Available online at www.sciencedirect.com

SciVerse ScienceDirect

System 41 (2013) 575–586

SYSTEM

www.elsevier.com/locate/system

Investigating the effects of multiple-choice listening test items in the oral versus written mode on L2 listeners' performance and perceptions

Anna C.-S. Chang [a,*], John Read [b]

[a] *Hsing-Wu University, Linkou, New Taipei, Taiwan*
[b] *University of Auckland, Auckland, New Zealand*

Received 11 February 2011; revised 8 June 2013; accepted 15 June 2013
Available online 19 July 2013

Abstract

This study explored the effects of varying the mode of presenting multiple-choice tests items on L2 listeners' performance and perceptions. In the study, 87 Chinese college students in Taiwan took a listening test with 30 written multiple-choice items and another 30 items presented orally. In addition to test-taking, the participants also completed a short questionnaire on their perceptions of the task, and participated in a post-test discussion. The test results showed that the students scored almost the same with the oral (66%) as the written mode (68%). However, an interaction between the modes and the students' listening proficiency was detected: lower level students scored significantly higher on written items, whereas higher proficiency students had similar scores in both modes. Although the students' overall listening performance did not vary significantly, in their questionnaire responses 78% of the students favoured the written over the oral mode. There was evidence that higher proficiency students were able to deal with the memory load of the orally presented items through effective test-taking strategies that lower level students did not employ. Thus, while presenting the test items orally may create a pure test of listening ability, it appears that there are good reasons for written presentation of multiple-choice listening items.

Keywords: Multiple-choice questions; L2 listening; Listening comprehension; Listening performance

1. Introduction

In assessing the proficiency of second language learners, there is a long tradition of testing each macroskill separately, despite a recent trend in favour of integrated testing across skill areas (see Plakans, 2012; for an overview of this development). Each skill poses its own challenges in making a valid assessment of learner abilities, but our focus here is on listening comprehension. The standard approach to the design of a listening test is to present the test-takers with one or more (usually pre-recorded) spoken texts and to require them to demonstrate their understanding of various aspects of the text content. Although it is possible to create response formats based on pictures or diagrams (see Heaton, 1988; for examples), a much more common practice, especially for intermediate and advanced level adult

* Corresponding author. Tel.: +886 2 26015310x2471.
E-mail address: annachang@livemail.tw (A.C.-S. Chang).

0346-251X/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.system.2013.06.001

learners, is to use response items which require a significant amount of reading and/or writing. The need to apply reading or writing skills can be seen as introducing what Messick (1996) called construct-irrelevant variance into a listening test, in that it no longer assesses just the learners' listening ability.

Reading is required in particular for the multiple-choice item, which is one of the most widely used formats in listening comprehension tests. These items are often criticized on a number of grounds: they test only recognition knowledge, they invite guessing, they are susceptible to cheating, they difficult to construct successfully, and they do not represent an authentic language use task (Buck, 2001; Hughes, 2003; Weir, 1990). However, multiple-choice items are valued by test developers because of their reliability, ease of marking, and suitability for large-scale testing (Thompson, 1995). In the context in which the present study was undertaken, multiple-choice tests are not just very familiar to learners but also the most feasible means of assessing the language knowledge and receptive skills of the typically large student populations in schools and universities.

The conventional approach in multiple-choice listening tests is to present the test items in written form but, in order to reduce if not eliminate the reading element, another possibility is to administer the items as well as the input text in spoken form. In their recent study (discussed further below), Yanagawa and Green (2008) experimented with presenting only the stem questions or only the answer options in written form. In this study we went a step further by comparing student performance on multiple-choice items that were written in full on the answer sheet (what we called the written mode) with items that were heard only on the audio-recording (the oral mode). To date, there appears to have been no published research providing empirical evidence for the common assumption that presenting listening test items orally creates too much of a cognitive challenge or memory burden for second language listeners, and so we set out to investigate whether this was really the case. In addition to analysing the test results, we were interested in the students' perceptions of the two forms of presentation after they had experienced both modes. We also wanted to investigate whether there were differences in the performance of higher and lower proficiency students according to the mode of the test items.

2. Literature review

A number of previous studies have looked at the relative difficulty of listening test items like the multiple-choice format, which involve selecting responses, as compared to constructed-response items like gap-filling or short answer questions. In general selected-response items, which draw on reading skills, have been shown to be easier than items requiring some writing. For example, Eykyn (1992) compared multiple-choice items with other task types (choose a picture, wh-questions, vocabulary lists) as measures of the comprehension of high school French learners listening to radio texts. The results showed students scored the best on the multiple-choice format. Teng (1998) studied three test methods (multiple-choice items, cloze, and short answer questions), and also found that her university level students scored significantly higher when responding to the multiple-choice items. Another study by Cheng (2004), investigating standard multiple-choice, multiple-choice cloze and open-ended questions, showed that her students performed best on the multiple-choice cloze, but worst on open-ended questions.

Chang (2005) explored the effect on test-takers' listening performance of allowing them to preview two types of test format – multiple-choice versus short answer questions – before taking the test. She found that all of the test-takers, whose language proficiency ranged from beginning to intermediate level, achieved higher scores on the multiple-choice items. In addition, lower-level students doing multiple-choice items outscored those of a higher level responding to short-answer questions. These consistent findings suggest that test formats which require test-takers to write words in gaps or to compose responses to short-answer questions impose a significant extra demand which is not relevant to the construct of listening ability.

Although written multiple-choice items have been shown to be less difficult overall than constructed-response formats, there is evidence from other research that lower proficiency learners may be disadvantaged by the reading requirement (Chang, 2005; Wu, 1998). Wu investigated test-taker performance on multiple-choice listening items by means of immediate retrospection. She found that, whereas being able to read the written test items in advance had a facilitating effect on the ability of more proficient listeners to understand the input text, those who were less advanced had difficulty in understanding the answer options in the items and tended to engage in a great deal of uninformed guessing. Wu saw this as threatening the construct validity of the listening test. Chang (2005) obtained supporting evidence for this point of view in her study, which showed that less proficient test-takers were not able to benefit from previewing multiple-choice listening items in the way that more proficient learners could.