The Effect of Four Formats of Multiple-choice Questions on the Listening Comprehension of EFL Learners
Fatemeh Hemmati, Esmaeil Ghaderi

Abstract
This study aimed at investigating the effect of four formats of multiple-choice questions (MCQs) on the listening comprehension of EFL learners. To achieve this goal, 60 homogeneous intermediate EFL learners were selected as the sample of the study. They were divided into four groups of fifteen. Each group received one format of the same listening test. The findings indicated that the full question preview format (FQP), answer option preview (AOP), and question stem preview (QSP) can have a facilitative effect on the listening comprehension of the EFL learners in comparison with the not-previewing format (NP); however, no significant difference was found between the three question previewing formats themselves (FQP, AOP, and QSP). These findings can be helpful for test developers, EFL teachers, and material designers.

© 2014 The Authors. Published by Elsevier Ltd. Open access under CC BY-NC-ND license.

Keywords: multiple-choice; listening comprehension; format

1. Introduction
Nunan (1997) has called listening a “Cinderella skill” in second language teaching and learning. He gave this label to this skill because it was neglected by teachers and learners. Listening was taken a back seat to the teaching of reading, speaking, writing, and grammar. Teachers either abandoned teaching listening or did so rather poorly. However, in the 1960s, the emphasis on oral language skills gave it a boost. It became popular again in the 1980s, when Krashen's (1982) ideas about comprehensible input obtained a prominent role. After a while, it was reinvigorated by Asher's (1988) Total Physical Response method which was based on the notion that learning a second language is similar to learning first language and L2 is learned most effectively in the early stages if the pressure for production is reduced on the part of the learners. Further pedagogical research refined the process of listening more. Rubin (1994) listed text, interlocutor, task, listener, and process characteristics as different contextual characteristics which influence the speed and efficiency of processing aural language.

*Corresponding author: Tel: +98 9141591696
Email address: ghaderi43@gmail.com
Richards (1985) states that, “current understanding of the nature of listening comprehension based on research in psycholinguistics, semantics, pragmatics, discourse analysis, and cognitive science has indicated many complex processes involved in the act of listening” (p.189). It is an active process involving various complex mental activities including receiving aural stimuli (Wolvin & Coakley,1988), attending to spoken words (Underwood, 1989; Wolvin & Coakley, 1988), attaching meaning to aural symbols (Chamot & Kupper, 1989; Wolvin &Coakley,1988), and reacting and responding to oral communication (Purdy, 1997). Rost (2002) presents a comprehensive account of the listening process. He states that “listening comprehension is the process of relating language to concepts in one’s memory and to references in the real world” (p.59). He describes in detail the neurological, linguistic, pragmatic, and psycholinguistic processes involved in the comprehension of speech.

Lynch and Mendelsohn (2002) discuss four models of listening including communication theory, information processing, social/contextual, and situated action models. They also maintain that skilful listeners use different types of learning strategies including meta-cognitive, cognitive, and social/affective strategies in combination and vary their use according to the needs of the specific situation.

Weir (2005) describes listening comprehension process in three broad terms. First, executive processing which includes goal setting, acoustic/visual input, audition, pattern synthesizer, and monitoring. Second, executive resources comprised of grammatical, discourse, functional (pragmatic), and sociolinguistic knowledge. The last one is content knowledge which is composed of internal and external knowledge. It is understandable that processes such as goal-setting, interpreting multimodal input, pattern synthesizing, and monitoring comprehension can become highly complex and challenging in the foreign language context (e.g. in lectures or discussions), putting great cognitive demands upon a student operating in their second language.

1.1. Testing listening

Richards and Renandya (2002) assert that for years there have been different theories and methods involved in developing a good and valid language test and there has been growing interest in use of assessment processes that are different from the traditional forms of assessment. In fact the shift from traditional form of assessment to the recent one reflects the changing paradigm in education in general and in language teaching in particular (Richards & Renandya, 2002). Brown (2002) emphasizes the need for approaches in testing and believes that language testing has flourished into a highly developed and sophisticated field with many facets.

Moreover, for many L2/FL students, listening is stressful (Chang & Read, 2006) and difficult undertaking. They believe that listening comprehension is more difficult than reading comprehension (Graham, 2006) because there is less opportunity or it is not easy to go back over previous input. Lynch and Mendelsohn (2002) assert that we should think of listening not as a single process but as a bundle of related processes such as recognition of sounds spoken by speakers and perception of intonation patterns conveying information. It shows that listening processes are complex and consequently testing listening comprehension is difficult in second language.

Tests of listening comprehension attempt to measure the examinee’s ability to comprehend and interpret the oral stimuli. In particular, listening comprehension tests for beginner and lower-intermediate levels attempt to assess the testees’ comprehension of the specific elements of the language—statement, questions, and brief conversational exchanges. However, Tests for higher-intermediate and advanced levels assess the examinees’ understanding of both informal and formal lectures and their comprehension of native speakers in different speech situations.

Tests of listening comprehension attempt to measure the examinee’s ability to comprehend and interpret the oral stimuli. In particular, listening comprehension tests for beginner and lower-intermediate levels attempt to assess the testees’ comprehension of the specific elements of the language—statement, questions, and brief conversational exchanges. However, Tests for higher-intermediate and advanced levels assess the examinees’ understanding of both informal and formal lectures and their comprehension of native speakers in different speech situations.

There are three main approaches to test listening. The first approach is the discrete-point testing which is based on behaviourism, the theory that language learning is a kind of habit formation through stimulus, response, and reinforcement. In the discrete-point testing, language is regarded as a combination of individual discrete “bits” and language proficiency as comprehending these individual bits. The basic technique for testing listening comprehension in this approach is presenting an utterance to the test takers in the target language and checking the test takers’ understanding of the utterance. Multiple-choice questions and true-false items are the two main methods to achieve this goal.
The second paradigm in testing, integrative testing, requires testees to combine many language elements in completion of a task. The major distinction between the discrete-point approach and integrative testing is that integrative testing examines the processing of language rather than individual bits of knowledge about language. Listening comprehension is identified as a process of decoding the meaning of words. Dictation and Cloze tests are the two prominent formats to test listening skill integratively.

Finally, communicative tests assess the use of the language for its communicative function rather than the usage. Proponents of this approach tend to provide a descriptive taxonomy rather than a testable theoretical model. Theorists have not proposed clear guidelines for construction of communicative tests. However, the concept of “authenticity” plays a major role in this type of testing. Therefore, a listening test following this paradigm should be reasonably authentic, i.e. taken from the real-life target language communication situation and possess the characteristics of the target language use situation (Buck, 2001).

Bachman and Palmer (1996) state that characteristics of test methods affect test scores to some extent and they further suggest that since it is impossible to avoid the effects of test methods, it is necessary to control them as much as possible so that the tests will be appropriate for what they are used for. Factors that commonly affect test takers performance of listening comprehension include text genre, topical knowledge, the text length, test item type, visual cues, question preview, and the number of times the text is presented. Wier (2005) included other components influencing the listening comprehension of test takers. Speech rate, variety of accent, acquaintanceship, number, and gender are the factors of input affecting listening comprehension.

Weir (2005) lists different techniques for testing listening comprehension including matching responses, dictation, short answer, and information transfer. The author states some advantages and some disadvantages for each technique. One of the techniques to assess listening comprehension is discussed briefly below.

1.2. Multiple choice tests

Multiple-Choice (MC) question formats include a stem, or prompt, and alternative responses. The stem is, in fact, the question. The alternatives that are not correct are called distractors. There are different types of MC examinations. The most prominent MC test uses MC questions with one stem and some choices; one of these choices is correct, the other ones are incorrect alternatives, called distractors (Bradbard, Parker, & Stone, 2004). Several other MC formats were designed to measure more complex thinking skills. The incorporation of more than one correct answer is a good way to reduce the chances of guessing a question correctly (Bush, 1999).

Hughes (2003) lists problems associated with MCQ. Testing the recognition knowledge, guessing, limitation in testing different components of a language, difficulty in preparing the successful items, harmful backwash, and cheating are among these problems. Moreover, the cognitive processing involved in determining an answer in this format bears little resemblance to the way we process texts for information in real-life, and it can harm the theory-based validity of the test method (Weir, 2005). However, due to some advantages including high marker reliability, ease of scoring, and objective scoring, MCQ is still a major tool in testing and assessment.

In sum, multiple-choice question (MCQ) formats are widely used to measure listening ability. Although MCQ is often criticized on the grounds that it may misrepresent natural listening conditions (Buck, 2001; Hughes, 2003), it is, probably because of its practical benefits, the most popular format for the large-scale testing of listening skill. Other test methods that require trained scorers (such as short answer questions or summary writing) would result in greatly increased costs for the large numbers of test takers. MCQ is definitely both familiar to the most test takers and highly practical as it readily allows for machine scoring.

2. Review of related literature

The effect that different test formats might have on language skills has been focus of some researches so far. Testing listening comprehension and the potential effects that different multiple-choice questions formats might have on it have been an area of interest for some researchers. Some researchers have paid attention to different ways of
presenting multiple-choice tests.

Underwood (1989) found that looking through a list of test questions before listening was helpful for the beginner learners. It removed the stress of suddenly hearing something unfamiliar and thus being distracted from the next part of the listening text. Moreover, a quick pre-listening look at the questions would probably refresh listeners' memories and force test takers to listen. Yang (1996) regarded test question preview useful in providing contextual cues to predict the framework of the listening texts. Sherman (1997) studied the effect of four test formats on listening ability of second language learners. The results showed that the sandwiched format (questions between two hearings) was the most facilitative method in comparison with other methods including question before hearing, question after hearing, and no questions.

Brindley and Slatyer (2002) reported on an exploratory study that examined the effects of task characteristics and task conditions on learner's performance in competency-based listening assessment tasks. They investigated major variables such as the nature of the input and the response mode, namely speech rate, text type, number of hearings, input source (live vs. audio-recorded), and item format. Quantitative and qualitative analyses of test scores indicated that speech rate and item format affected task and item difficulty.

Chung (2002) studied the effect of question previewing and vocabulary pre-teaching on listening comprehension. The author examined the effects of advanced organizers on 188 Taiwanese college students' listening comprehension of English language videotapes. Results showed that the group having received a combined treatment of vocabulary pre-teaching and question viewing between the two video viewings outperformed the groups which only received only vocabulary pre-teaching or no treatment on both multiple-choice and open-ended tests.

Yangawa and Green (2008) studied the effect of three different formats of multiple-choice questions including full question preview, question stem preview, and question option preview on the listening comprehension of the participants. The results indicated that the group which was allowed to preview answer options produced significantly fewer correct answers than the other two groups. They did not find a significant difference between the other two formats. They attributed the fewer correct responses for AOP to the exploiting the word-matching strategy by the listeners which is not compatible with listening strategies used in real-life situations.

Jafari and Hashim (2012) explored the impact of using two types of written advance organizers, key sentences and key vocabulary, on the improvement of EFL learners’ listening comprehension. They exploited one control and two experimental groups. The results of the study indicated that the students who received advance organizers demonstrated significant improvement in the listening comprehension post-test while the control group did not. They found significant effect of participants’ listening proficiency on their listening comprehension performance, but there was no interaction effect between the use of advance organizers and the listening proficiency level.

While most of the studies show the positive effect of question previewing on the listening comprehension, some researchers like Ur (1984) and Weir (1993) have disagreed. They believe that previewing questions before listening changes the nature of the task, probably distracts the learners’ attention, and puts a greater burden on them.

3. The present study
This study investigated the effect of four formats of multiple choice tests on listening comprehension of the EFL learners. Four formats under investigation in this study are 1. No preview (NP) in which the examinees did not receive any preview of questions. They saw the questions and options after listening to the materials. 2. Full question preview (FQP) in which the subjects saw both the stems and answer options and then listened to the listening material and chose the correct answer. 3. Answer option preview (AOP) which allowed the participants to preview the answer options (and not the question stems) and then answer the questions, and finally 4. Question stem preview (QSP) which made it possible for the examinees to look at the question stems but not the answer options before listening. After listening to the materials, the participants saw the options and chose the correct answer.
More specifically, the study was designed to answer the following main research question:
Is there a significant difference between EFL learners’ performance in listening comprehension tests with four different formats of multiple-choice tests i.e. NP, FQP, AOP, and QSP?

4. Method

4.1. Participants
A total of 103 EFL students (male and female) at Payame Noor University of Meshginshahr in Iran volunteered to participate in this study. Their major was English translation and their age ranged from 19-24. The listening subsection of the TOEFL test was administered to them. Sixty participants whose score was 1 standard deviation above and below the mean were chosen as the subjects of the study. Their proficiency level can be taken as intermediate or lower intermediate.

4.2. Instrumentation
Two listening tasks were exploited in the study. The first one was a 20-item listening test taken from the listening section of the TOEFL. It was used to identify the listening ability of the participants in EFL and to have four homogeneous groups in this skill. The second one was a listening test devised by the researchers to measure the listening comprehension of the participants. This 20-item test consisted of short conversations from Interchange Series and short passages taken from VOA. The reliability of the test was calculated using test-retest method (α=0.78).

4.3. Procedure
To achieve the goal of the research, first, 60 EFL students of English in the same level of listening ability in English (intermediate) were chosen using the listening section of the TOEFL. Then they were divided randomly into four groups. The groups were comparable in terms of listening ability in English. Each group was tested on listening comprehension using four different formats of multiple-choice tests. Necessary instructions were given but the participants were kept in the dark regarding the real purpose of the tests. The researchers analysed and interpreted the data collected from administering the tests.

5. Results
After administering the tests, the collected data were analyzed. One-Way ANOVA was used to explore the significance of the difference between the four formats of presenting the MCQ listening tests. Table 1 shows the descriptive statistics for the collected data.

<table>
<thead>
<tr>
<th>Formats</th>
<th>N</th>
<th>Mean</th>
<th>Std Deviation</th>
<th>Std Error</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP</td>
<td>15</td>
<td>13.933</td>
<td>1.387</td>
<td>.358</td>
<td>11.00</td>
<td>16.00</td>
</tr>
<tr>
<td>FQP</td>
<td>15</td>
<td>16.066</td>
<td>1.667</td>
<td>.430</td>
<td>13.00</td>
<td>19.00</td>
</tr>
<tr>
<td>AOP</td>
<td>15</td>
<td>15.666</td>
<td>1.543</td>
<td>.398</td>
<td>12.00</td>
<td>17.00</td>
</tr>
<tr>
<td>QSP</td>
<td>15</td>
<td>15.800</td>
<td>1.320</td>
<td>.340</td>
<td>13.00</td>
<td>17.00</td>
</tr>
<tr>
<td>Total</td>
<td>60</td>
<td>15.366</td>
<td>1.677</td>
<td>.216</td>
<td>11.00</td>
<td>19.00</td>
</tr>
</tbody>
</table>

In the above table, the mean and standard deviation of the four formats (NP, FQP, AOP, QSP) were presented. NP format had the minimum mean (11) and FQP had the maximum mean (19).

To find that whether the difference between the four formats was significant or not, One-Way ANOVA
was exploited. The results of the analysis are presented in the table 2.

Table 2. ANOVA for four formats

<table>
<thead>
<tr>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>42.333</td>
<td>3</td>
<td>14.111</td>
<td>6.395</td>
</tr>
<tr>
<td>Within Groups</td>
<td>123.600</td>
<td>56</td>
<td>2.207</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>165.933</td>
<td>59</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As shown in table 2, the difference between the four formats was significant. (p < 0.05). However, in order to investigate the pairwise difference between the two formats, Post hoc tests (Scheffe procedure) were adopted. Table 3 illustrates the findings.

Table 3. Multiple Comparisons

<table>
<thead>
<tr>
<th>(I) format</th>
<th>(J) format</th>
<th>Mean Difference (I-J)</th>
<th>Std. Error</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP</td>
<td>FQP</td>
<td>-2.133*</td>
<td>.542</td>
<td>.003</td>
</tr>
<tr>
<td>NP</td>
<td>AOP</td>
<td>-1.733*</td>
<td>.542</td>
<td>.024</td>
</tr>
<tr>
<td>NP</td>
<td>QSP</td>
<td>-1.866*</td>
<td>.542</td>
<td>.013</td>
</tr>
<tr>
<td>FQP</td>
<td>AOP</td>
<td>.400</td>
<td>.542</td>
<td>.909</td>
</tr>
<tr>
<td>FQP</td>
<td>QSP</td>
<td>.266</td>
<td>.542</td>
<td>.970</td>
</tr>
<tr>
<td>AOP</td>
<td>QSP</td>
<td>-1.133</td>
<td>.542</td>
<td>.996</td>
</tr>
</tbody>
</table>

* The mean difference is significant at the 0.05 level.

Table 3 shows that the difference between NP and three other formats (NP-FQP, NP-AOP, NP-QSP) was significant (p<0.05). However, the pairwise difference between the three other groups (FQP-AOP, FQP-QSP, AOP-QSP) was not significant. As shown in above tables, the mean of the listeners in FQP format (16.06) was more that the mean of the AOP (15.66) and QSP (15.80), but these differences were not statistically significant.

6. Discussion

Listening process is complex. Therefore, listening comprehension and particularly its testing are difficult in a second or foreign language. This paper investigates the effect of four formats of multiple-choice questions on the listening comprehension of the EFL learners.

The collected data and their analysis led us to the conclusion that previewing multiple choice questions before listening can facilitate listening comprehension. This finding is in line with what has been found by researchers such as Underwood (1989) and Buck (1991). The reasons for this facilitation are not investigated in this paper. However, the reasons can be, as being proposed by Underwood (1989), decreasing the stress of hearing something suddenly, refreshing listeners’ memories, and providing readiness on the part of the listener. Buck (1991) proposed other reasons for this advantage. He believes that previewing the questions helps listeners find the purpose of listening material and motivates them to get necessary information to answer the questions. Planning before listening, using contextual cues, and using different learning strategies can be other justifications. Of course, some researchers (Ur, 1984; Weir 1993) contend that question previewing in listening task changes the nature of the task and makes it more difficult for the listeners to comprehend.

However, the findings of the study does not show a pairwise significant difference for the three other
formats. In other words, previewing full questions (stem and options), previewing only answer options, and previewing only question stems facilitated the listening comprehension in comparison with not seeing the questions before listening. However, there was not a significant difference between the FQP-AOP, FQP-QSP, and AOP-QSP. This finding is to some extent against what has been found by Yanagawa and Green (2008). They contended that AOP format produced significantly fewer correct answers by the listeners. Of course, they did not find a significant difference between the two other formats.

This study shows that FQP, QSP, and AOP can improve the listening comprehension of the EFL listeners. However, the three formats don’t have a significant advantage with respect to each other in spite of a little difference in their means. The lowest mean among these three formats belongs to the AOP. This format, as claimed by Weir (2005), does not reflect the real-life communication and the validity of this format of test is jeopardized. Another interpretation for the low mean of AOP can be the difficulty of memorizing of four options for the listeners.

The number of participants and the short listening test can be the major limitations of the study. Other researchers interested in this field can replicate it using more participants and in different contexts. Moreover, the fundamental issue of validity of these formats should be scrutinized by researchers.

7. Conclusion
Listening comprehension is considered as a polestar of second language acquisition in theory building, research, and pedagogy (Dunkel, 1991). In spite of its importance, listening has been overlooked by L2/FL teachers and researchers and received less attention than the other language skills and also has often been referred to as “passive or receptive” skill. However, for the past decades, there has been an increased focus on L2/FL listening comprehension and many had acknowledged its importance in language acquisition and research. Therefore, a considerable effort should be devoted to prepare valid and reliable listening tests. This study investigated the impact of four formats of presenting multiple-choice questions on listening comprehension of the EFL learners. The findings lead us to the conclusion that previewing questions (only stem, only options, or both) can help EFL listeners in listening comprehension in comparison with not previewing the questions at all. Of course, three other formats (FQP, AOP, QSP) do not have a considerable difference in their effect on listening comprehension of the EFL listeners with respect to each other. Test developers, teachers, material designers, and curriculum developers can benefit from the findings of this study. Of course, the issue of validity of these formats in testing listening comprehension needs further investigation.

References


