

## **EAP LISTENING ASSESSMENT**

### **Academic Listening as a discrete skill – is this valid and how can it be assessed in an examination?**

#### **Introduction**

The two main problems that most international students face when embarking on a course of study in the UK are reading and listening. With Listening, the greatest challenge the students face is the academic lecture – the major medium of instruction at universities across all disciplines. All the standard academic English exams seek to address this in some fashion, despite the fact in most cases, high school graduates taking these exams will have no experience or preparation for such a task.

On the other side of the tracks, there are many pre-sessional and direct entry courses designed specifically to teach integrated skill sets, but are put under pressure by universities who want to see IELTS style 4-skill grades. Coursework will rarely assess listening in isolation but will look at how that skill informs overall performance and will encourage students to see it in the same way, and yet when it comes to summative assessment, it is still necessary to produce grades for listening in isolation.

This paper will examine some common factors arising from academic listening taxonomies, and how these are addressed in current assessment design. In particular, it will examine the academic lecture and how this is presented in four exams, the paper-based IELTS and TEEP, and the computer-based TOEFL iBT and Pearson Academic. The factors of face validity and construct validity will be examined in each case, looking at the question of authenticity vs. reliability; and whether Academic Listening and lecture comprehension in particular should be examined in this fashion, or whether greater reliability could be obtained through the assessment of micro-skills, removed from the construct.

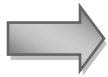
In conclusion, the paper offers no solutions but is written to promote further debate and research in this area, and hopes to inform teaching practice and classroom assessment design.

#### **Academic Listening taxonomy – the Listening process**

The first and most widely referred to taxonomy of Academic Listening was done by Richards (1983, cited in Buck, 2001, Young, 1994 and others), and is given below.

1. Ability to identify purpose and scope of lecture
2. Ability to identify topic of lecture and follow topic development
3. Ability to identify relationships among units within discourse (e.g. major ideas, generalisations, hypotheses, supporting ideas, examples)
4. Ability to identify role of discourse markers in signalling structure of a lecture (e.g. conjunctions, adverbs, gambits, routines)
5. Ability to infer relationships (cause and effect, conclusion)
6. Ability to recognise key lexical items related to subject/topic
7. Ability to deduce meanings of words from context
8. Ability to recognise markers of cohesion
9. Ability to recognise function of intonation to signal information structure (e.g. pitch, volume, pace, key)
10. Ability to detect attitude of speaker towards subject matter
11. Ability to follow different modes of lecturing: spoken, audio, audio-visual
12. Ability to follow lecture despite differences in accent and speed

13. Familiarity with different styles of lecturing: formal, conversational, read, unplanned
14. Familiarity with different registers: written versus colloquial
15. Ability to recognise relevant matter: jokes, digressions, meanderings
16. Ability to recognise function of non-verbal cues as markers of emphasis and attitude
17. Knowledge of classroom conventions
18. Ability to recognise instructional/learner tasks (e.g. warnings, suggestions, recommendations, advice, instructions)



**Q1 – What commonly taught listening skill is not represented above?**

### **Key features of the lecture**

Looking at this list, we can also glean some of the features that make an authentic lecture different from a written text (compiled from Lebauer, 1984, Young, 1994 and Flowerdew, 1994):

- Rhetorical structure of monological discourse (classical, problem-centred, sequential, comparative, thesis, cause-and-effect) – mostly genre dependent
- Phasal structure of discourse - sentential level and up(not dependent on genre) (Young, 1994)
- Meta-talk to indicate overall structure (including evaluative function markers *I mean, for example, in fact*)
- Use of micro-level discourse markers (*okay, and, so, now*)
- False starts, redundancies, repeats
- Thematic redundancies to indicate structure and to control pace of information flow
- Use of asides to indicate importance/intention of speaker
- Rhetorical questions
- Active verbs
- Use of personal pronouns and *it*
- Empathy (*we, let's*) as compensation for lack of direct interaction and to counter affect
- Checking
- Short phrases punctuated by pause fillers – idea units (Hansen and Jensen, 1994) of 11 words, single intonation contour followed by pause
- Use of extended pauses to allow information ordering for speaker/listener
- Contractions
- Not lexically dense (low verb frequency, low syllable/minute count, limited use of tenses)

Considering both lists, it is self-evident that the use of scripted texts as are found in some text-books or assessments serves little purpose, as in Flowerdew (1994);

It is clear from the above discussion of lexico-grammatical features of lectures that there is a need for the use of authentic lectures...scripted texts are obviously likely to present to learners a type of language that is not authentic (p.20).

Hansen and Jensen (1994) addressed the issue of assessment in particular, and whilst acknowledging that academic lectures shared more features with written discourse than normal spoken discourse, after completing such an analysis as that outlined above, concluded;

...in order to separate statistically test takers' listening comprehension skill from their reading comprehension skill on a language proficiency test, the listening discourse must have the features of oral discourse listed above. To present a lecture that has been scripted and read aloud in a listening comprehension teaching or testing situation does a disservice to the students. This is not the type of material that they will have to grapple with when they attend a lecture (p.246).

### The question of validity

Following this argument, and also making reference to Bachman and Palmer's (1986) emphasis on validity of the construct (where an examination element comes as close as possible to the Target Language Use), to be valid, any assessment of a student's ability to comprehend a lecture should use an authentic lecture and the task should reflect authentic usage. Here it is necessary to draw a distinction between these terms, *face validity*, *content validity* and *construct validity*.

**Content validity** – 'the degree to which the test accurately samples from the EAP course of study and/or some future study domain' (Fulcher 1999, p.222)


**Construct validity** – how the closely the construct – the content and task – relates to the Target Language Use, which in turn indicates greater interpretive reliability of the test scores (Bachman and Palmer, 1990; Messick, 1996)

**Face Validity** – the degree to which the test 'looks authentic' (Fulcher, 1999)

Test designers should be keen to make their tests as authentic and valid as possible. Face validity is also an important factor in student motivation (Fulcher, 1999), and the more '*unified*' validity a test has, the greater the positive washback effect that test will have on the students and teaching practice (Messick, 1996). However, with authenticity, a compromise must be reached between using texts that can also produce enough raw material on which elements can be based and that are reliable and practical for use.

In the main standardised tests examined here, IELTS, TOEFL, Pearson Academic and the TEEP (see attached), face validity seems to be employed mainly as a marketing tool and to make up for the use of inauthentic texts or elements which lack construct validity. The TEEP is the closest it can get to construct validity, having a wealth of source material to assist comprehension, but still it is not authentic as this is not practically possible – and it does come back to the application of integrated skill sets, as the tasks require the production of written answers and for the information to be processed.

Bearing this in mind, the TOEFL is the only test that examines the student's ability to apply meta-linguistic analysis to the text, to explicitly identify discourse markers, attitude and inferential markers, and the purpose of asides and repeats. While this appears to move further away from construct validity, it acknowledges the purpose of the text in this context – an exam – and in this way, by explicitly examining micro-skills outside of the pretence of the construct, does this make the test more valid in its own right?

 **Q2 – If we teach skills explicitly, for students to consciously apply these skills, is it not valid to assess them explicitly?**

Such a test could arguably be no longer a test of pure language ability – but does assess directly the knowledge of the students. If the student is expected to process the topical content of the text by

conscious application of these skills, then direct examination of the knowledge of these skills and how they apply to the text surely becomes a valid test of the ability of the student to comprehend that lecture – but now is no longer dependent on the understanding of the content.

If construct validity and true face validity is practically unattainable – would an examination of indirect skills, with the exam as the Target Language Domain, not be more valid and would not the results of that exam be more indicative of a student's 'listening' ability?

## References

- Alexander, O., Argent, S. and Jenifer Spencer (2008), *EAP Essentials – A teacher's guide to principles and practice*. Reading: Garnet Education.
- Bachman, L.F., Palmer A.S., (1986), *Language Testing in Practice*. Oxford: Oxford University Press.
- Buck, G. (2001), *Assessing Listening*. Cambridge: Cambridge University Press.
- Chaudron, C, Loschky, L and Janice Cook (1994), 'Second language listening comprehension and lecture note-taking' in Flowerdew, J. (ed.) (1994), *Academic Listening Research Perspectives*. Cambridge: Cambridge University Press.
- Flowerdew, J. (1994), 'Research of relevance to second language comprehension – an overview' in Flowerdew, J. (ed.) (1994), *Academic Listening Research Perspectives*. Cambridge: Cambridge University Press.
- Fulcher, G. (1999), 'Assessment in English for Academic Purposes: Putting Content Validity in Its Place' *Applied Linguistics* 20/2, pp. 221-236.
- Hansen, C., Jensen, C, 'Evaluating lecture comprehension', in Flowerdew, J. (ed.) (1994), *Academic Listening Research Perspectives*. Cambridge: Cambridge University Press.
- Lebauer, R.S., 'Using lecture transcripts in EAP lecture comprehension courses', *TESOL Quarterly*, Vol 18 No.1 pp.41-54.
- Messick, S. (1996), 'Validity and washback in language testing'. *Language Testing* 1996: 13, pp.241-256.
- Powers, D. E. (1986), 'Academic demands related to listening skills' *Language Testing* 1986;3;1 pp. 1-38.
- Weir, C.J., (2005), *Language Testing and Validation – An Evidence-Based Approach*. New York: Palgrave Macmillan.
- Young, L. (1994), 'University lectures – macro-structure and micro-features' in Flowerdew, J. (ed.) (1994), *Academic Listening Research Perspectives*. Cambridge: Cambridge University Press.