

## Can we test listening authentically?<sup>1</sup>

---

**Richard Watson Todd**

King Mongkut's Institute of Technology Thonburi

### **Abstract**

With the present move towards authenticity in language teaching, there should be a similar shift in language testing, but such a move has not been apparent. This paper addresses why testing has not incorporated authenticity to a greater extent by examining the issue of authenticity in the testing of listening. The extent of authenticity in tests is considered along four dimensions : materials, testing situation, task and testee's response. It is concluded that it is possible to create authentic listening tests, but this can only be done at the expense of the test's reliability and practicality.

One of the many movements associated with Communicative Language Teaching is the move towards using authentic materials in the classroom. By authentic materials is meant materials which were "created to fulfill some social purpose in the language community in which [they were] produced" (Little et al., 1989: 25), where the language community is not associated directly with language teaching. Widdowson (1979) builds on this definition by arguing that the reader's/listener's response to the materials must also be authentic in nature for materials to be considered authentic. A further dimension to authenticity is what Nunan (1989: 60) terms "activity authenticity" whereby the situation in which the learner uses the materials and the task accompanying the materials should also reflect how such materials

are used in the real world. Thus truly authentic use of authentic materials comprises four dimensions: the materials themselves, the situation in which they are used, the task accompanying the materials and the learner's response to the materials.

Parallel to the movement towards authenticity in teaching, there is a similar desire to increase the amount of authenticity in testing (Cohen, 1994; Weir, 1990), where authenticity is seen as one aspect of validity (Bachman, 1990). This desire to increase authenticity in testing is rooted in two reasons. Firstly, based on the dictum "Test what you teach", if we use more authentic materials in teaching, this should presumably be reflected in the methods used for testing. Secondly, tests which closely match the

real world give a better picture of how testees can purposefully use language (Carroll and Hall, 1985). More authentic tests reduce the extent to which a test measures ability to cope with the testing method, and focus more on real linguistic ability. In spite of this, tests are still largely inauthentic along at least one of the four dimensions above. The presence of inauthenticity in tests is, perhaps, an artefact of testing, since testing situations where scores are awarded for performance introduce anxiety and artificiality. Tests can therefore never truly reflect the real world and all tests contain some measure of inauthenticity (Spolsky, 1985). We can, however, attempt to increase the level of authenticity in a test, in other words, the extent to which it matches the real world.

In this paper, I will examine the problems which arise when you try to create authentic tests, with specific reference to listening tests, as these are notoriously difficult to 'authenticise'. In doing this, I will focus on the problems arising in each of the four dimensions of authenticity in turn, and finally I will look at the level of authenticity of the methods typically used to test listening.

### **Authentic materials**

Authentic materials for testing listening typically take the form of tapes of lectures, news broadcasts and conversations. It is a moot point whether such recordings should be considered authentic as the simple acts of recording and replaying in a different situation introduce an element of inauthenticity. Since using recordings is a very practicable way of testing listening, however, let us assume that recordings of authentic speaking are themselves authentic.

Using recordings of authentic speaking presents a large number of problems for test designers, largely due to the nature of speaking. Speaking involves clustering, redundancy,

reduced forms, hesitations, backtracking, colloquial language, stress, intonation, and frequently a fast rate of delivery (Brown, 1994). These characteristics make listening to authentic speech a difficult task for most learners.

However, the extent to which these characteristics are manifested, and thus perhaps the extent to which a test may be considered difficult, often depends on the discourse type of the text. Formal lectures and news broadcasts can display similarities to written language and might be considered less marked in terms of speech characteristics and thus less difficult. Conversation, on the other hand, presents all the above features of speech in abundance together with others which can cause further problems. Conversations include 'ungrammatical' sentences, interruptions, references to context which the listener cannot know, in-jokes, overlapping speech and so on, all of which can make listening to an authentic conversation difficult even for a native speaker.

The use of authentic materials, and especially authentic conversations, as input for listening test would therefore seem to be restricted to students at an advanced level. Such an attitude smacks of defeatism, and one way around this, so that authentic materials can be used even with beginners, is to make the task accompanying the materials very simple. For example, Thai learners listening to an authentic news broadcast can be asked to identify the names of Thai politicians mentioned. This, however, as we shall see, may cause other dimensions of authenticity, namely the authenticity of the situation and task, to be reduced.

### **Authentic situations**

Attempting to create an authentic situation for a listening test can be problematic. While people do sit back and listen to, say, news

broadcasts and lectures, other discourse types, such as conversations, are frequently tested in situations which are inauthentic.

Listening to tapes of conversations, even where the conversation is authentic, is a situation which rarely occurs in real life. We may eavesdrop on other people's conversations (though this is hardly a skill which should be promoted) and we may listen to conversations in soap operas on television, but most listening to conversations in the real world occurs in situations where we are present and have a chance to participate.

Thus to create authentic listening situations, either we can use transactional monologues, such as lectures, and videos where testees would not be expected to participate in the real world, or we can use conversational situations where testees have a chance to participate, such as interviews or group discussions. The latter are frequently used to test speaking but usually overlook the listening aspects of the situation. This imbalance could easily be redressed by assigning marks specifically for listening, and thus an authentic situation could be created for a listening test.

### **Authentic tasks**

The authenticity of a listening task hinges on the extent to which it matches what people really do when listening. The main problem with creating an authentic task for a listening test is that some response, either written or spoken, is needed from the testees as a basis on which to award marks. While this presents no problem with, say, an interview, with transactional listening such as a news broadcast some measure of inauthenticity is imposed when we require testees to give a response.

Most listening tests require the testee to listen and give a written response (Rost, 1990; Weir, 1990). But when, outside the language classroom, do people listen and write? Secretaries take dictation, students take notes in lectures, people take messages on the telephone, journalists take notes, and applied linguists transcribe tapes! Although this may seem a wide range of situations, it does impose severe restrictions on the kinds of listening available for tests if we emphasise the need for authentic tasks in listening tests.

Even further restrictions are placed on the options available if we require testees to write responses to written questions. Thinking of situations in the real world where people undertake a task such as listening and answering multiple-choice questions quickly becomes an exercise in creativity and is ultimately futile. If, for whatever reason, it is essential to use a multiple-choice listening test, then we have to accept that the listening task is inauthentic. Indeed, it would appear that authenticity of task is not a major consideration in most listening test design, even of such highly developed tests as TOEFL and IELTS. Thus, in the area of task at least, the practicalities of testing seem to override any desire for authenticity.

### **Authentic responses**

The last aspect of authenticity, that of the testee's response, is closely related to that of the task. Before I discuss the authenticity of the testee's response in detail, I would like to digress and consider models of listening.

Rost (1990) identifies two main models of listening. The first, the traditionally-oriented information processing model, regards the listener as a tape recorder whose role is to reconstruct the speaker's message as accurately as possible. The second, relevance theory,

argues that the responsibility for interpreting the message rests on the listener, and that the listener will pay attention to those aspects of the speaker's message which are most relevant to him/her and may ignore or completely reinterpret other aspects of the message (Sperber and Wilson, 1986; Wilson, 1994). There is a general move in linguistics at present away from the first of these models and towards the second where the listener becomes the main factor in listening.

This shift in how we view listening has profound implications for the authenticity of the testee's response in listening tests. Let us examine note-taking in lectures to see what these implications involve. Many listening tests which focus on note-taking in lectures take the form of controlled note-taking such as completing a table (Rost, 1990). While this may aid the marking of the tests, it also introduces inauthenticity of task and response. For most lectures in the real world, the audience starts off with blank paper or complete handouts, and so giving testees a set of partially completed notes to fill in is an inauthentic task. Furthermore, the partially completed notes assume that the test designer's and the testees' interpretations of the lecture are the same, which relevance theory suggests may not be true. The controlled note-taking forces the testee to follow the test designer's interpretation of the lecture, leading to possible inauthenticity of response.

One way to solve this problem is to allow free note-taking, where testees start the lecture with blank paper. This, however, leads to other problems, especially of setting up marking criteria. The testees' different perspectives could mean that they will be interested in different aspects of the lecture and thus take different notes. Even for testees who fully comprehended the lecture, the range of possible responses is vast and may even include no notes taken at all if the testee found the lecture uninteresting or knew the content well already. The notes taken by the testees, then, could bear very little relationship to the extent to which they understood the lecture, and thus valid marking, frequently the main purpose of a test, becomes impossible. Such a situation is clearly unacceptable as a listening test and, as a consequence, most listening tests do not encourage authenticity of response.

### **The authenticity of typical listening tests**

Having looked at the four dimensions of authenticity, we can now consider the extent to which the methods typically used to test listening exhibit authenticity. Table 1 shows ten methods of testing listening and the extent of the authenticity of the methods in each of the four dimensions discussed above. The levels of authenticity identified are only intended as a rough guide since the exact format and content of the test may influence these greatly.

	Are the materials authentic?	Is the situation authentic?	Is the task authentic?	Is the response authentic?
Minimal pair discrimination	No	No	No	?
Stress/intonation identification	No	No	No	?
Following instructions	?	?	?	?
Short-answer questions on a lecture	Yes	Yes	No	No
Multiple-choice on a conversation	Yes	No	No	No
Interview	Yes	Yes	Yes	Yes
Dictation	?	?	?	?
Taking notes from the news	Yes	Yes	No	No
Gap-filling while listening	?	?	No	No
Group discussion	Yes	Yes	Yes	Yes

Table 1 Some methods for testing listening and their authenticity

### Conclusion

Most of the typical methods of testing listening involve at least some degree of inauthenticity. This is largely because, as with other aspects of authenticity (Bachman, 1990), the more authentic you try to make a test the less reliable it frequently becomes (Carroll, 1980; Weir, 1993), and since reliability is one of the benchmarks of a 'good' test, it is not surprising that most tests lack authenticity. This payoff between authenticity and reliability, one specific instance of the general payoff between validity and reliability in the testing field (Underhill, 1982), suggests that we may have to

quench any desire for authenticity in testing and accept it as an unattainable goal.

There are, however, situations where authenticity is of such importance that it may take precedence. For example, a company offering a position requiring specific listening skills may need clear details of how well applicants can be expected to perform in this position, and thus may require a test which closely mirrors the skills required in the actual position. In situations such as this, where authenticity is paramount, the options available to the test designer are few. The test may need

---

to forsake reliability in favour of authenticity, or it may require a lot of time and effort to be devoted by the tester as happens with interviews and group discussions. Thus test designers must

be prepared to give up some measure of reliability or practicality if they wish to make their tests authentic, or they must accept the inauthenticity of their tests.

---

<sup>1</sup> Paper originally presented at CULI's 1996 National Seminar: "Language assessment: quality assurance in internationalization", December 2-3, 1996.

## References

- Bachman, L. F. (1990) *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Brown, H. D. (1994) *Teaching by principles: an interactive approach to language pedagogy*. Englewood Cliffs, NJ: Prentice Hall Regents.
- Carroll, B. J. (1980) *Testing communicative performance: an interim study*. Oxford: Pergamon.
- Carroll, B. J. and Hall, P. J. (1985) *Make your own language tests: a practical guide to writing language performance tests*. Oxford: Pergamon.
- Cohen, A. D. (1994) *Assessing language ability in the classroom*, 2nd edition. Boston, Mass.: Heinle & Heinle.
- Little, D., Devitt, S. and Singleton, D. (1989) *Learning foreign languages from authentic texts: theory and practice*. Dublin: Authentik.
- Nunan, D. (1989) *Designing tasks for the communicative classroom*. Cambridge: Cambridge University Press.
- Rost, M. (1990) *Listening in language learning*. Harlow, Essex: Longman.
- Sperber, D. and Wilson, D. (1986) *Relevance: communication and cognition*. Oxford: Blackwell.
- Underhill, N. (1982) The great reliability/validity trade-off: problems in assessing the productive skills in B. Heaton (Ed.) *Language testing* (pp. 17-23). London: Modern English Publications.
- Weir, C. J. (1990) *Communicative language testing*. Hemel Hempstead, Herts.: Prentice Hall.
- Weir, C. J. (1993) *Understanding and developing language tests*. Hemel Hempstead, Herts.: Prentice Hall.
- Widdowson, H. G. (1979) The authenticity of language data in H. G. Widdowson *Explorations in applied linguistics* (pp. 163-172). Oxford: Oxford University Press.
- Wilson, D. (1994) Relevance and understanding in G. Brown, K. Malmkjoer, A. Pollitt and J. Williams (Eds.) *Language and understanding* (pp. 37-58). Oxford: Oxford University Press.