

Research Summary: Once or twice? A critical review of current literature on the question how many times the audio recording should be played in listening comprehension testing items

Glyn Jones
Pearson

The purpose of this paper is to review the currently available literature on the question of whether, in listening comprehension tests using recorded speech, the audio signal should be played only once or should be repeated. Surprisingly little has been written about this issue, in view of its relevance to listening test design.

Such previous literature as there is on this topic can be divided into two categories: theoretical arguments for one position or another (listen once or listen twice) and empirical studies.

1. Theoretical arguments

Theoretical arguments fall into four broad categories.

1.1 Arguments from authenticity

These have been put forward for both positions. In relation to "real life" listening experience they maintain either that one generally only gets to hear an utterance once, or, conversely, that one can generally elicit a repetition in one way or another. In an example of the first position Fortune (2004) states: "In virtually all real-world listening situations we hear the text only once." Murray (2007), on the other hand, argues that "the concept of a once-heard task could become less relevant ... given changes in technology which mean that people can generally listen to online materials, e.g. radio programmes, as many times as they wish."

Buck (2001), after weighing the arguments, simply points out that "...playing the text a second time may significantly change the nature of the listening construct."

In relation to the "second chance" argument it is perhaps worth considering what actually happens in real life. In situations in which one can replay all or part of a mechanical recording – such as a recorded telephone message, podcast or a DVD – one is not restricted to *two* listenings but can generally listen as many times as one likes.

In those situations where one can ask an interlocutor to repeat, such as in normal conversation, one rarely hears an exact repetition. The author conducted an informal study to test this hypothesis by requesting a random sample of twenty concordance lines the word "Pardon?" in the spoken sub-corpus of the British National Corpus (BNC). In no case was there a complete verbatim repetition of the previous utterance. In only one case was there a repetition of a single grammatical clause. Five responses contained two or more words from the previous utterance, and remainder contained only one repeated word or none at all. In most cases speakers give only a partial reproduction or a paraphrase of what they previously said.

It is not clear, then, that allowing test takers to hear a listening text twice (and only twice) is true to "real life". Moreover, it can be argued that systematically

playing every utterance twice makes for conditions which are very far from authentic.

1.2 Arguments from constraints of the testing situation

These arguments centre on the supposed need for repetition as a way of compensating for some aspect of the testing situation that places the test taker at a disadvantage relative to real life language use, such as:

- absence, in the testing situation, of some support which is normally available in real life, such as visual cues or an understanding of the context.
- presence in the testing situation of some obstacle, such as the cognitive load imposed by some task types.
- variable conditions of presentation, such as sound quality or background noise.

Geranpayeh and Taylor (2008) invoke all three of these arguments and offer a fourth: unexpected noise: "...unexpected noise may occur at any moment during the listening test (e.g. due to road/air traffic, building works, or even a candidate coughing); this can be intrusive and/or disruptive and risks impacting on candidate performance." This is arguably a reason for repeating the affected part of the text, but it hardly justifies playing the whole text twice.

Boroughs (2002), citing John Field, advances similar arguments and adds that test takers need time to adjust to different voices. Quite apart from the fact that we frequently have to respond quickly to unfamiliar voices in real life, it is not how this justifies repeating every listening passage.

1.3 Arguments from practicality/economy

These favour listening once. A test in which every audio text is played twice is bound to take much longer to administer (nearly twice as long, in fact) than a once-heard test. As the time available for testing is inevitably limited, a once heard test can include more items, which makes for higher reliability. Fortune makes this point: "...given the limited time involved in any particular listening test, the use of twice-heard texts only, precludes the use of a greater variety of texts and text types that might increase the sampling and coverage of the construct, and thus the generalisability of the test score produced."

1.4 Arguments from tradition

These maintain that it is best to keep to the conditions that stakeholders are used to unless there is a compelling reason to change. Geranpayeh and Taylor argue, for example, that "A test's origins or 'heritage' also understandably shape its design." This is stated as a reason why listening passages are played twice in Cambridge ESOL First Certificate and Proficiency exams; the practice has been carried over from former times when examiners used to read listening passages aloud (twice).

2. Empirical studies

Such empirical research as has been reported on the once/twice issue has mostly been a secondary aspect of some larger investigation. Some studies have considered repeated listening as a way of facilitating comprehension in the classroom as rather than in testing (Berne, 1995; Cervantes & Gainer, 1992; Chang & Read, 2006; Dupuy, 1999; Lund, 1991). Boroughs (2002) examined the differences between once-heard and twice heard tasks in the course of revising the listening section of the Cambridge CPE examination. Brindley and Slayter (2002) and Sherman (1997) considered number of listenings as one among several factors that influence task difficulty.

In all of the above cases repeated listening was found to reduce difficulty. Otsuka

(2004), in one of the few studies that focussed solely on the effects of repeated listening, found that difficulty actually *increased* slightly with repeated listening, but this was with a small sample (N = 38). In a follow-up study (N = 169) he found that difficulty decreased significantly with repeated listening.

Lund, Otsuka and Chang & Read found that repeated listening benefitted high ability subjects more than low ability ones. None of the other studies reviewed here mentions any interaction with person ability or item difficulty.

Regarding item type, Cervantes & Gainer found that repeated listening had more effect with "top down" (gist) tasks than with "bottom up" (detail) tasks. This would seem to accord with Borroughs, who found a stronger effect with summary completion tasks than with discrete multiple choice items.

Few studies report results relating to item properties other than difficulty, such as reliability or discrimination. Otsuka found reliability decreased with repeated listening in his first study (with a small sample, see above), whereas it increased in his second, larger study (but was very low under both conditions). He found no significant effect on item discrimination in either study. Borroughs reports point-biserial statistics for his CPE comparisons, showing slightly higher discrimination for twice-heard tasks, however it is not clear how many subjects were involved in the comparison, nor is it reported whether these differences are significant.

Fortune (2004) focussed exclusively on the contrast between once-heard and twice heard tasks, measuring discrimination as well as difficulty, and with a very useful discussion of both the previous literature and the theoretical arguments. His findings (N = 62) are, in summary:

- Once-heard tasks were between 10% and 14% more difficult than twice-heard ones.
- Most once-heard items showed higher discrimination than their twice-heard equivalents.
- There was no significant interaction between number of exposures and item type (although according to Fortune this may be due to limitations of the study).

3. The need for further research

Research so far supports the unsurprising conclusion that repeated listening tends to make comprehension tasks easier. However, this finding by itself does not serve to inform decisions about test design. Arguments that claim greater authenticity, and hence validity, for either once-heard or twice heard tasks on theoretical grounds are generally inconclusive. Evidence from some studies suggests that repeated listening affects other psychometric properties, notably reliability (Otsuka, 2004) and item discrimination (Borroughs, 2002; Fortune, 2004), and that its effects vary with test taker ability (Lund, 1991; Otsuka, 2004; Chang & Read, 2006) and item type (Cervantes & Gainer, 1992). The importance of these considerations test design, and paucity (and relatively small scale) of the relevant studies, make further research into these issues highly desirable.

References

- Berne, Jane E. (1995). How Does Varying Pre-listening Activities Affect Second Language Listening Comprehension? *Hispania*, **78**(2).
- Borroughs, R. (2002). Authenticity in listening texts. In Cyril Weir & M Milanovic (Eds.), *Continuity and Innovation: Revising the Cambridge Proficiency in English Examination 1913-2002*. Cambridge: CUP.

- Brindley, Geoff, & Slayter, Helen (2002). Exploring task difficulty in ESL listening assessment. *Language Testing*, 19(4), 369.
- Buck, G (2001). *Assessing listening*: Cambridge Univ Pr.
- Cervantes, R, & Gainer, G (1992). The effects of syntactic simplification and repetition on listening comprehension. *TESOL Quarterly*, 26(4), 767-770.
- Chang, ACS, & Read, J (2006). The effects of listening support on the listening performance of EFL learners. *TESOL Quarterly*, 40(2), 375-397.
- Dupuy, B. C. (1999). Narrow listening: An alternative way to develop and enhance listening comprehension in students of French as a foreign language. *SYSTEM*, 27(3), 351-361.
- Fortune, Alistair (2004). Testing listening comprehension in a foreign language - does the number of times a text is heard affect performance?. Unpublished M.A. dissertation.
- Geranpayeh, Ardeshir, & Taylor, Lynda (2008). *Examining Listening: developments and issues in assessing second language listening*.
- Lund, RJ (1991). A comparison of second language listening and reading comprehension. *Modern Language Journal*, 75(2), 196-204.
- Murray, Steve (2007). *Reviewing the CAE Listening test*.
- Otsuka, Ken'ichi (2004). *How to determine the optimal number of listening opportunities for listening comprehension tests among Japanese high school learners of English* Paper presented at the 9th Conference of Pan-Pacific Association of Applied Linguistics. from proceedings/PAAL9/pdf/OtsukaKe.pdf
- Sherman, Jane (1997). The Effect of Question Preview in Listening Comprehension Tests. *Language Testing*, 14(2), 185-213.