2000) also found that interlocutor behaviour in oral assessment tasks influenced both the quality and quantity of the language produced by learners and, hence, the scores that they obtained. So, when deciding to choose a teacher or fellow-learner as an interviewer, it is necessary to look for someone who will have the least negative impact on learners' performance. However, when the teacher is the interviewer, the literature suggests there may be some problems. For example, Farrell and Mallard (2006: 340) found evidence to support Pica (1987) and Liebscher and Dailey-O'Cain's (2003) previous studies of teacher-learner interactions, indicating that in most FL classroom interaction, teachers and learners "engage in a social relationship which affords them unequal status as classroom participants" and that this relationship, in turn, inhibits successful FL comprehension. In addition, learners did not use unspecified repairs with the teacher when they experienced comprehension problems.

In China, where education is deeply influenced by Confucian ideology, Chinese students normally do not interrupt a teacher for clarification, as they feel it is rude, or they may be reluctant or feel embarrassed. These factors suggest that a teacher is not very suitable as an interviewer in an interactive listening assessment. At the same time, a lot of studies have found that learner-to-learner interaction will produce a non-threatening atmosphere to lessen learners' anxiety, which in turn gives learners more confidence to ask for clarification and for them to correct each other (e.g., Crookes & Chaudron, 2001; Doughty & Pica, 1986; Long & Porter, 1985; Rulon & McCreary, 1986; Varonis & Gass, 1985). Based on this research, to lessen learners' anxiety and ensure learners' normal performance, a patient student from each class with very good spoken English was invited to be the interviewer and interview other students individually.

3.2 Self-assessment questionnaire

In addition to the interactive listening assessment and standard pre- and post-tests, a self-assessment questionnaire was administered to the students. Self-assessment, as used in second language learning contexts, is a technique utilized in some placement procedures that involve asking second language learners to make judgments about their own abilities or self-rating (Heilenman, 1991; Rost, 2002; Schwartz, 1985; Smith & Baldauf, 1982). LeBlanc and Painchaud (1985) found positive correlations between self-assessment questionnaires and standardized proficiency tests for listening and reading. They concluded that "self-assessment must be considered a very valuable tool as a placement instrument" (p. 673). Birckbichler et al. (1993) utilized a self-assessment questionnaire as a part of the placement exam administered at Ohio State University in the US and reported that self-assessment correlated higher than any other variable with the placement scores. Participants also reported accurate ability levels. Brantmeier (2005) found that levels of self-assessed abilities positively correlated with levels of enjoyment. These investigations lend support to the hypothesis that self-assessment can be accurate for placement. However, using self-assessment instruments with Asian students to rate their listening is still untried. Nevertheless, it forms part of the assessment used in this instance.

To keep teacher informed of learners' needs, questions on learners' attitudes towards

the interactive listening teaching method were also included. Thus, the questionnaire consisted of two parts. Part A of the questionnaire gathered students' opinions about the listening teaching methods using a 5-point Likert scale after which students were invited to write their own comments and suggestions on the listening teaching method. Part B of the questionnaire included students' self-assessment related to their L2 listening abilities. (See Tables 1 and 2.)

The design of the self-assessment questions used in Part B of this study was based on a Spanish reading questionnaire by Brantmeier (2006). The self-assessment factors were measured by four questions, two completed before the interactive listening assessment and two completed after the assessment. A 5-point Likert scale was used in to encourage more precision in rating and to encourage respondents to show greater discrimination in their judgements. The first two self-assessment rating questions for completion before learners listened to the passages were based on the instrument utilized by Birckbichler et al. (1993) and Deville and Deville (1999) for language placement at a university. The questions were about L2 listening, e.g., "How do you rate yourself as a listener of English?" with five possible choices ranging from 1 (I do not think so) to 5 (I think so). (See Table 3 for the self-assessment questions.) The objective of pre-self-assessment items was to obtain a general assessment of students' perceived proficiency. The latter two self-assessment rating questions were created for use after the students completed interactive listening assessment in order to contextualize items and obtain an assessment of comprehension of the specific listening tasks utilized for the study. The two questions for this assessment were taken from a questionnaire created by Schraw et al. (1995) and Tobias and Everson (1998) and Everson (Chief Research Scientist, College Board of USA, personal communication, November, 2004) respectively. The questions are "I found the passage I just listened to easy to understand" and "How much of the passage did you understand?" (See, Table 4.)

4. Research methodology

Having implemented a probing interactive listening teaching format, the previously described package of listening assessments was used to examine its effectiveness.

4.1 Participants
A quasi-experimental research design was developed in which two classes of 30 students majoring in industrial design in Year 2 at Beijing Information Science and Technology University were randomly designated as: (a) an experimental group which received listening instruction using an interactive listening teaching format, and (b) a control group which was taught with the traditional testing listening format. The two classes had the same teacher, the same material and the same number of class hours dedicated to listening.

4.2 Procedures
This study ran over an 18 week semester, and both the experimental and control group took the same pre-test at the beginning of the semester and post-test at the end of

semester, those tests being the final university-wide class examinations for the previous semester and for the current semester. To have a closer look at the change in the learners' standard test results, the results of a semi-final exam held in the 9th week were also included. Interactive listening procedures were employed in one class, and both oral interview assessments and self-assessment questionnaires were held throughout the course. Scores on these informal tests provided the teacher with an objective means of checking learners' progress in the course. To check how well listeners could perform in an interactive listening situation, which requires actual communication between the interviewer and other students, 10 of the 30 students in each class were randomly assigned to one of three groups to be interviewed in the 6th, 12th and 18th weeks, thereby testing all the students. Questionnaire A and part of Questionnaire B were distributed to students before the interactive listening assessment, and the remaining part of questionnaire B, after the assessment. Each assessment lasted approximately 10 minutes.

To reduce the variability in the interactive listening assessment, the materials used for interactive listening assessments were taken from the model test for CET 4 (College English Test, Band 4) in 2006, with questions largely convergent or on specific content but with a few global questions at the end. This provided a way to ensure the same level of difficulty of the listening materials for each of the three assessment periods.

In terms of the interviews themselves, three pilot interviews were conducted to gain experience and to judge the effectiveness of having students interview each other. It was found that these interviews save time, and it is much more relaxing when two learners were tested at the same time rather than just one. The two student format elicited more negotiation between the peer learners, than was the case with the external interviewer. During this procedure, the teacher watched the interview and marked the types of negotiation the students used, allocating a score to each student.

5. Results

5.1 Standard tests results
On the traditional standard tests, that is, the pre-test (Test 1), semi-final test (Test 2) in the 9th week and post-test (Test 3) in the 18th week, with 25 being a full grade for listening comprehension, the average grades and standard deviations for the two classes are shown in Table 1. All of the three tests were held university-wide and the exam questions were issued by the Beijing municipal education commission.

Table 1. The means and standard deviations of standard tests

	Test 1	Test 2	Test 3
Control group	10.9 (1.689)	10.3 (3.346)	14.9 (2.815)
Experimental group	9.7 (1.658)	10.1 (3.652)	15.3 (3.711)

Comparing the t-test results for the three tests, at the pre-test (Test 1), the control group was significantly better ($t = 2.777$, $p = 0.0037$) than the experimental group, while at semi-

final test (Test 2) and post-test (Test 3), there was no significant difference ($t = 0.221$, $p = 0.8287$; $t = -0.470$, $p = 0.6801$) between the groups. The data in Table 1 shows that the experimental group made more improvement (5.6 points or 22.4%) than the control group (4 points or 16%) over the 18 weeks, steadily increasing its scores across the three testing times. In addition, the relatively big standard deviation suggests that some students were benefiting more from instruction than others.

5.2 Interactive listening assessment

Interactive listening assessments were held three times during the semester. Before interactive listening began, both experimental and control groups were taught how to negotiate with their interviewer and the negotiation types the students used were recorded. Three passages and 10 SAQs (short answer questions) were presented, and students' average grades are shown in Table 2.

Table 2. The means and standard deviations of interactive listening assessments

	Test 1	Test 2	Test 3
Control group	4.8 (2.201)	4.5 (1.841)	4.0 (1.699)
Experimental group	4.6 (1.713)	4.7 (2.163)	5.1 (2.378)

Comparing the t-test results for the three tests, at the pre-test (Test 1) ($t = 0.393$, $p = 0.6959$) and semi-final test (Test 2) ($t = -0.386$, $p = 0.3506$), the experimental and control groups were not significantly different, while at the post-test (Test 3), the experimental group had significantly higher results than the control group ($t = -2.062$, $p = 0.0219$).

Table 2 shows that the experimental group showed clear improvement in listening skills compared with the control group over the 18 weeks, steadily increasing its scores across the three testing times. However, this result could be due to the experimental group's greater familiarity with the testing technique as it was imbedded in their interactive teaching, since the control group's result declined over the same period.

What is important in understanding these results is to have a look at the negotiation type these students adopted in the oral interview. Due to time constraints, the student interviewer interviewed two students at a time during each interactive listening assessment. During the assessments, students in both the experimental and control group were found to be quite reluctant to interrupt the student interviewer to negotiate when no or lack of -understanding occurred. In the control group, only 27% or 8 of the 30 students ever implemented negotiations. This frequency did not change much during the three assessment periods with requests for repetition of sentences or words accounting for 62%, and requests for clarification accounting for 38% of interruptions. In the experimental group, the students seemed to feel more at ease with negotiation as 40% or 12 of the 30 students negotiated with the interviewer to get more information about the listening material. In the first interactive listening assessment for the experimental group, only one negotiation type was used—repetition. In the second assessment, the students had learned

to control the interviewer's speed and asked for clarification and paraphrase. Requests for repetition accounted for 40% and clarification 40%, with the remaining 20% being paraphrases. However, in the third assessment, the ten students did not ask any questions, and some mentioned that they could master the meaning, and they did not think questioning was necessary. Perhaps their listening proficiency skills had increased to the point that they did not need to rely on negotiation.

The experimental group showed a pattern similar to that of previous studies that had investigated the relationship between types and frequency of negotiation and learners' listening proficiency levels (e.g., Bremer, Roberts, Vasseur et al., 1996; Farrell & Mallard, 2006; Rost & Ross, 1991; Vandergrift, 1997b), that is, intermediate learners will use more types and a higher frequency of negotiation than learners of higher or lower proficiency levels. This result across testing periods provides further evidence that learners' listening proficiency is improved with the implementation of interactive listening teaching, and supports the improvements found in the interactive listening assessment and standard listening test.

5.3 Self-assessments questionnaire

5.3.1 Pre-self-assessment

The self-assessment questionnaire was also administered three times. Both the experimental and control group were given the same questions to obtain a general idea of students' perceived listening proficiency. There were four self-assessment questions, two of which were given as pre-self-assessment items. The mean scores for the two statements on a five-point Likert scale ("not very well" to "very well") are given in Table 3 for the experimental and control groups.

Table 3. Experimental and control groups on the pre-self-assessment

Statement	Time 1	Time 2	Time 3
1. How well can you understand English? (Control)	2.6	2.3	2.1
1. How well can you understand English? (Experimental)	2.2	2.6	2.5
2. How do you rate yourself as a listener of English? (Control)	2.5	2.1	2.2
2. How do you rate yourself as a listener of English? (Experimental)	2.0	2.6	2.2

Table 3 shows that the result of the pre-self-assessment is very different from those of the standard test and interactive listening assessments, in which the experimental group's grades increased the second and third times of assessment. What is clear is that although the experimental group was not as confident as the control group at the beginning, their confidence increased after interactive listening was implemented and reached a peak in the 12th week. However, the learners' confidence seemed to decline in the last 6 weeks. Perhaps this can be explained by a realization of the effectiveness of interactive listening at the very beginning of its implementation, when students showed an increasing passion and confidence in their listening skills. However, as the semester progressed, their poor

oral English, shyness, the multiple-choice format of the up-coming CET 4 examination, and many years of playing passive roles in the classroom gradually wore down their patience and enthusiasm. In much of Asia (and elsewhere), exams are very dominant, and when faced with a choice of a "safe" study procedure or an unknown one, students will choose the safer one.

5.3.2 Post-self-assessment

To have a closer look at the relationship between their confidence and their real proficiency, self-assessment ratings were also collected for use after students completed the interactive listening assessment in order to contextualize items and gain an assessment of comprehension of the specific listening material utilized in the study. Responses to these two statements on a five-point Likert scale ("strongly disagree" to "strongly agree") are given in Table 4.

Table 4. Experimental and control groups on the post-self-assessment

Statement	Time 1	Time 2	Time 3
3. I found the passage I just listened to easy to understand. (Control)	2.3	2.5	1.9
3. I found the passage I just listened to easy to understand. (Experimental)	2.0	2.5	2.2
4. How much of the passage did you understand? (Control)	2.6	2.6	2.3
4. How much of the passage did you understand? (Experimental)	2.6	2.6	2.9

Table 4 shows that although the experimental group ranked themselves as poor listeners, their perceived understanding of listening materials did improve, and this suggests that there was some improvement in their listening proficiency, although it was not noted by them. This result may be explained by the Confucian ideal which values modesty, and which has a strong influence on Chinese people's self-evaluation. Normally, people make negative remarks about themselves, no matter how capable they are.

6. Conclusions

An examination of the three types of assessment shows that in the experimental group, interactive listening assessment has produced the same trend as the standard test result, i.e., scores of both steadily increased after interactive listening was implemented. Even though learners' perceived confidence in their listening proficiency declined after it peaked mid-term, their perceived understanding of listening material gradually improved, that is to say, the results of these three assessments produced parallel trends.

These parallel trends suggest that interactive listening assessment can be used as a reliable measure of learners' interactive listening skills, while at the same time the self-assessment questions can be used as a compensatory instrument, which can be used to get information about learners' needs and their emotion changes so as to adjust classroom teaching practices.

References

Anderson, A. & Lynch, T. 1988. *Listening*. Oxford: Oxford University Press.

Bahns, J. 1995. There's more to listening than meets the ear (retrospective review article). *System*, 23(3), 531-547.

Birckbichler, D., Corl, K. & Deville, C. 1993. The dynamics of language program testing: Implications for articulation and program revision. In D. P. Benseler (ed.), *The Dynamics of Language Program Direction*, 162-178. Boston, MA: Heinle & Heinle.

Brantmeier, C. 2005. Non-linguistic variables in advanced L2 reading: Learner's self-assessment and enjoyment. *Foreign Language Annals*, 38(4), 493-503.

Brantmeier, C. 2006. Advanced L2 learners and reading placement: Self assessment, CBT, and subsequent performance. *System*, 34, 15-35.

Bremer, K., Roberts, C., Vasseur, M., Simonot, M. & Rroeder, P. 1996. *Achieving Understanding: Discourse in Intercultural Encounters*. London and NewYork: Longman.

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

Brindley, G. & Ross, S. 1997. Trait-method comparisons across three language test batteries using exploratory, MTMM and structural equation modelling approaches. Paper presented at 19th Annual Language Testing Research Colloquium. Orlando, Florida, March.

Brown, A. 2002. Interviewer variation and the co-construction of speaker proficiency. *Language Testing*, 20(1), 1-25.

Buck, G. 2001. *Assessing Listening*. Cambridge: Cambridge University Press.

Crookes, G. & Chaudron, C. 2001. Guidelines for language classroom instruction. In M. Celce-Murcia (ed.), *Teaching English as a second or Foreign Language*, 29-42. Boston, MA: Heinle & Heinle.

Deville, M. & Deville, C. 1999. Computer adaptive testing in second language contexts. *Annual Review of Applied Linguistics*, 19, 273-299.

Doughty, C. & Pica, T. 1986. "Information gap" tasks: Do they facilitate second language acquisition? *TESOL Quarterly*, 20, 305-325.

Ellis, R., Tanaka, Y. & Yamazaki, A. 1994. Classroom interaction, comprehension and the acquisition of L2 word meanings. *Language Learning*, 44, 449-91.

Faerch, K. & Kasper, G. 1986. The role of comprehension in second-language learning. *Applied Linguistics*, 7, 257-274.

Farrell, T. & Mallard, C. 2006. The use of reception strategies by learners of French as a foreign language. *The Modern language Journal*, 90, 338-352.

Feyten, C. M. 1991. The power of listening ability: An overlooked dimension in language acquisition. *Modern Language Journal*, 75, 173-180.

Field, J. 2000. "Not waving but drowning": A reply to Tony Ridgeway. *ELT Journal*, 54, 186-195.

Field, J. 2002. The changing face of listening. In J. C. Richards & W. A. Renandya (eds.), *Methodology in Language Teaching* (Chapter 22). New York: Cambridge University Press.

Heilenman, K. 1991. Self-assessment and placement: A review of the issues. In R. V. Teschner (ed.), *Assessing Foreign Language Proficiency of Undergraduates, AAUSC Issues in Language Program Direction*, 93-114. Boston: Heinle & Heinle.

Krashen, S. 1982. *Principles and Practice in Second Language Acquisition*. Oxford, England: Pergamon.

Lam, W. & Wong, J. 2000. The effect of strategy training on developing discussion skills in ESL

classroom. *ELT Journal*, 54, 245-255.

Lazaraton, A. 1996. Interlocutor support in oral proficiency interviews: The case of CASE. *Language Testing*, 13(2), 151-172.

LeBlanc, R. & Painchaud, G. 1985. Self-assessment as a second language placement instrument. *TESOL Quarterly*, 19(4), 673-687.

Liebscher, G. & Dailey-O'Cain, J. 2003. Conversational repair as a role-defining mechanism in classroom interaction. *Modern Language Journal*, 87, 375-390.

Long, M. 1985. Input and second language acquisition theory. In S. M. Gass & C. G. Madden (eds.), *Input in Second Language Acquisition*, 377-393. Rowley, MA: Newbury House.

Long, M. & Porter, P. 1985. Group work, interlanguage talk and SLA. *TESOL Quarterly*, 19(2), 207-228.

Lund, R. 1991. A comparison of second language listening and reading comprehension. *Modern Language Journal*, 75(2), 197-204.

Lynch, T. 1995. The development of interactive listening strategies in second language academic settings. In D. Mendelsohn & J. Rubin (eds.). *A Guide for the Teaching of Second Language lListening*, 166-185. San Diego, CA: Dominie Press.

Lynch, T. 2002. Listening: Questions of level. In: R. B. Kaplan. (ed.), *Oxford Handbook of Applied Linguistics*, 39-48. Oxford: Oxford University Press.

Macaro, E. 2003. *Teaching and Learning a Second Language: a Review of Recent Research*. London / New York: Continuum.

McKay, P. 2007. *Assessing young Language Learners*. Cambridge: Cambridge: University Press.

McNamara, T. F. 1996. *Measuring Second Language Performance*. London: Addison Wesley Longman.

Morley, J. 2001. Aural comprehension instruction: principle and practices. In M. Celce-Murcia (ed.), *Teaching English as a Second or Foreign Language*, 69-85. Boston, MA: Heinle & Heinle.

Murphy, M. J. 1985. Examining ESL listening as an interpretive language process. *TESOL Newsletter*, xix(6), 23-24.

Oprandy, R. 1994. Listening/speaking in second and foreign language teaching. *System*, 22(2), 153-175.

Pica, T. 1987. Second language acquisition, social interaction, and the classroom. *Applied Linguistics*, 8(1), 3-21.

Pica, T., Lincoln-Porter, F., Paninos, D. & Linnell, J. 1996. Language learners' interaction: How does it address the input, output, and feedback needs of L2 learners? *TESOL Quarterly*, 30, 59-84.

Ridgeway, T. 2000a. Listening strategies—I beg your pardon? *ELT Journal*, 54, 179-185.

Ridgeway, T. 2000b. Hang on a minute! A reply to John Field. *ELT Journal*, 54, 196-197.

Ross, S. 1992. Accommodative questions in oral proficiency interviews. *Language Testing*, 9(2), 173-186.

Ross, S. & Berwick, R. 1992. The discourse of accommodation in oral proficiency interviews. *Studies in Second Language Acquisition*, 14, 159-176.

Ross, S. & Langille, J. 1997. Negotiated discourse and interlanguage accent effects on a second language listening test. In G. Brindley & G. Wigglesworth (Eds.), *Access: Issues in English Language test Design and Delivery*, 87-116. Sydney: National Centre for English Language Teaching and Research, Macquarie University.

Rost, M. 1990. *Listening in Language Learning*. London: Longman.

Rost, M. 2002. *Teaching and Researching Listening*. New York: Longman.

Rost, M. & Ross, S. 1991. Learner strategies in interaction: Typology and teachability. *Language Learning*, 41, 235-273.

Rulon, K. A. & McCreary, J. 1986. Negotiation of content: Teacher-fronted and small-group interaction. In R. R. Day (ed.), *Talking to Learn*, 182-199. Rowley, MA: Newbury House.

Schrafnagl, J. & Cameron, D. 1988. Are you decoding me? The assessment of understanding in oral interaction. In P. Grunwell (ed.), *Applied Linguistics in Society*, 88-97. London: Centre for Information on Language Teaching and Research.

Schraw, G., Bruning, R. & Svoboda, C. 1995. Sources of situational interest. *Journal of Reading Behavior*, 27(1), 1-17.

Schwartz, L. 1985. The California study on foreign language articulation. *ADFL Bulletin*, 17(1), 22-28.

Sherman, J. 1997. The effect of question preview in listening comprehension tests. *Language Testing*, 14, 185-213.

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

Smith, K. F. & Baldauf, R. B., Jr. 1982. The concurrent validity of self-rating with interviewer rating on the Australian Second Language Proficiency Ratings Scale. *Educational and Psychological Measurement*, 42(4), 1117-1124.

Swain, M. 1985. Communicative competence: Some roles of comprehensible input and output in its development. In S. Gass & C. Madden (eds.), *Input in Second Language Acquisition*, 235-253. Rowley, MA: Newbury House.

Swain, M. 2000. The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In J. P. Lantolf (ed.), *Sociocultural Theory and Second Language Acquisition*, 97-114. New York: Oxford University Press.

Tobias, S. & Everson, H.T. 1998. Research on the assessment of metacognitive knowledge monitoring. Paper presented at a Symposium on "Metacognition: Assessment and Training." Annual conference of the American Educational Research Association, San Diego, CA.

Vandergrift L. 1997. The Cinderella of communication strategies: reception strategies in interactive listening. *The Modern Language Journal*, 81(4), 495-505.

Vandergrift, L. 2004. Listening to learn or learning to listen? *Annual Review of Applied Linguistics*, 24, 3-25.

Varonis, E. M. & Gass, S. M. 1985. Non-native/non-native conversations: A model for negotiation of meaning. *Applied Linguistics*, 6, 71-90.

Weir, C. J. 1993. *Understanding and Developing Language Testing*. New York: Prentice Hall.

Weir, C. J. 2005. *Language Testing and Validation: An Evidence-based Approach*. New York: Palgrave Macmillan.

Wigglesworth, G. 2000. Issues in the development of oral tasks for competency based assessments of second language performance. In G. Brindley & G. Wigglesworth (eds.), *Studies in Immigrant English Language Assessment. Volume 1*, 81-124. Sydney: National Centre for English Language Teaching and Research, Macquarie University.

Young, R. 1995. Conversational styles in language proficiency interviews. *Language Learning*, 45(1), 3-42.

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

(Copy editing: Janet ROBERTS)