

# 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, London, UK  
Glyn.jones@pearson.com

July 2011

## 1. Introduction

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. Considering that this issue constitutes a fundamental design decision in the course of development of any listening test, surprisingly little has been written about it.

The instances in which the issue is touched on in previous literature fall into two categories: arguments supporting one position or another (listen once or listen twice) on theoretical grounds, and reports of experimental studies.

## 2. Theoretical Arguments

Theoretical arguments in support of either position, in turn, fall into four broad categories.

### 2.1 Arguments from authenticity

Interestingly, arguments from authenticity are advanced for both once and twice; that is, they can take the form:

"In real life you only get one chance to hear something". Fortune (2004), for example, states: "In virtually all real-world listening situations we hear the text only once."

OR

"In real life you can always ask someone to repeat, or play the recording again."

Murray (2007), for example, writing about the revision of the Cambridge ESOL CAE exam, argues that "the concept of a once-heard task could become less relevant in the CAE context, 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."

To consider each of the "second chance" scenarios in turn:

In those where one can ask an interlocutor to repeat, one rarely hears a verbatim reproduction of the previous utterance. In a randomly selected sample of twenty occurrences of the word "Pardon?" in the spoken sub-corpus of the British

National Corpus (BNC), only one is followed by a repetition of as much a single grammatical clause from the previous utterance. A further five contain two or more words from the previous utterance, while the remainder contain only one repeated word or none at all. In short, when prompted to repeat, speakers generally provide only a partial reproduction of their previous utterance or else a paraphrase.

In situations - becoming increasingly common with advances in technology - in which one can listen again to all or part of a mechanical recording – such as recorded telephone message, a podcast or a DVD - one is not restricted to two listenings but can generally listen as many times as one likes.

In short, it is by no means clear that allowing test takers to hear a listening text twice (and only twice) is an authentic representation of real life “second chance” situations. Besides, a language learner can hardly be said to be at a given level in listening ability if in real life they always need to ask for repetition when confronted with spoken language at that level.

## 2.2 Arguments from constraints of the testing situation

These arguments also invoke a comparison between the testing situation and real life, but rather than being concerned with repetition (or its absence) as a reproducible feature of real life listening they focus 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 in relation to real live language use, such as:

- absence, in the testing situation, of some support available in real life, such as visual cues or prior understanding of the context of situation.
- presence in the testing situation of some obstacle not present in real life, such as the cognitive load imposed by some task types or the competing demands on attention made by having to record answers on the test paper while the audio continues to play.
- variable conditions of test presentation, such as sound quality and acoustics, or intrusive background noise.

However, it is not obvious how repeated listening can correct for any of these factors, other than by making the task easier. It certainly does not make the testing situation any more like real life.

Geranpayeh and Taylor (2008) advance all three of these arguments and offer a fourth: the possibility that a sudden extraneous noise will obscure part of the listening text: “...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 true; however, while it is arguably a reason for repeating the affected part of the text (the missed phrase or sentence, say) it hardly justifies playing the whole test twice.

Boroughs (2002), citing remarks made by John Field at a seminar organised by Cambridge ESOL, advances similar arguments and adds the consideration that test takers need time to adjust to different voices. This may be an argument for including a preamble in which the recorded voice or voices are heard without any requirement to demonstrate comprehension, but not for playing everything twice. Besides, in authentic listening situations language users frequently do have to respond quickly to unfamiliar voices.

### 2.3 Arguments from practicality

If you don't repeat the audio text you can get twice as many items in the same time, thus making for a more reliable test. This is not a trivial consideration, given that the time available for testing is inevitably restricted by the test takers' ability to sustain concentration as well as by institutional constraints.

### 2.4 Arguments from tradition

These run along the lines of "our stakeholders are accustomed to having the listening passages played twice so unless there is a pressing reason to change we prefer to keep it that way"). According to Geranpayeh and Taylor, for example, "A test's origins or 'heritage' also understandably shape its design." This is advanced as a reason why listening passages are heard twice in Cambridge ESOL First Certificate and Proficiency exams; they were read out twice in the past, when examiners used to read the texts aloud, and this practice has simply been carried over to the audio recordings used in the present day tests.

## 3. Empirical Studies

Such empirical research as has been reported on the issue has been as a by product of some larger project. Some studies have considered repeated listening as one of a number of possible strategies for facilitating comprehension in the classroom as distinct from testing contexts (Berne, 1995; Cervantes & Gainer, 1992; Chang & Read, 2006; Dupuy, 1999; Lund, 1991). Borroughs (2002) examined the differences between one-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 exposures as one of a number of factors affecting task difficulty.

In all of the above cases repeated listening was found to reduce difficulty, sometimes considerably so. In one of his studies Otsuka (2004), whose research was focussed solely on the effects of repeated listening, found that difficulty actually decreased slightly with repeated listening, but this was with small sample (N = 38), the mean score for both formats was very low with a strong floor effect. In a repeated study (N = 169) 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 so far 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.

Findings concerning properties other than item difficulty are very sparse indeed. Otsuka found reliability decreased with repeated listening in his first study (but see above for the limitations of this), whereas it increased (but was very low for both formats) in his second, larger study. Item discrimination was not significantly affected in either of his two studies. Borroughs reports point-biserial statistics for his CPE comparisons and these show slightly higher discrimination for twice-heard tasks, however it is not clear how many subjects were involved in the comparison between once-heard and twice-heard versions, nor is it reported whether these differences are significant. Otsuka claims that the standard error of measurement (SEM) was "positively affected" by repeated listening, but does not report the statistics. Berne, on the other hand, reports higher SEMs for twice heard tasks (but in conjunction with higher standard deviations).

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).

#### **4. 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-hear 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](http://proceedings/PAAL9/pdf/OtsukaKe.pdf)
- Sherman, Jane (1997). The Effect of Question Preview in Listening Comprehension Tests. *Language Testing*, *14*(2), 185-213.