A New Matrix for ELT Listening Testing

Sheila Thorn

www.thelisteningbusiness.com and sthorn@clara.net
Outline

1. Introduction
2. Two ELT Englishes
3. Construct validity
4. Cognitive validity
5. Reliability
6. My dealings with ELT exam boards
7. A new matrix for testing listening using authentic recordings
8. A possible way forward
9. Conclusion
1. Introduction: my background

Languages (fluent French, German and Swedish through authentic listening)

DoS (On being appointed, I insisted on the college getting a language lab so that we could regularly expose our students to authentic off-air recordings.)

ARELS/FCE (Having been prepared for the former ARELS exams by teachers using authentic off-air recordings, our students found FCE relatively easy.)

Hospital work (I mix with health professionals from all over the world who achieved good grades in international ELT exams, but who are astounded to discover they cannot understand anything anyone says when they first arrive.)

Current areas of work (ELT item writer, editor, materials writer, author)

My objective (less of this and more of this)
2. Two ELT Englishes

**Scripted coursebook English v. real world English:**
Listening texts are used to model new language (grammar and lexis) as opposed to teaching students how to cope with listening to the authentic spoken English they will encounter outside the classroom.

**Pronunciation:**
productive (taught very well) v. receptive (generally ignored in the ELT classroom – too messy)
3. Construct validity

The degree to which a test measures what it claims, or purports, to be measuring.

Part 1: The recordings in ELT exams

a) performed by actors who are told to speak slowly and articulate clearly
b) recorded in a studio with no background noise
c) uniform speech rate – artificially slow with no speeding up and slowing down, unlike in real life
3. Construct validity

The degree to which a test measures what it claims, or purports, to be measuring.

Part 2: The content

a) scripted and carefully graded in terms of grammar and lexis
b) transactional v. interactional (informational rather than social)
c) convoluted v. direct
d) limited repetition and redundancy
4. Cognitive validity

The extent to which candidates use the same processes as native speakers when listening.

a) Not representative of how we listen as NESs
b) Need to test listening using other skills, but this often results in a huge reading/memory load
c) Tasks are largely unrepresentative
d) Constraints of computer marking – have to have definitive answers and no open-ended questions
e) Synonyms are used widely in listening tests, but these test language ability rather than listening ability
To sum up the current situation

a) We use simplified and graded recordings with difficult items → I would like to see this reversed, so that we use authentic recordings with less challenging items.

b) There is a negative washback effect on ELT coursebooks and classrooms because students do not need to be exposed to authentic listening to pass their exams.

c) Candidates perform well in official ELT exams, but not in real-life listening situations.
5. Reliability of mainstream ELT listening exams

Fantastic! But...
6. My dealings with exam boards

1. Permissions comes up again and again as an issue because exam boards would have to pay a fortune to use authentic off-air recordings.

2. Knee-jerk reaction when authentic listening texts are mentioned: ‘impossible’, ‘too difficult’, ‘unreliable’.

3. In my experience, item writers tend to produce listening texts that they know will work rather than experiment with something more authentic-sounding and risk losing their fee.

4. Instructions to item writers (we are told to only use grammar and lexis at the students’ productive level, according to the CEFR)
7. **A new matrix for listening testing using authentic recordings** © Sheila Thorn, The Listening Business

1. Speed of delivery (sps) and prevalence of streamlining features
2. Frequency/Familiarity of lexis
3. Frequency/Familiarity of grammatical structures
4. Number of speakers
5. Concrete → abstract
6. Familiarity with the topic
7. Formality (written English read aloud → spontaneous)
8. Background noise
9. Length of pauses and number of pauses
10. Neutral, standard accent → varieties of NS/NNS accents

NB I have deliberately not included length in the list. Exam boards tend to believe that increased length = increased difficulty, when in fact the reverse is often true, due to redundancy and repetition.
7. Matrix

Using the matrix overleaf, award points, so we have 10 categories graded 1 (easy) to 4 (difficult), with a top score of 40.

10–15 A2 (KET)
15–20 B1 (PET)
20–30 B2 (FCE)
30–35 C1 (CAE)
35–40 C2 (CPE)
8. A possible way forward

a) Use a mixture of scripted/spontaneous texts

b) Issue guidance to actors to speak more naturally and vary their speed of delivery

c) At least place more emphasis on authentic-sounding texts if 100% authentic is too problematic

d) Use a mixture of transactional and interactional texts

e) Tap into listening as a sense – more use of eavesdropping, voicemails and one-sided phone calls

f) Insist item writers use authentic recordings as their sources rather than written texts, as at present

g) Invite item writers to supervise recording sessions

h) Insert longer pauses at discourse boundaries to give candidates more processing time

i) Allow repetition and redundancy, as used in the real world
9. In conclusion

We should be striving after increased content and cognitive validity in our listening tests, while maintaining reliability.

This would lead to a positive washback effect in ELT classrooms, as can be seen from the following quote:

‘If a high-stakes test uses unscripted spoken texts for L2 listening assessment, then it is more likely that curriculum planners, materials developers and classroom teachers will likewise use unscripted spoken texts in their materials for L2 leaners.’ (Wagner and Toth, 2017, 78)

It’s happening already. Exam boards in Germany, Austria and The Netherlands use authentic listening texts in their ELT exams.
Useful references

Sheila Thorn  *Mining listening texts*  Modern English Teacher Vol. 18  No. 2 2009

Sheila Thorn  *Real Lives, Real Listening*  Collins 2013

Sheila Thorn  *Debunking authentic listening*  Modern English Teacher Vol. 21 No. 2 2012

Sheila Thorn  *Listening Testing: time for a rethink?*  Modern English Teacher November 2016


http://research-information.bristol.ac.uk/files/96659318/ Isaacs_TrofimovichOA.pdf
Useful references


Richard Cauldwell (2002) *Grasping the nettle: the importance of perception in listening comprehension*
http://www.developingteachers.com/articles_tchtraining/perception1_richard.htm

Richard Cauldwell *Phonology for Listening – Teaching the stream of speech* 2013

Richard Cauldwell *A Syllabus for Listening – Decoding* 2018

John Field *Listening in the Language Classroom* CUP 2008

Emma Rodero (2012) *A comparative analysis of speech rate and perception in radio bulletins*