Effects of Test Procedures on EFL Learners' Listening Comprehension

英文科　李宜軒老師

Definition of Key Words

Test Procedures

Questions Before: It is the procedure that allows students to preview listening questions prior to the two hearings of the text.

Questions After: It is the procedure that gives students the listening questions after two hearings of the text.

Sandwich: It is the procedure that allows students to read the listening questions between the two hearings.

Question Types

Global Questions: Questions require the test taker to obtain the gist of the text.

Local Questions: Questions require the test taker to locate details, such as name, time, and location.

Inferential Questions: Questions demand the test taker to make inference from the stated facts.

Introduction

While scholars are growing to understand the significant role played by listening, they recognize more and more importance of teaching listening comprehension. This recognition has resulted in an increase in the approaches to enhance students' listening comprehension. These approaches encompass listening strategy training (O'Malley, Chamot, and Kupper, 1989;; Rubin, 1990;), the use of visual aids (Mueller, 1980; Chaudron, 1983; Chiang and Dunkel, 1992), and incorporation of previewing questions (Herron, 1994; Berne, 1995). While all of these prove their effectiveness in fostering students' comprehension, the field of previewing questions, is relatively less highlighted. However, in the instructional settings, the use of previewing questions which takes little preparation time and class time is most

accessible. The present study is thus primarily interested in the previewing questions in the listening.

The Effect of the Test Procedures

Brown and Yule (1983) affirm the benefits of the previewing questions, claiming the previewing would drive listeners to listen for a purpose. This purpose would enhance both of their motivation and strategy use. Yet Sherman (1997) upholds an opposite opinion by proving that previewing questions do not necessarily result in the positive influence. In her study, Sherman (1997) proposes another procedure named Sandwich version, which yields better outcome compared with Questions Before and Questions After. Sandwich version not only shows the most satisfactory result but also wins the supportive evaluation in the follow-up questionnaire.

The exceeding result of Sandwich raises the importance of procedures. The scope of the study originally confined to the previewing itself is now extended to the locus of the previewing in the whole test procedure. At the first glance Sandwich and Question-Before are mutually exclusive. However, Sandwich version, which gives the participants the questions between the two hearings, can also be regarded as another form of Question-Before for its second hearing. It is the locus of the questions being given in the whole testing procedure that makes Question-Before and Sandwich dissimilar from each other. So to speak, it is the testing procedure that affects participants' listening comprehension in Sherman's (1997) research.

The Effect of Proficiency Levels on Listening Comprehension

Except for the sole influence of the test procedures, another issue to be explored is the relationship of test procedures and proficiency levels. Mueller (1980) in his research points out that the effects of visual aids are closely related to the students' level of proficiency. His result reveals that, for less proficient students, visual aids play a pivotal role in affecting their performances. But they exert much less influence on more proficient students' listening. Thereby, the effects of the advance organizers may be different for learners depending on their proficiency level.

The Influence of Question Types on Listening Comprehension

The third issue under discussion is the relationship between the question types and the test procedures. Shohamy and Inbar (1991) conduct a study examining how the question types affect the outcome of students' listening comprehension. The global, local, and trivial questions are brought to be focus in their discussion. Despite the fact that local questions and trivial ones overlap in terms of their nature, the result generally shows that participants are more capable of handling the local questions than the global ones. However, they suggest that pre-listening questions, especially referring to global ones, will promote "the overall theme of the passage."

Another question type that is not mentioned by Shohamy and Inbar (1991) but is more and more frequently seen: inferential questions. Nowadays, the ability to infer is more and more highlighted by a bunch of scholars (Nunan, 1999; O'Malley, Chamot, & Kupper, 1989). Making inference assumes the essential role in strategy training. Employing inferential questions in the listening test is one way to boost students' skills in making inference. And it would be quite helpful for the test takers to handle the listening text if they can preview this type of questions.

Research Questions

Four main research questions related to above discussion are posed as follows.

1. How do Questions Before, Questions After, and Sandwich differ in affecting participants' listening comprehension?
2. How do participants of high, intermediate, and low proficiency level diversify in the performance of the listening test with different procedures?
3. How do the effects of global questions, local questions, and inferential questions differ with different procedures?
4. What do participants feel and think about Questions Before, Questions After, and Sandwich?

Research Design

Participants

The participants were Taiwanese college students (20 male and 151 female) from 10

departments of National Taiwan Normal University. Among 171 participants, most of them were English majors. In the current study, 171 participants were sampled from 9 listening classes. Five classes were Basic Aural-oral Training in English, three were Intermediate one, and one was Advanced. However, with a view to clearly distinguishing participants' proficiency level, all the participants were required to take a TOEFL listening test beforehand. By re-grouping participants based on their TOEFL scores, the study would have more confidence in the validity of the proficiency variable. On the whole, participants were distinguished as three levels: low, intermediate, and high level.

The Listening Comprehension Test

Three monologues included in the listening comprehension test were extracted separately from the TOEFL tests in 1999, 1997 and 1996. The monologues were similar in the light of length and difficulty. All the questions in the present study were re-designed. There were two reasons to do so. First, the original questions in the TOEFL test were multiple-choice. The current study preferred using short-answer questions and thereby the original questions were not suitable. Secondly, since the present study was also interested in the relation between testing procedures and question types, the questions employed here needed to be divided into global, local, and inferential questions. For the three texts, nine questions were devised in total, with each text containing three questions. Among three questions, one pertained to be global, another was local, and the other was inferential.

Questionnaire

Questionnaires were designed to probe into participants' perception of the various effects. The questions were formulated as personal interrogation which required the respondents to tick, with the choices of Question Before, Question After, Sandwich and None. There was a brief definition of the terms at the top of the questionnaire. Six questions were involved. General questions attempted to detect the factors of habit, anxiety, distraction, difficulty, enhancement, preference.

The Experimental Procedure

Participants' proficiency levels had been distinguished before the major testing was

conducted. Unlike the TOEFL test which was held in class, the researcher implemented the major testing out of class by recruiting the participants on her own. In order to fortify their motivation for attending the test, the researcher offered NT 100 to each participant as a reward. 171 out of 216 who had participated in TOEFL agreed to take the test. Since this test was conducted out of class, the researcher had to cooperate with participants in terms of their available time. Yet with a view to centralizing the testing time, the researcher distributed a note, allowing the participants free to choose their available time.

In the testing, participants were required to listen to three monologues (texts1, 2, and 3). In order not to blend the effect of the test procedure with that of the texts, the researcher decided to alternately implement each text with each procedure. Thus, the sequence of the test procedures encompass six possibilities, including QB→S→QA, QB→QA→S, QA→QB→S, QA→S→QB, S→QB→QA, and S→QA→QB. These six sequences were manipulated in the formal tests by turns. The following table is the time allotment of sequences.

Table 1. Time Allotment of Sequences

	10/17 Monday	10/18 Tuesday	10/19 Wednesday	10/20 Thursday	10/21 Friday
9:10	S→QA→QB		QA→S→QB		QA→QB→S
10:10					
12:10	QA→QB→S		QB→S→QA		S→QA→QB
1:10	S→QB→QA	QB→S→QA	S→QB→QA	QB→QA→S	QB→S→QA
2:10	QB→QA→S		S→QA→QB		QA→QB→S
3:10		QA→QB→S	QB→QA→S		QB→QA→S
4:10	QB→QA→S	S→QA→QB	QA→QB→S	QA→S→QB	S→QB→QA

Although participants underwent different sequences of procedures, each of them definitely experienced all the procedures. The questions for the three monologues were printed in three respective paper sheets. As soon as they finished the questions of one monologue, they needed to turn in that paper sheet before the next monologue started. Hence, participants might not have the chance to change their answers.

Once the participants completed their questions, they would be given a questionnaire.

Then, the researcher made a lucid announcement to secure that all the participants understood the terms of Questions Before, Questions After, and Sandwich.

Result

The Interaction of Test Procedures and Proficiency Levels

The present study adopted linear mixed-effects model and ANOVA to analyze the data. In this model, the total score was regarded as response, whereas proficiency levels and procedures served as fixed effects. And each participant (id) was considered as a random effect. The present study was interested in probing into the interaction between proficiency levels and test procedures. Hence, the applied model looks like the following.

total score = id + procedure + level + procedure*level + error

$$Y_{jk} = \mu + \beta_j + \gamma_k + (\beta\gamma)_{jk} + \varepsilon_{jk} \text{ , } j,k = 1,2,3$$

After being analyzed by ANOVA, Table 2 indicates that each single variable reached the acceptable degree of significance, but this significance did not occur in the interaction of procedures and proficiency levels. That's to say, the performances of the three proficiency levels were not distinctively differed with the different procedures.

Table 2. The Fixed Effect of Interaction between Proficiency Levels and Test Procedures

Source of Variation	DF of Numerator	DF of Denominator	F-value	p-value
Procedure	2	334	25.189	**.0000**
Level	2	168	66.967	**.0000**
Procedure*Level	4	334	.401	**.8231**

Since the significance did not occur in the interaction of procedures and proficiency levels, this effect was dropped. As for the rest of the fixed effects, they were preserved.

total score = id + procedure + level + error

$$Y_{jk} = \mu + \beta_j + \gamma_k + \varepsilon_{jk} \text{ , } j,k = 1,2,3$$

ANOVA also proved this model was suitable. With a view to further investigating whether significant differences occurred in the three procedures, two-sample t-test was utilized. According to Table 3, differences among the three procedures all achieved the degree of significance. It could be generally observed that Sandwich was the best-performing procedure, followed by Questions Before and Questions After.

Table 3. The Two-Sample t-test of Differences in Three Procedures (for Total Scores)

Contrast (in group)	Mean Difference	Std. Error	t-test	p-value
Question Before —Question After	1.182	.345	3.426	**.001**
Question Before —Sandwich	-1.203	.345	-3.487	**.001**
Question After—Sandwich	-2.384	.346	-6.890	**.000**

The Interaction of Test Procedures and Question Types

In this section, the linear mixed-effects model and ANOVA were employed in the same manner. But now, it was interested in examining the interaction between procedures and question types. Thus, the adopted model should be like the following.

total score = id + procedure + level + question + procedure*question + error

$$Y_{jkl} = \mu + \beta_j + \gamma_k + \delta_l + (\beta\delta)_{jl} + \varepsilon_{jkl}, \; j,k,l = 1,2,3$$

Based on ANOVA, the result indicates that the scores of different question types were not significantly affected by procedures (Table 4). All the question types yielded highest grades in Sandwich, followed by Questions Before and Questions After. In terms of the question types, the grades of global questions took the lead, surpassing both the local questions and the inferential questions. In addition, global questions clearly showed higher grades in both of Questions Before and Sandwich procedures, implying that the scores of global questions were better heightened when being manipulated with these two procedures.

Table 4. The Fixed Effect of Interaction of Test Procedures and Question Types

Source of Variation	DF of Numerator	DF of Denominator	F-value	p-value
Procedure	2	1358	19.581	**.000**
Level	2	168	67.934	**.000**
Question	2	1358	12.281	**.000**
Procedure* Question	4	1358	1.115	**.413**

Since the interaction of test procedures and question types did not yield the significant outcome, the model was rearranged. In this new model, the variable of procedures*question types was eliminated. The model was adjusted as follows:

total score = id + procedure + level + type + error

$$Y_{jkl} = \mu + \beta_j + \gamma_k + \delta_l + \varepsilon_{jkl}, j,k,l = 1,2,3$$

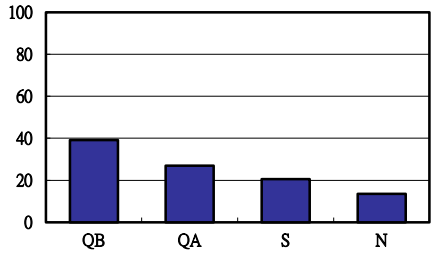
Because fixed effect of question types was proven significant, two-sample t-test was required again to detect whether the significant differences lay between any two of the three question types. Table 5 indicates that significant differences occurred to global vs. local and global vs. inferential. Nevertheless, the difference between local question and inferential question did not attain to the significant degree (p= .326).

Table 5. The Two-Sample t-test of Differences in Question Types

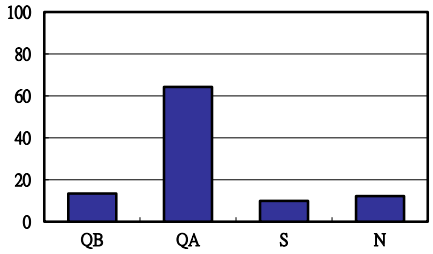
Contrast (in group)	Mean Difference	Std. Error	t-test	p-value
global question—local question	.472	.127	3.717	**.000**
global question— inferential question	.596	.127	4.693	**.000**
local question —inferential question	.125	.127	0.984	**.326**

Participants Responses to Questionnaire

1. Habit



2. Anxiety



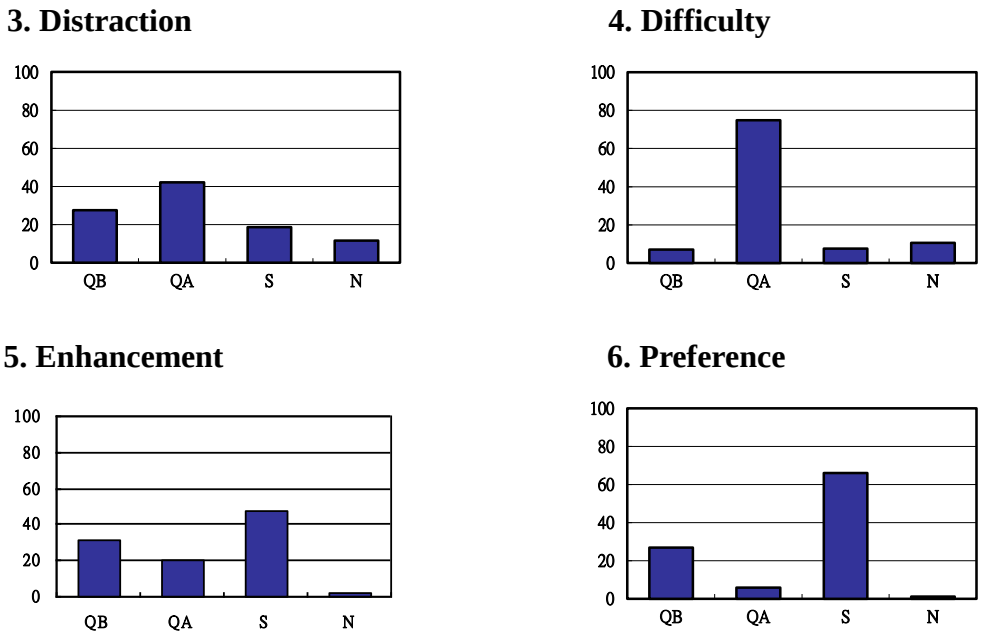


Figure 1. The Results of Questionnaire in Six Categories (%)

Holistically speaking, Questions After was by far most anxiety-provoking, most distracting, most difficult, least facilitative, and least preferable procedure in the eyes of participants. The evaluation on Questions Before was better. Sandwich, on the whole, was a procedure worth recommending. No matter in anxiety or distraction category, Sandwich begot least negative responses. But in both of enhancement and preference categories, Sandwich earned most positive evaluation. But it needed to note that Sandwich was also the procedure which was least used for daily listening activities. Hence, it would be worthwhile for teachers to introduce Sandwich in their listening class as a way to improve students' listening ability.

Conclusion

Discussion of the results

The major findings of the research are presented as follows. First, the researcher was interested in looking into how Questions Before, Questions After, and Sandwich differ in affecting participants' listening comprehension. The outcome clearly verified that Sandwich was the best performing procedure, followed by Questions Before and Questions After. The fact that Sandwich yielded leading scores could be elucidated from the aspect of the way listeners listened. Sandwich tended to drive the listeners to listen in an all-around manner first. Having a rough picture of the text, listeners could further deepen their impression through the

following previewing. Yet regarding participants in Questions Before, they might search for the answers urgently right at the beginning, failing to construct an overall picture of the text.

Secondly, the researcher also aimed to investigate how participants of high, intermediate, and low proficiency level diversify in the performance of the listening test with different procedures. Based on the study outcome, there was no interaction occurring between variables of procedures and levels. It was speculated that participants' levels were probably not heterogeneous enough. In the present study, most of the participants were English majors. Compellingly grouping them into three levels based on their TOEFL risked the possibility that low level participants were not really low at all. As a result, the expectation of different efficacy on different levels brought by the three procedures was not realized.

The third research question was concerned with how the effects of global questions, local questions, and inferential questions differ with different procedures. The result indicated that there was no statistically significant interaction between procedures and question types. Although the significant interaction was not found, there was a tendency that the scores of global questions were much higher in Questions Before and Sandwich. Besides, the outcome also showed that global questions were relatively simpler, followed by local questions and inferential questions. It seemed to signify that the simpler the question types were, the more meaningful the previewing would be. The local and inferential questions, especially the latter ones, were relatively harder so that the effect of previewing was abated.

The final research question inquired what participants feel and think about Questions Before, Questions After, and Sandwich. The questionnaire result showed the overwhelmingly high percentage of participants preferred Sandwich to Questions Before. This could be explained by the fact that most of participants were English majors whose proficiency and confidence were higher than the ordinary. They didn't rely on the previewing at the first beginning to secure their sense of security. They were comfortable to be directly engaged into the first hearing.

Pedagogical Implication

In view of the study result, teachers are suggested to incorporate Sandwich into listening training. Sandwich is a procedure worth recommending. Based on the overall outcome, better listening comprehension is not the sole advantage of Sandwich. From participants' affective

perspective, Sandwich is also less anxiety-provoking than Questions Before. Since listening has long been regarded as tension-stimulating activity, adopting Sandwich to train or test students would help reduce the possibility that tension degrades their performances. Given the fact that Sandwich is the procedure with most enhancement and least anxiety, the present study strongly urges teachers who used to adopt only Questions Before and Questions After to add it in the listening class.

Another suggestion to make is that teachers should alert the sequence of conducting these procedures. Especially for freshmen, teachers need to implement Sandwich or Questions Before first, which appear not as difficult as Questions After. If students are rather unaccustomed to listening, teachers may first introduce Questions Before since students are more familiar with it than with Sandwich. After they get on the track of listening training, teachers may shift to Sandwich by postponing previewing questions until the first hearing. When students develop more proficiency and confidence in listening, teachers may progress to Questions After. Employing Questions After to deepen more difficulty in the listening test should be the ultimate goal rather than the initial one for the listening training.

Bibliography

Berne, E. (1995). How does Varying Pre-Listening Activities Affect Second Language Listening Comprehension? *Hispania* 78, 316-29.

Brown, G. and Yule, G. (1983). *Teaching the spoken language.* Cambridge: Cambridge University Press.

Mueller, G.. (1980). Visual Contextual Cues and Listening Comprehension: An Experiment. *Modern language Journal* 64, 335-340.

Nunan, D. (1999). *Second Language Teaching and Learning.* Boston: Heinle & Heinle.

Nunan, D. (2002). *Listen In: A three-level listening series.* Second edition. Boston: Heinle/ Thomson Learning.

Rankin, P. T. (1978). Listening Ability: Its Importance, Measurement, and Development. *Chicago Schools Journal* 12, 177-9, quoted in W.M. Rivers and M.S. Temperley.

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

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