

# A Historical Review of the Development of Listening Assessments: Pedagogical Implications to English Teaching and Testing

*Suthathip Thirakunkovit*

English Program, Faculty of Liberal Arts

Mahidol University, Thailand

suthathip.thi@mahidol.edu

## Abstract

Developing listening skills is an essential part in acquiring a second language. Without listening skills, learners cannot communicate face to face with one another. For over 50 years, several approaches were used to teach listening e.g., the audio-lingual, direct-method, and communicative approach. However, the process of developing listening skills are still difficult to understand. The role of listening comprehension skills in classrooms began to receive attention from researchers a few decades ago. In the case of listening assessment, it received only a limited range of coverage in the language testing literature. Moreover, a lack of a conclusive definition of listening constructs contributes to the difficulty in measuring listening skills. In spite of the imbalance in the literature, in this study, I will investigate a number of research studies related to teaching and assessing listening skills and explored how listening comprehension and listening process have been defined, how L1 and L2 listening comprehension are different, and what are the common difficulties experienced by L2 learners. Ultimately, I hope that this review can be a guideline to provide ways for further research on listening comprehension which will, hopefully, enhance teaching and assessing listening comprehension ability.

**Keywords:** Listening assessments; English learners; historical review

## พัฒนาการของการทดสอบทักษะการฟัง

บทบทวนวรรณกรรมและนัยที่มีต่อการเรียนการสอน และการสอบภาษาอังกฤษ

### บทคัดย่อ

การพัฒนาทักษะการฟังเป็นสิ่งสำคัญในการเรียนรู้ภาษาที่สอง เพราะถ้าผู้เรียนไม่มีทักษะการฟังที่ดี ผู้เรียนจะไม่สามารถสื่อสารแบบต่อหน้าได้ ในช่วงเวลาที่ผ่านมา เทคนิคการสอนฟังได้มีการพัฒนาอย่างหลากหลาย เช่น วิธีสอนแบบฟัง-พูด (*audio-lingual method*) วิธีสอนแบบตรง (*Direct Method*) และ วิธีสอนเพื่อการสื่อสาร (*Communicative Method*)

ถึงแม้กระนั้น กระบวนการพัฒนาทักษะการฟังก็ยังเป็นเรื่องที่มีความซับซ้อน หัวข้อการพัฒนาทักษะการฟังในห้องเรียนเริ่มได้รับความสนใจจากนักวิจัยเมื่อไม่กี่ทศวรรษที่ผ่านมา ยิ่งในกรณีของการทดสอบทักษะการฟัง ก็มีงานวิจัยกล่าวถึงน้อยมาก อีกทั้งคำนิยามของทักษะการฟังก็ยังไม่ได้มีข้อสรุปที่แน่ชัด ดังนั้นจึงทำให้งานวิจัยทางด้านนี้ยากยิ่งขึ้นไปอีก ถึงแม้จะมีข้อจำกัดดังกล่าวมา ในบทความชิ้นนี้ ผู้เขียนได้สำรวจงานวิจัยจำนวนหนึ่งที่เกี่ยวข้องกับการเรียนการสอน และการทดสอบทักษะการฟัง และได้ค้นหาคำนิยามของทักษะการฟังที่ได้มีผู้กล่าวไว้ นอกเหนือไปจากนี้ ผู้เขียนได้ค้นคว้า ข้อมูลเกี่ยวกับทักษะการฟังในภาษาแม่กับในภาษาที่สอง ว่ามีความแตกต่างกันอย่างไร และผู้เรียนภาษาที่สองส่วนมากประสบปัญหาในการฟังอย่างไรบ้าง ในกรณีนี้ ผู้เขียนหวังเป็นอย่างยิ่งว่า บทความชิ้นนี้จะมีส่วนช่วยในการพัฒนาการเรียนการสอน และงานวิจัยชิ้นต่อไป ที่เกี่ยวเนื่องกับทักษะการฟัง

**คำสำคัญ** การทดสอบทักษะการฟัง ผู้เรียนภาษาอังกฤษ การทบทวนวรรณกรรม

## 1. Introduction

In many ways, it is not surprising that a number of English learners perceive listening comprehension as one of the most challenging skills in language learning. A study of Graham (2006) shows that listening comprehension is the skill that most students feel they have achieved the least. The students have highlighted some of their major difficulties that their problems tend to be related to the speed of delivery and the lack of listening practices. In addition, they could not help themselves trying to identify every individual word when they listen to a stream of spoken language in order to interpret the full message.

As Buck (2001) has emphasized in his book, listening skills involve complex processing. The learners must use a wide variety of language knowledge, including linguistic and non-linguistic knowledge when interpreting the incoming speech. Even though it is generally accepted that listening requires a combination of language skills, the factors contributing to effective listening is still not clearly understood (Graham, 2006).

In order to understand the development of teaching and assessing listening skills, it is important to discuss the definition of listening and the significant features of listening comprehension, how the listening comprehension skill differs from other language skills. What are the common pedagogical models for L2 listening comprehension and the frameworks for testing listening comprehension? What are different test formats used in assessing listening comprehension? And what could be potential issues and challenges related to listening assessments?

## 2. What is the Listening Comprehension Skill?

Traditionally, researchers believed that comprehension was a construct that could be applied to both listening and reading (Osada, 2004). Listening was viewed and taught alongside reading because they were usually defined as receptive skills. Since many characteristics of listening comprehension skills actually involve many forms of reading

comprehension (Buck, 2001), listening is seen as the receptive skill in the oral mode, comprising both an interpretive and analytical process (Anderson & Lynch, 1988; Lundsteen, 1979). When we speak of listening comprehension, what we refer to is listening and understanding what we particularly hear and listen to. However, there is considerable research evidence showing that listening ability is a unique process in comparison with reading comprehension (Buck, 1992). However, what is listening comprehension process exactly, and what is the crucial part in the listening process that creates a distinction between listening and other language skills?

### **2.1. How is Listening Comprehension Process Defined?**

Rost (1994) defines listening as a skill that underlies all verbal communication. However, a perusal of the literature reviews indicates that there is still no generally-accepted definition of the term *listening comprehension* because listening is a complex skill, involving different kinds of knowledge, and many researchers in different disciplines have defined more than fifty different ways, depending on how researchers intend to apply the definition (Purdy, 1997). For example, Rost (1994) defines the term listening as “a process that is triggered by our attention. ... and purpose of attention is to help us organize and use what we see and hear” (p. 2). Purdy (1997) defines “listening as the active and dynamic process of attending, perceiving, interpreting, remembering, and responding to the needs, concerns and information offered by other human beings” (p. 8). He points out that when listening to something, we need to attach meaning to the aural input because we have to consciously doing it. That is, listening does not just simply happen. Similarly, Glenn (1989) also states “listening is the process of receiving, constructing meaning from, and responding to spoken message” (as cited in Purdy, 1997). Lundsteen (1979) defines ability in listening comprehension as “generalized knowledge-acquisition skills exhibited as a consequence of hearing spoken material” (p. 81). According to Anderson and Lynch (1988), successful listeners must be thought of active ones because “understanding is not something that happens because of what a speaker says: the listener has a crucial part to play in the process, by activating various types of knowledge, and by applying what he knows to what he hears and trying to understand what the speaker means” (p. 6).

In 2002, Rost has given a very clear and comprehensive description of listening process that has taken details in the neurological, linguistic, pragmatic, and psycholinguistic views into consideration.

The briefest and broadest definition ... is: Listening = experiencing contextual effects. But this definition is not much better than any of the others. It merely defines listening as a neurological event (experiencing)

overlaying a cognitive event (creating a change in a representation).

(Rost, 2002, p. 3)

Comprehension is often considered to be the first-order goal of listening, the highest priority of the listeners, and sometimes the sole purpose of listening. Although the term 'listening comprehension' is widely used to refer to all aspects of listening, the term 'comprehension' is used in a more specific sense here. Comprehension is the process of relating language to concepts in one's memory and to references in the real world. ... 'Complete comprehension' then refers to the listener having a clear concept in memory for every referent used by the speaker... The most fundamental aspect of communication is the integration of the information conveyed by the text with information and concepts already known by the listener. Comprehension occurs as an internal model of the discourse by the listener, in which information in the text only plays one part... while attending to speech over a period of several intonation units, the listener has to store a mental representation of the discourse, and continuously update the representation with new information.

(Rost, 2002, p. 59-60)

According to the aforementioned definitions, *listening comprehension* is more than just hearing what is said. It is someone's ability to understand the meaning of the utterances he hears and relate these utterances together. Good listening comprehension would involve cognitive processes which will enable the listeners to understand a message, remember, discuss, and even retell it in their own words. In addition, prior knowledge of the topic also plays a role in aiding comprehension since it can encourage listeners to participate in the conversations actively.

Nonetheless, in my opinion, these definitions should be expanded to include higher-order or critical listening skills such as analysis and synthesis of some nonverbal features of spoken language such as the abilities to comprehend the meaning from the tone of voice, intonation patterns, facial expressions, and gestures since all these skills are very important in facilitating active listening in different day-to-day conversations.

## **2.2. What is Unique to Listening?**

Even though listening and reading skills are two receptive skills that are believed to be related, Buck (1992) has made a clear distinction between reading and listening that

“listening involves on-line processing of acoustic input, whereas reading requires processing of graphic input and allows back tracking and review” (p. 314). However, Buck (2001) and Flowerdew and Miller (2005) argue that the unique characteristics of listening comprehension stems from the differences between spoken and written language. They have summarized five major differences between spoken and written language.

First, the idea units in spoken language are shorter, and the syntactic structures are also simpler, while written language tends to be more complex and complete in terms of information and grammatical structures. For example, Flowerdew and Miller (2005) find that passive voice is used more frequently in written language, and the average word length tends to be longer as well.

Second, ideas in spoken language tend to be connected by coordinators, whereas those in written language are connected in a more complex way.

Third, natural impromptu speech tends to contain pauses, false starts, hesitations, fillers, and repetitions. These features can give the speaker some time to think and the listeners to follow along. On the other hand, in written language, overt transition markers and certain language features such as punctuations tend to be used.

Fourth, spoken language is likely to contain non-standard features such as phonological contractions, slang, and dialect, whereas the language of the written form tends to be more formal.

Fifth, spoken language can be more personal and have some emotional involvement while people talk. Speakers might express their feelings toward a particular topic with expressions such as ‘I think’ or ‘I mean.’ In contrast, these expressions were rarely seen in most academic texts (Biber, 2006).

However, Buck (2001) notes that the above differences are only the matter of degree. These features should be put on a continuum where the significant differences can vary. As Buck (2001) has reviewed the differences between spoken and written language, a question can arise. What would be the differences between reading and listening comprehension then? Flowerdew (1994) points out some features that make these two skills different and put them into two major categories.

The first distinctive feature involves the way the input has been processed. Listening comprehension involves real-time processing, whereas readers have control over the text they are reading. That is, listeners must understand the text as they listen to it (Buck, 2001; Flowerdew, 1994; Thompson, 1995). They have to be able to retain the information, integrate it, and then adjust to their understanding of what they hear to their prior knowledge (Thompson, 1995). They cannot skip over or dwell on a specific part of the

text. Also, they cannot always have the whole conversations repeated (Flowerdew, 1994).

The second difference involves the phonological features. Whereas the readers can see the word, phrase, sentence boundaries visually, the listeners must be able to recognize the phonological boundaries in a flow of speech by paying attention to pauses, false starts, hesitations, fillers, and stress and intonation patterns (Flowerdew, 1994).

These differences between listening and reading skills have practical implications for testing listening comprehension since the likelihood of including test items that are less suitable for listening than for reading can be excluded (Thompson, 1995). It is my belief that paying particular attention to these features will help us with defining the listening construct which can later on enhance the development of good listening assessments.

### **2.3. A Traditional View and an Alternative View of Listening**

Traditionally, a listener's comprehension was viewed as one's ability to remember or memorize the entire message that one received (Anderson & Lynch, 1988). The learner should be able to pay attention to what was just said to him and reproduce it. From this viewpoint, a second language learner is seen only as "a sub-standard cassette recorder that can function only in ideal listening/recording conditions" (Anderson & Lynch, 1988, p. 9). However, based on this view, the problem can arise when we want to decide whether the learner comprehends the message. The problem of viewing learners as a tape recorder is that it cannot explain or capture the features of the listening comprehension process. Oftentimes, listeners can understand the message a lot more than what they can recall the given input. Clearly, a wide range of strategies allow listeners to demonstrate their listening comprehension abilities without relying solely on memorizing the text verbatim. Accordingly, this traditional view of listening was seen as inappropriate and inadequate (Anderson & Lynch, 1988).

Anderson and Lynch (1988) reject the conceptualization of listening as a passive skill. They have therefore proposed an alternative view of listening comprehension. Based on their new point of view, the listener is seen as an active model builder because he has to "construct his own coherent interpretation of any spoken message" in order to be a successful listener (Anderson & Lynch, 1988, p. 11). They emphasize the central role of the previous knowledge or what they call "background knowledge" or "knowledge of the world" in helping the listeners understand the language, either through listening or reading. Based on this viewpoint, the essential part of listening comprehension process lies in the combination of the new information that the listener hears and his knowledge of the world and previous experience. The teachers should, therefore, encourage the students to become active listeners by asking them to make a conscious effort to hear not only the words that

the person is saying but, more importantly, to understand and recap the messages as well.

#### **2.4. Similarities and Differences between First and Second Language Listening**

If someone asks a question, “What is the difference between first and second language listening?” here is what Rost (2002) would answer, “There are many more similarities than differences” (p. 3) for example, the order of acquiring listening competency. Both L1 and L2 learners have to be able to first identify different sounds in the language and then understand how sounds blend together to create a meaningful word. Also, in order to understand the message well enough, the learners need to know the meanings of words.

However, Anderson and Lynch (1988) have discussed several differences between L1 and L2 listening in their book. They suggest that thinking about the listening process in the L1 can help language teachers understand the areas of listening that appear to be problematic for L2 listeners since it can help learners be more conscious of the listening process in the L2.

All of the children’s early language development is through oral rather than written communication, while the order is reverse for L2 adult learners (Rost, 1994). Anderson and Lynch (1988) claim that when children are listening in their first language, identifying sounds in the speech does not seem to be problematical because they usually learn to imitate sounds or distinguish two different sounds since the first few years of their language development. In addition, children mostly learn to understand the meanings of new things through situations such as ‘mealtime’, ‘going to bed’, and ‘saying goodbye to daddy’ (Rost, 1994, p. 107). This learning process is called assimilation (Rost, 1994). Most children also learn to understand the incoming messages through the process called ‘learning-by-selection’ (Rost, 2002). They can hear many things at once in the environment, but they usually only pay attention to some of them. Paying attention is often an intentional act. Oftentimes, they do not wait for the incoming speech to be complete. Even though acquiring listening ability in the first language is a gradual process, most children who have no receptive disorders eventually reach the ultimate attainment at about the same time, regardless of what our first language is (Rost, 2002).

On the other hand, acquiring listening comprehension skills can be extremely challenging for the L2 listeners, especially for the listeners at a lower proficiency level. Even though nearly everyone can achieve a high level of listening comprehension abilities in their L1, only a small percentage of people can become highly proficient in a second language due to several factors (Rost, 1994). The first factor is related to the amount of exposure of the learners to the target language. Generally, L2 learners are exposed only to a restricted set of linguistic knowledge, and the conversations around them do not always provide repeated input (Anderson & Lynch, 1988; Rost, 2002). Also, the nature of input

occurring between native/non-native speakers tends to be different from that occurring between two native speakers. Furthermore, the gaps in the knowledge of the target language cultures can be obstacles to the comprehension of L2 listeners, depending on the degree of familiarity with the foreign cultural system. Some concepts in the second language are fuzzy or unfamiliar to the learners, so they might try to use the translation strategy to help them understand (Rost, 1994). However, Rost claims that this strategy may only help them temporarily and can weaken the process of L2 acquisition. The final factor is associated with the motivation of the learners. If a student has lost his/her motivation to learn the language, the language development can slow down.

Accordingly, the teachers should help their students develop certain strategies that can help them participate successfully in conversations, for example, the ability to identify main ideas of the conversation and learning useful phrases to help them ask for clarification. Also, in a noisy environment, the listeners can be distracted or fatigued, and they can lose their listening ability in the real world; therefore, they should practice listening by consciously paying attention to the sounds and speech they hear around them.

### **3. Current Frameworks of Testing Listening Comprehension**

The relative low profile of research in listening assessments can reflect the difficulties involving the models of listening comprehension which could be particularly used to guide the field of testing (Brindley, 1998; Buck, 2001). Even though there are several models of listening comprehension, many researchers have pointed some common grounds on the nature of listening processes (Brindley, 1998; Lundsteen, 1979). The first assumption that is frequently made by educators and language testers is that listening skills usually involve different hierarchies of language processing, and the second assumption is based on the idea to move away from decoding of the sounds and words towards a much more complex and interactive model (Brindley, 1998; Lundsteen, 1979).

#### **3.1. A Listening Taxonomy**

Numerous researchers have proposed different taxonomies used in identifying listening operations. However, a number of perspectives on the process of listening comprehension is based on the studies in reading comprehension (Weir, 1993). The concern is now with testing a higher order of thinking or testing the communication as against the decoding in the word level. A high-order hierarchy here includes distinguishing facts from opinions, inferencing, evaluating the bias or prejudice of the speaker, making a critical evaluation, whereas a lower order hierarchy involves understanding of utterances only in the literal level such as following oral directions, identifying the main ideas, summarizing a story, and paraphrasing spoken message (Brindley, 1998; Lunsteen, 1979).



Since the listeners sometimes do not have a direct access to the speakers' actual intention, they have to rely on a process of inference to arrive at an accurate interpretation (Chaudron & Richards, 1986). As Kintsch (1998) has pointed out, this process is what psychologists call inferencing even though it has nothing to do with real inferences (as cited in Chaudron & Richards, 1986). According to Kintsch (1998) and Lundsteen (1979), this high-order thinking process can be part of cognitive or problem-solving processes when some conclusions have to be drawn. And the use of making an inference can include general knowledge about the world in which small pieces are retrieved.

### **3.2. Listening as an Interactive Process**

This second common ground from many previous studies provides evidence that an interactive approach can reflect the test takers' true abilities to comprehend messages in authentic contexts (Brindley, 1998; Buck, 2001; Thompson, 1995; Weir, 1993). The interactive model that seems to have the most influence on language testing practices is the one developed by Buck in 1990.

There seems to be a noticeable trend in the literature on language processing away from simple serial models, through more complex parallel interactive models. ... Interactive models, in which different knowledge sources simultaneously influence the developing interpretation. This has been supplemented by work in philosophy and linguistics, which has shown that what is communicated, is often very different from what the linguistic form of the utterance would suggest. Linguists now study how listeners extract meaning from texts by utilizing the shared knowledge of the participants to make inferences about what is meant.

(Buck, 1990, p. 57, 72)

Buck (2001) believes that listening comprehension involves the utilization of different levels of language processing. Strictly speaking, this model explains the ways listeners apply different elements of the language. The listeners have to utilize various types of knowledge in order to understand the input, and the knowledge of the language can be applied in any order, or even simultaneously.

Many studies have found that less-experienced or weaker learners tend to rely heavily on the identification of every single word in a stream of speech. Apparently, sound and word recognitions are the major sources of their difficulties (Buck 2001; Field, 2004; Gernsbacher, Varner, & Faust, 1990; Goh, 2000). They pay too much attention at the word

level, trying to understand every word they have heard. This could occupy much of their memory. When their memory capacity has been exceeded, it could prevent them from moving onto the higher-level interpretation.

In summary, according to the first view of listening skills, the process of developing listening comprehension involves both lower-order and higher-order thinking skills. However, in order to become competent listeners, they have to be able to employ their prior background knowledge or experience to understand the oral input. Based on different perspectives on listening comprehension, teachers can use them as the points for consideration when designing listening assessments and creating classroom materials to help students develop their listening skills.

## **4. Listening Assessments**

Rost (1994) strongly argues that measuring the development of listening comprehension is not an easy task because the test developers need to have a comprehensive understanding of how listening ability develops and what they mean by improving listening abilities. He always mentions that listening involves psychological skills, such as recognizing words and processing the discourse in terms of cohesion, logic, and relevant underlying schemas. It also involves the speaker's intentions and numerous social skills such as back-channeling signals and making repairs when misunderstandings occur. Accordingly, if language testers would like to describe the development of listening ability, they need to consider all the following factors and be able to explain them.

### **4.1. Test Purposes**

Buck (2001) says that when a test is designed, the test developers normally create it for a particular purpose and for a specific group of examinees. And the types of listening tests can vary depending on the purpose of the test. There are a number of common test purposes such as language placement, language achievement, and diagnostic testing. For example, in placement tests, the main aim is to discover what the learners know about the language. Therefore, the test items should require the test takers to demonstrate overall listening abilities (Flowerdew & Miller, 2005). For achievement and diagnostic tests, the test should be designed to measure what the students know after a course of instruction (Flowerdew & Miller, 2005). Once the purpose of the test has been decided upon, the test developers can begin the item writing process.

## 4.2. Approaches to Assessing Listening

Generally speaking, there are three major approaches in assessing listening. They are the discrete-point approach, the integrative approach, and the communicative approach. Buck (2001) and Flowerdew and Miller (2005) mention that these three historical developments correspond to the theories of pedagogical practices over the past 60 years.

### 4.2.1. The Discrete-Point Approach

The discrete-point approach is based on the audio-lingual teaching method which considers the language learning process as a habit formation. Flowerdew and Miller (2005) mention that this approach prevailed when the concepts of structuralism and behaviorism were the dominant paradigms in pedagogical practices. In the discrete-point approach, the components of the language have to be measured separately. Segmental phonemes, grammatical structures, and lexis are treated as separate entities, and each particular segment is measured one at a time (Farhady, 1983; Flowerdew & Miller, 2005). “For Lado (1964), testing listening comprehension means testing the ability to recognize elements of the language in their oral form” (as cited in Buck, 2001, p. 63). After this approach had been introduced, it soon became the target of attacks from many scholars in the field because “the primary function of language is communication” (Farhady, 1983).

The common formats of discrete-point tests are true/false items and some four-option multiple-choice questions. The most common techniques in testing listening comprehension would be phonemic discrimination tasks, paraphrase recognition, and response evaluation (Flowerdew & Miller, 2005).

In sound discrimination tasks, the test takers will hear minimal pairs in different words, and they are asked to identify the correct sounds. Normally, they are usually not given any clue to the correct interpretation.

#### **Example 1:**

Choose the words with the /p/ sound in them:

Pen    Ben

Ball   Paul

Bat    Pat

(Flowerdew & Miller, 2005, p. 199)

For paraphrase recognition, the test takers have to paraphrase what they hear, and the contexts might be given.

**Example 2:**

Test takers hear:

Mary asked her mother for some money to go to the cinema.

After hearing the statement, they will read and choose one answer from the three options:

- a) Mary wanted money to buy some new clothes.
- b) Mary wanted to see a movie so she asked for some money.
- c) Mary asked her mother to go with her to the cinema.

(Flowerdew & Miller, 2005, p. 199)

For response evaluation, the test takers have to give an answer to the questions they hear.

**Example 3:**

Test takers hear:

Would you like to go shopping?

After hearing the statement, they will read and choose one answer from the three options:

- a) Yes, okay.
- b) Yes, I have.
- c) It cost \$200.

(Flowerdew & Miller, 2005, p. 199)

When compared with other assessment approaches, this type of assessment is relatively easy to grade (Flowerdew & Miller, 2005). The graders can easily mark the students' responses by assigning the numbers to each item and sum the total correct responses. Nonetheless, Buck (2001) argues that the discrete-point approach can be "somewhat unnatural task because, in most listening situations, listeners do not just use the phonetic information to determine what was said, they also use the context" (p. 63). This was the major reason why the integrative approach had been proposed in the early 1970s.

### 4.2.2. Integrative Approach

The arguments against the discrete-point approach were mainly led by John Oller in 1983. He argued that the integrative approach could be more effective as it attempts to assess the students' capacity to use different elements of the language at the same time (Buck, 2001). Douglas (1989) also called for an assessment that is integrative and integrated since it can challenge the test takers to deal with a variety of listening skills. From this perspective of language use, the language processing becomes the main focus of assessments since the listeners must process the input simultaneously in order to understand the literal meanings of what was said (Flowerdew & Miller, 2005). Common testing techniques include gap-filling exercises (or cloze tasks), dictation, statement evaluation, and translation (Flowerdew & Miller, 2005; Farhady, 1983).

For gap-filling exercises, the test takers are assumed to be able to comprehend the message by filling the missing part of language based on their knowledge of the target language. One criterion of this measurement activity is the students' language ability to make predictions. For gap-filling tests, the test writers can either remove certain words (often content words) or select every 5<sup>th</sup> or 10<sup>th</sup> words to be removed from the text.

Next, it is claimed that dictation tests measure more than spelling or word recognition (Flowerdew & Miller, 2005). The technique is seen as integrative because the test takers have to be able to synthesize the incoming speech by recognizing the phonological, syntactic, and semantic components of the language. They will listen to a passage more than once and then they have to write down what they hear. This type of test can be useful for low-proficiency learners whose writing skills might not be high enough (Flowerdew & Miller, 2005).

For the statement-evaluation technique, the test takers have to evaluate the truth of the sentence they hear.

#### **Example 5:**

Test-takers hear:

When you go to the beach on a hot day, you should drink plenty of water.

Then the test-takers have to agree or disagree with the statement by evaluating the truth of the sentence.

(Flowerdew & Miller, 2005, p. 200)

For the translation technique, Flowerdew and Miller (2005) say that it is quite similar to a dictation test since the test takers are asked to listen to a text and write what they have heard. The only difference between the translation and dictation techniques is in the fact that the test takers have to write their answer in their L1 in a translation task, whereas they have to write in the L2 in a dictation task.

By using the integrative approach, the test takers must be able to process the language relatively simultaneously and demonstrate the ability to comprehend the literal meaning of what is said (Flowerdew & Miller, 2005). However, the main criticism of the integrative approach is that there is little or no focus on the link between the linguistic competence and the contexts of language use (Buck, 2001; Flowerdew & Miller, 2005). In addition, the integrative tests tend to involve a large amount of scoring time; accordingly, they are rarely used in large-scale tests (Farhady, 1983).

#### **4.2.3. Communicative Approach**

Responding to the trend towards the communicative approach in language teaching, the communicative approach takes a wide range of contexts into consideration when assessing listening comprehension (Flowerdew & Miller, 2005). Based on this, it is suggested that the tests should be authentic, so they can predict the real-life listening abilities of the learners (Buck, 2001; Flowerdew & Miller, 2005; Hymes, 1972; Thomson, 1995). That is, the texts must be taken from authentic sources or contain language features in real-life conversation. An example of communicative task would be that the test takers are asked to listen to news on the radio and summarize the story. For example, the test takers are given a task to fill in the missing information while they have to listen to an announcement at the Tourist Information Office about the schedule of day tours (Flowerdew & Miller, 2005, p. 200). Even though the communicative approach seems to fit the current pedagogical practices most, Flowerdew and Miller (2005) have argued about the impracticality of this approach that even though a test taker can perform well in one test situation, it does not necessarily mean that he/she will be able to perform well in other communicative situations. Furthermore, communicative tests are more difficult to prepare and to evaluate when compared with discrete-point or integrative tests. A number of different test items are absolutely required for the learner to demonstrate different aspects of listening abilities. Despite the impracticality of the communicative approach, it has been an influential approach in listening assessments for more than 30 years.

### **5. Considerations in Selecting Suitable Texts for a Listening Test**

When developing tests of listening comprehension, the test developers should consider different factors that can affect the comprehensibility of the test takers. This is

mainly because, unlike the readers, “the listeners cannot review and reevaluate the information that has been presented to them” (Thompson, 1995, p. 35). A number of factors that may affect the difficulty of listening assessments have been identified by many researchers (Buck, 2001; Brindley & Slatyer, 2002; Thompson, 1995). Based on the categorization mentioned in Brindley and Slatyer’s study (2002), I have categorized the factors into 3 aspects: the nature of the input, the nature of the assessment tasks, and the differences in individual listeners.

## **5.1. The Nature of Input**

The nature of input can include the text type whether it is a listening or reading text, speech rate, text difficulty, test length, and number of hearings.

### **5.1.1. Text Type**

There is enough evidence indicating that the degree of ‘orality’ of the listening text can affect the test takers’ listening scores (Brindley & Slatyer, 2002). For example, a study of Chaudron and Richards in 1986 shows that texts are much easier to understand if they include conversational discourse signals (e.g., *well*, *so*, and *now*), discourse markers (e.g., *because*, *at that time*, and *of course*), redundancies in the form of repeated nouns, elaboration in the form of synonyms and paraphrase, and frequent use of transitions and prepositions. This is because these discourse markers help the listeners organize the major ideas and understand sophisticated contents.

In addition, a study of Shohamy and Inbra (1991) has demonstrated that the texts that are like the oral end of the continuum yield a higher score than the ones that are closer to the written end. The students in their study have reported the greater difficulty when they listened to the news broadcasts (prewritten edited monologues) than when they listened to lectures even though both texts contained similar factual information.

Accordingly, Thompson (1995) has suggested the test developers to avoid using the written materials for listening comprehension assessments since it is rather difficult to modify a written text to sound like a real spoken one, and many studies have already shown that spoken texts differ from reading texts in a number of features. For instance, as Thompson (1995) has pointed out, unlike the reading text which is usually densely packed with information, idea units in the spoken language are typically presented in shorter clauses, and they are loosely connected together. In addition, the spoken language oftentimes contains repetitions and bounded by pauses or hesitations because the speakers do not usually have time to plan their speech.

### **5.1.2. Speech Rate**

It is not surprising to find that numerous studies have shown the effect of speech rate on listening comprehension (Brindley & Slatyer, 2002; Griffith, 1990; Hayati (2010); Thompson, 1995). For example, in 1990, Griffith conducted a study on the effects of three speech rates--200 words per minute, 150 wpm, and 100 wpm--on listening comprehension of NNS adult learners. He found that faster speed rates could significantly reduce the comprehension of the listeners, particularly those from low-intermediate levels. Similarly, a study of Hayati (2010) examined the effect of speech rate on listening comprehension of Iranian EFL learners. He also found that slower speech rate could improve EFL learners listening comprehension. However, both studies have indicated that natural speech rate could demonstrate greatest improvements than slow speech rate in NNS learners' listening comprehension.

### **5.1.3. Text Difficulty**

Establishing the level of text difficulty can be very crucial for the teachers and test developers who wish to select the appropriate materials for the students (Fulcher, 1997). However, finding an authentic text at the desired level of difficulty can be difficult (Thompson, 1995) since the test takers can be at a variety of ability levels. Clearly, the writers of texts for various audiences would need to consider many factors, including students' interest, background knowledge, inferring ability, and language experience into consideration (Brown, 1994; Thompson, 1995). They may also need some guidance from the teachers or experts in the field in order to make a decision of what would make the texts more accessible to the test takers (Fulcher, 1997). However, I believe that all the decisions that have been made could be very much based on intuition of people who are involved with language teaching and test development.

### **5.1.4. Test Length**

While the test takers are listening to a conversation or someone talking, they must comprehend as they listening to it (Buck, 2001; Flowerdew, 1994; Thompson, 1995). They basically have to be able to retain the incoming information and incorporate it with their prior knowledge in order to interpret the incoming information (Thompson, 1995). Given this as a heavy processing load of information, people recall less information from what they hear than from what they read; therefore, the heavy load of listening information can cause the test takers to lose concentration rather quickly (Brown, 1994; Thompson, 1995). From a study conducted by the ETS during the field testing of ETS Advanced Russian Listening/Reading test (1986), its results show that the test takers did poorly in the 50-



minute listening sub-section than the 50-minute reading portion, and they have reported the great difficulties in maintaining their concentration during the listening period than the equivalent period of reading (as cited in Thompson, 1995). Therefore, the length of a test can vary depending on the proficiency levels of the test takers.

### **5.1.5. Number of Hearings**

As Skehan (1998) has suggested, allowing only one hearing may cause too much cognitive load on the listeners by creating greater demands of language processing (as cited in Brindley & Slatyer, 2002). Accordingly, many tests allow the test takers to hear the recording more than one. For example, the revised Cambridge First Certificate in English (FCE) of the Cambridge examinations for the lower levels allows two hearings (Brindley & Slatyer, 2002), and the test developer is now no longer free to change this because it has become a standard practice that the test takers would expect when they take this test (Buck, 2001).

## **5.2. The Nature of Assessment Tasks**

The characteristics of assessment tasks that have been found to have an effect on the listening test difficulty can include the amount of context provided in the spoken texts, clarity of task, and test formats (Brindley & Slatyer, 2002). However, the effects of test format will be explained in the next section.

Generally speaking, the more demanding a text is in terms of the amount of context given, the more difficult it would be. How much context should be given to the test takers depends largely on students' proficiency levels. Individual teachers and test developers will need to make their own judgments according to the teaching and learning context they find themselves in. For example, a study of Alderson (1983) shows that changing the deletion rate in the text of a cloze listening test can result in the different levels of difficulty. The more contexts the test takers have, the more they can reconstruct textual coherence and restore the missing words (Klein-Braley, 1983).

Moreover, the instructions of a test task should give a clear indication of what the test takers have to do. No examinees should misinterpret the task because this can pose a threat to reliability and construct validity. One way to avoid this is to provide some test examples of the tests for the test takers to try before the test. Moreover, the test should be carefully proofread and free of all errors and typos. During the test administration, it is always wise for the test administrators to explain to the test takers in the mother tongue if necessary. Also, the test writers should avoid trick questions or anything that is designed to mislead.

### **5.3. Individual Differences**

The degree of learning or comprehending the new information from listening alone can largely vary from individual to individual. Listening comprehension does not only require acuity of hearing, but it also involves attention span, memory storage including retention of mental images or pictures, abilities in recognizing linguistic features, levels of vocabulary knowledge, the organization and retrieval of facts, and the understanding of the speaker's purpose (Lundsteen, 1979). Moreover, entry of new information into long-term can be limited for some people because some people might take a longer time to activate necessary memory links to retrieve a specific piece of information while they listen (Rost, 1994). This depends on individuals' listening techniques, years of language study, and the opportunity in exposing to the target language. Accordingly, L2 listeners can have a wide range of performance in listening skills. Therefore, I believe that teachers should be aware of students' individual differences and be able to create a variety of classroom materials and activities that can potentially help different students develop their listening skills.

## **6. Formats of Listening Assessments**

Some researchers have found that different formats of listening tests can affect the performance of the students. For example, from an extensive literature review of In'nami and Koizumi (2009), they have found that the subjects in many studies were likely to receive higher scores on multiple-choice questions than on open-ended or cloze tasks. However, to what degree that a multiple-choice test seems to be easier, the answer still depends on different variables related to the test itself and individual listeners.

### **6.1. Multiple Choice Questions and Answers**

The most widely-used assessment technique in many standardized tests of ESL such as the TOEFL, IELTS, and TOEIC is in the multiple-choice format (Brindley, 1998; Henning, Gary, & Gary, 1983; Thompson, 1995). Typically, the test format involves listening to a text and then choosing the most appropriate response to what has been heard (Henning et al., 1983).

Brindley (1998) and Thompson (1995) have identified several advantages of using multiple-choice questions. First of all, a multiple choice test can offer the ease of scoring. It is fast and reliable because no judgment on the part of scorers is required. Secondly, the multiple-choice items generally require a minimal amount of time to complete. When more test items can be included in the test, the reliability of the test can be enhanced. Next, the test developers might be able to control some confounding variables more effectively since

the language production is not the requirement. Finally, multiple-choice tests can be more practical in situations where there is a large number of test takers taking the test at the same time.

However, multiple-choice tests can have several disadvantages (Brindley, 1998; Thompson, 1995; Yian, 1998). The most controversial issue is that multiple-choice tests are too open to students' guessing, especially the tests with three options. Second, good multiple-choice items with effective distractors are often time-consuming and difficult to write. Third, answering the multiple-choice questions does not resemble normal language use.

Another consideration related to the use of multiple-choice format is whether the test takers are allowed to see the questions before they listen to the script. In Yian's study in 1998, she found that allowing the test takers to view the questions can help them anticipate the incoming input. This can make the test less difficult for them. In some cases, the test takers have to listen to the entire script without seeing the comprehension questions in advance. The differences in the format decisions can increase the difficulty relatively quickly because the demands of listening processes and memory capacity are different.

Despite all advantages and disadvantages that have been mentioned, some test developers have argued that "it is possible to write the items that can test the meaningful use of language in context" (Brindley, 1998, p. 178). For example, the revised Cambridge First Certificate in English (FCE) that is developed by the University of Cambridge Local Examinations Syndicate (UCLES) includes item types that can test a wide range of communicative skills because most of the test questions can be found in day-to-day conversations (Please see <http://www.cambridgeenglish.org/exams-and-qualifications/advanced>).

## 6.2. Listening Recalls

Henning et al. (1983) argue that while the multiple-choice technique is easy to administer and grade, what they have found in one of their studies is that such tests are too difficult and do not have enough discriminating power in low proficiency learners as the students' scores have shown to be quite random. Based on their argument in the ineffectiveness of multiple-choice tests, they have proposed a new test technique called *Listening Recall*. It can be seen as a type of listening cloze tests. However, unlike the cloze tests, the deletion of words listening recall tests is not random. The excerpt below is an example of the listening recall test that the researchers used to conduct a study on a group of EFL students in Egypt.

**Directions:** In this passage you will fill in the missing words in the blank spaces after the passage is read to you. First you will have a chance to read the passage. Look at the passage now.

The Porters live in Dartmouth, a \_\_\_\_\_ (small) English town at the \_\_\_\_\_ (mouth) of the River Dart. Mr. Porter is a lecturer in (engineering) at the local \_\_\_\_\_ (college). Mrs. Porter is a \_\_\_\_\_ (nurse) and work in a \_\_\_\_\_ (nearby) hospital. The Porters have their \_\_\_\_\_ (friends), the Saids visiting them from Egypt. They've been there a \_\_\_\_\_ (week) now. ...

(Henning et al., 1983)

From the given text, some content words were purposely deleted in the way that the test takers would not be able to guess based on the context alone. They have to comprehend the total passage in order to be able to fill in the missing words. Based on the findings, the listening recall test appears to be highly correlated with the traditional multiple-choice listening comprehension test. Also the result from a statistical analysis indicates that the listening recall task is capable of discriminating students of a wide range of proficiency than the traditional multiple choice task. Thus, the authors argue that the listening recall test can be more suitable for evaluating students with a wide range of language abilities, particular those who are from the low proficiency levels.

### 6.3. Cloze Tests

The first and general definition of the cloze test is “the systematic deletion of words from text” (Alderson, 1983). Typically, the missing parts are single words, and the cloze procedure is normally done systematically. For instance, a test writer might decide to remove every 7<sup>th</sup> word in the text (Alderson, 1983). It has been claimed that the cloze test measures students' comprehension abilities by giving them a text with several blanks and asking them to fill in those blanks.

Typically, the listening cloze test task requires students to build an internal representation after they have listened to an aural text. They have to fill in the words in a meaningful way (Alderson, 1983). Some researchers mention that this technique has been originally used as measure of reading comprehension, and it is now used in some American universities as a placement examination for incoming international students because it is proved to have a high reliability and construct validity in the field of aural testing when compared with standardized multiple choice tests (De Jong, 1984; Templeton, 1977). As De Jong has pointed out, the test takers “are required to have ability to understand

the foreign language at the level of native speakers of comparable age and educational background" (De Jong, 1984, p.99).

Many researchers have mentioned several advantages of this technique. Developing a cloze test allows flexibility in the texts and in the topics chose. It is more time-efficient when compared with the multiple-choice technique because a large number of items can be developed, the test can be given to a large group of examinees at the same time, and many problems involving the item writing process can be avoided (Brindley, 1998; Templeton, 1977). Despite numerous advantages, Templeton (1977) has mentioned a few issues that are related to the use of cloze test. For example, the scoring of cloze tests can be carried out in many ways. In some cases, only exact word will be scored, or students can receive a partial credit score if they use synonyms or semantically acceptable words. This can affect the validity and reliability of the test (Alderson, 1983). In addition, Alderson argues that the cloze test can be very sensitive to the deletion of individual words. That is, changing the deletion frequency of a test can create a different level of test difficulty since the amount of context given to the test takers will vary. Currently, it is still controversial whether the cloze test is measuring lower-order or higher-order skills. Some researchers believe that a cloze test can test global knowledge which is considered a higher-order skill (Oller, 1972, 1973; Oller & Inal, 1971; Oller & Conrad, 1971), whereas others claim that the cloze test is only testing a short-term memory and measuring the skills in sentence level, which is clearly not a higher-order skill (Alderson, 1983).

However, the most important issue related to the cloze test would be the fact that most findings on the cloze procedure still involved only small-scale studies; therefore, I think it is very difficult to make a generalization of the results. Absolutely, this technique requires careful pretesting (Brindley, 1998), and further investigations (De Jong, 1984).

#### **6.4. Short Answer Questions**

The short answer questions or open-ended questions allow the test developers to determine whether the students have actually understood the text (Brindley, 1998). Open-ended questions are usually preferred to multiple-choice tests by many test developers because it can avoid some of the issues related to multiple-choice tests (Thompson, 1995). Above all, it can eliminate guessing. Second of all, it allows the test writers to ask any questions, in which the answers are not limited to four multiple-choice options. However, since more than one answer is plausible, a careful, detailed scoring key listing all acceptable responses needs to be constructed, and a comprehensive set of possible responses can be added up as the scoring procedure proceeds (Brindley, 1998). Accordingly, the answers should be kept very short and precise, and thus should not

depend too heavily on the test takers' writing abilities (Brindley, 1998).

### **6.5. True-False**

The true-false format is widely used in a number of language tests such as the Cambridge Certificate of Proficiency in English (Brindley, 1998). The most obvious drawback of this test format is that the test takers have a fifty percent chance of getting the correct answer without truly knowing the correct answers. However, this issue can be addressed by including the third option "No information available" in the answer. Another disadvantage of true-false items comes from the fact that the test writer usually bases an item only on one single idea, and the test takers can easily find the answer by paying attention to that particular detail.

### **6.6. Dictations**

Dictation is also one of the common measures of listening comprehension because it is frequently reported to be a valid and reliable method of listening comprehension test (Farhady, 1983). Dictation is typically involved the procedure where the sentences are read independently of one another (Farhady, 1983). He mentions that dictation can be a good measure of language skill when compared with the discrete-point tests since the abilities to create interrelationship among the sentences are somewhat similar to natural language behavior. However, he argues that dictation might not be suitable when used as a testing device since the scoring procedure can be inaccurate and time-consuming. Accordingly, the scores can be unreliable.

Moreover, even though many researchers suggest that dictation can be used to measure listening ability, their studies have shown that dictation clearly involves skills other than just listening (Oller & Inal, 1971; Rahimi, 2008). This includes grammatical knowledge, spelling, auditory memory, and lexical knowledge. In the worst case, the test takers might be tested on their ability to "speed write" (Farhady, 1983). This problem can influence the validity of the dictation test since it is not clear whether the test is assessing listening comprehension or the speed with which a message is written (Farhady, 1983). Therefore, it is recommended not to use dictation as a substitute for a listening test (Brindley, 1998).

In summary, there are different kinds of test formats that can assess listening abilities and it is oftentimes difficult to select a suitable test format. Therefore, it would be best if the teachers can select the task based on classroom learning objectives, levels of students' language proficiency, and the contexts of assessment.

## **7. Issues and Challenges in the Assessments of Listening**

### **7.1. Confounding of Skills**

In the field of listening assessment, it is still controversial whether listening is considered one unitary trait, or it involves several language sub-skills. A perusal of the literature reveals that there is no general agreement on what listening comprehension actually involves. Some studies show that there are a number of different types of knowledge involved in the process of listening comprehension, including both linguistic and non-linguistic knowledge (Buck, 2001; Lundsteen, 1979; Oller, 1983). Among these skills are phonology, lexis, syntax, semantics, and discourse structure (Lundsteen, 1979).

Accordingly, trying to define the listening construct based on language competence can pose several problems since it is very difficult “to come up with a suitable description for listening competence” (Buck, 2001, p. 102). Currently, there have been a number of models of communicative competence, for example Canale and Swain (1980), Bachman (1990), and Bachman and Palmer (1996). Even though these models complement one another in terms of their communicative goals, the areas of emphasis can be different. Weir (1993) has emphasized the importance of the decisions on conditions and contexts that have to be included in listening assessments. The test developers must first “identify the situational features and important facets of interaction involved in the desired performance” (Weir, 1993, p. 98) in order to make sure of what their test is measuring and what their scores mean. This can in turn help the test developers make decisions whether their tests and scores are valid and fair to the test takers.

### **7.2. Dealing with Authenticity**

The most basic definition of authentic materials refers to any materials that are “produced by native speakers for native speakers” (Chavez, 1998). Back in 1985, Breen (1985) subdivided authenticity of language into four types: authenticity of the texts used as input data for learners, authenticity of the learners’ own interpretation of such texts, authenticity of tasks conducive to language learning, and authenticity of situation that can reflect the actual social situation of the language classroom.

Breen (1985) characterizes authentic texts as the materials for genuine communication. As the first step, the teachers need to consider the degrees of authenticity of the materials they use in their classrooms and ensure that the materials are acceptable and appropriate. Those materials should not cause our learners to develop patterns of interaction that would mark them as deviant or simply odd. According to Breen (1985), by using authentic texts, learners can be exposed to direct contact with the input data, and then “these texts may serve as the *means* through which the learners can gradually uncover the

conventions which underlie the use of the target language" (p. 63). However, someone might ask that if any text that has been taken out of its original context has been altered or adjusted to the level of its intended audience, would the text become inauthentic? Chavez (1998) responds to this question that the texts should at least represent pedagogical intent. For example, he gave one example of a situation when learners may find movie show-time schedules from a foreign city and printed in a textbook inauthentic because they would not consider going to see a movie playing thousands of miles away. This can lead to another issue of authenticity which relates to the learners themselves.

The second issue of authenticity relates to the question "For whom it is authentic?" (Breen, 1985). In this case, he argues that it is the consideration of how the learners go about interpreting the texts in their own particular way. They might interpret those texts based on their own state of prior knowledge and experiences.

Using real-life situations or tasks is another facet of the authenticity. Traditionally, many teachers raised the questions whether the tasks led the students to believe that the teachers had imported reality into their classrooms. Perhaps, the most authentic tasks or situations are those which required the learners to participate in a real communication (Breen, 1985).

However, Buck (1990) has raised the concern about the state of uncertainty in establishing the relationships among skill levels necessary for real-life communication. In most contemporary approaches in listening assessments, the spoken texts can no longer be broken down into smaller elements of the language due to the issue of applicability. In terms of the test tasks, even though testing global understanding of the message has been considered the center of listening assessments, both Buck (1990) and Brindley (1998) have raised the issues associated with the integrated testing approach that lie in the application of the approach and the authenticity of the listening materials. The teachers, test developers, and raters would absolutely need a special training when adopting this approach.

### **7.3. Assessing Higher-Level Skills**

Traditionally, the importance was first placed on the test takers' ability to discriminate sounds and ability to recognize stress and intonation patterns (bottom-up processing). The students' scores on this type of discrete tests are claimed to be equated with proficiency in listening comprehension, and the General Council's PLAB test could be an example of this type of test that is still in use (Weir, 1993). However, many teachers argue that the ability to discriminate sounds no longer reflect the true ability of the learners to comprehend the messages. Instead, the most skilled or most successful listeners are those who "consistently, in the least time and in the greatest variety of circumstances, most



closely comprehend the speaker's meaning in the widest variety of spoken material ... and use the widest variety of thinking processes ... highly complex conscious, creative, critical, and problem-solving skills" (Lundsteen, 1979). Accordingly, the interest has been shifted to the top-down listening assessments that have incorporated all levels of language elements that can facilitate the language comprehension (Weir, 1993).

## 8. Conclusion

I hope that this review will help language teachers and test developers find the way of designing appropriate listening tasks by being aware of challenges and issues related to each approach of listening assessment. Based on the notion that it is impossible to measure everything without defining what it is as an investigation into the nature of listening comprehension, they need to consider what to be tested and how to test it. However, when compared with the number of research studies in reading comprehension, studies in listening during the 1950s were 8,400 studies behind those in reading (Chaudron & Richards, 1986; Flowerdew & Miller, 2005; Osada, 2004). Therefore, there is obviously a need for more research into second language listening comprehension and ways of testing it. This present research is intended to address this challenge and hopefully, it can shed some light on the keys and ways to teaching and assessing communicative listening ability.

## References

- Alderson, C. J. (1983). The cloze procedure and proficiency in English as a foreign language. In J. Oller (Ed.), *Issues in Language Testing Research* (pp. 205-217). Massachusetts: Newbury House.
- Anderson, A., & Lynch, T. (1988). *Listening*. Oxford: Oxford University Press.
- Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (1996). *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford: Oxford University Press.
- Biber, D. (2006). *University Language: A corpus-based study of spoken and written registers* (Vol. 23). John Benjamins Publishing.
- Breen, M. (1985). Authenticity in the language classroom. *Applied Linguistics*, 1, 60-70.
- Brindley, G. (1998). Assessing listening abilities. *Annual Review of Applied Linguistics*, 18, 171- 191.
- Brindley, G., Slatyer, H. (2002). Exploring task difficulty in ESL listening assessment.

- Language Testing*, 19 (4), 369-394.
- Brown, G. (1994). Dimensions of difficulty in listening comprehension. In D.J. Mendelson & J. Rubin (Eds.). *A Guide for the Teaching of Second Language Learning* (pp. 31-56). San Diego: Dominie Press.
- Buck, G. (1990). *The Testing of Second Language Listening Comprehension*. Unpublished PhD Thesis, University of Lancaster.
- Buck, G. (1992). Listening comprehension: Construct validity and trait characteristics. *Language Learning*, 42 (3), 313-357.
- Buck, G. (2001). *Assessing Listening*. Cambridge: Cambridge University Press.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1-47.
- Chaudron, C., & Richards, J. C. (1986). The effect of discourse markers on the comprehension of lectures. *Applied Linguistics*, 7 (2), 113-127.
- Chavez M. T. (1998). Learner's perspectives on authenticity. *IRAL*, 36 (4), 277-306.
- De Jong, J. H. A. L. (1984). Testing foreign language listening comprehension. *Language Testing*, 1, 97-100.
- Douglas, D. (1989). Testing listening comprehension in the context of the ACTFL proficiency Guidelines. *Applied Language Learning*, 1 (1), 53-73. Educational Testing Service. *Advanced Russian Listening/Reading Proficiency*. Princeton, ETS, 1986.
- Farhardy, H. (1983). New directions for ESL proficiency testing. In J. Oller (Ed.), *Issues in Language Testing Research* (pp. 253-269). Massachusetts: Newbury House.
- Field, J. (2004). An insight into listeners' problems: too much bottom-up or too much top down?. *System*, 32(3), 363-377.
- Flowerdew, J. (1994). *Academic Listening: Research Perspectives*. Cambridge: Cambridge University Press.
- Flowerdew, J., & Miller, L. (2005). *Second Language Listening: Theory and Practice*. Cambridge: Cambridge University Press.
- Fulcher, G. (1997). Text difficulty and accessibility: Reading formulae and expert judgement. *System*, 25(4), 497-513.
- Goh, C. C. (2000). A cognitive perspective on language learners' listening comprehension = problems. *System*, 28(1), 55-75.
- Gernsbacher, M. A., Varner, K. R., & Faust, M. E. (1990). Investigating differences in general

comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(3), 430.

Glenn, E. C. (1989). A content analysis of fifty definitions of listening. *International Journal of Listening*, 3(1), 21-31.

Graham, S. (2006). Listening comprehension. *System*, 42, 165-182.

Griffith, R. (1990). Speech rate and nonnative speaker comprehension: A preliminary study in the time-benefit analysis. *Language Learning*, 40(3), 311-336.

Hayati, A. (2010). The Effect of Speech Rate on Listening Comprehension of EFL learners. *Creative Education*, 2, 107-114.

Henning, G., Gary, H., & Gary, J. (1983). Listening recall-Listening comprehension test for low proficiency learners. *System*, 11(3), 287-293.

Hymes, D. (1972). On communicative competence: Extracts. In J. B. Pride and J. Holmes (Eds.), *Sociolinguistics* (pp. 269-293). Harmondsworth: Penguin.

In'nami, Y., & Koizumi, R. (2009). A meta-analysis of test format effects on reading and listening test performance: Focus on multiple-choice and open-ended formats. *Language Testing*, 26(2), 219-244.

Kintsch, W. (1998). *Comprehension: A paradigm for Cognition*. Cambridge: Cambridge University Press.

Klein-Braley, C. (1983). A cloze is a cloze is a question. In J. Oller (Ed.), *Issues in Language Testing Research* (pp. 218-227). Massachusetts: Newbury House.

Lado, R. (1964). *Language Teaching, A Scientific Approach*. ERIC.

Lundsteen, S. W. (1979). *Listening: Its Impact at all Levels on Reading and other Language Arts*. Urbana: the ERIC Clearinghouse.

Oller Jr, J. W., & Inal, N. (1971). A cloze test of English prepositions. *TESOL Quarterly*, 5(4), 315-326.

Oller, J. W., & Conrad, C. A. (1971). The cloze technique and ESL proficiency. *Language Learning*, 21(2), 183-194.

Oller, J. W. (1972). Scoring methods and difficulty levels for cloze tests of proficiency in English as a second language. *The Modern Language Journal*, 56(3), 151-158.

Oller, J. W. (1973). Cloze tests of second language proficiency and what they measure. *Language Learning*, 23(1), 105-118.

- Oller, J. W. (1983). *Issues in Language Testing Research*. MA: Newbury House Publishers.
- Osada, N. (2004). Listening comprehension research: A brief review of the past thirty years. *Dialogue*, 3, 53-56.
- Purdy, M. (1997). What is listening? In M. Purdy & D. Borisoff (Eds.), *Listening in Everyday Life: a Personal and Professional Approach* (pp. 1-20). Maryland: University Press of America.
- Rahimi, M. (2008). Using dictation to improve language proficiency. *Asian EFL Journal*, 10(1), 33-47.
- Rost, M. (1994). *Introducing Listening*. London: Penguin.
- Rost, M. (2002). *Teaching and Researching Listening*. London: Longman.
- Shohamy, E., & Inbar, O. (1991). Validation of Listening Comprehension Tests: The Effect of Text and Question Type. *Language Testing*, 8, 23-40.
- Skehan, P. (1998). *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press.
- Templeton, H. (1977). A new technique for measuring listening comprehension. *ELT Journal*, 31(4), 292-299.
- Thompson, I. (1995). Assessment of second/foreign language listening comprehension. In D.J. Mendelson & J. Rubin (Eds.), *A Guide for the Teaching of Second Language Learning* (pp. 31-56). San Diego: Dominie Press.
- Weir, C. J. (1993). *Understanding and Developing Language Tests*. New York: Prentice Hall.
- Yian, Wu. (1998). What do tests of listening comprehension test? – A retrospection study of EFL test-takers performing a multiple-choice task. *Language Testing*, 15 (1), 21-44.