Re-engineering a speaking test used for university admissions purposes: considerations and constraints

The case of IELTS

Overview

- A bit of historical background
- The evolution of the speaking test component in ELTS and IELTS
- The burden of expectation laid on a test
- Considerations and constraints
- The future?

A bit of historical background…

- 1913 – Certificate of Proficiency in English (Cambridge)
- 1960s – EPTB, TOEFL
- 1980 – introduction of ELTS (4-yr devt)
  - reflecting new communicative paradigm in teaching/testing + trends in ESP/LSP
  - incl. face-to-face speaking test – Interview (M3)
  - based upon discussion of material in Source Booklet (6 academic study domains – LifeSc, SocSt, PhySc, Tech, Med and GenAc)

The ELTS Speaking Test (1980)

- 3 parts (approx.15 mins)
  - Part 1: general questions
  - Part 2: questions about Source Booklet text
  - Part 3: discussion of future plans
- Assessment
  - Holistic scale
  - Range-finding after Part 1 (3 out of 9 bands)
  - Final decision on band assignment after Part 3
  - Required qualified and trained EFL teachers for interview and rating

ELTS Revision Project (1986–89)

- Extensive consultation with test stakeholders (questionnaire, focus group)
- Review of various considerations and constraints
  - interviewer/rater recruitment
  - interviewer/rater training
  - length
  - effectiveness of training (self-access)
  - standardisation
  - reliability of rating (single marker)
  - choice of subject module
  - discipline specificity/familiarity
  - administrative complexity

Revision proposals in 1988

- Reduction to 3 (instead of 6) subject modules – PST, LMS and ASS/BSS
- Speaking Test moved from M3 to G3
- The need to balance considerations and constraints

Considerations and constraints

- Task design
- Band levels
- Discrimination
- Security
- Timing
- Interview conduct
- Assessment procedures
- Administration
- Interviewer/rater training

A speaking test…

- 'designed to measure general proficiency in speaking'
- 'to interact in social, survival and training or academic contexts without focusing specifically on technical or academic features of the language'
- *(Ingram and Wylie 1997)*

A more structured speaking test

- 5 phases (10–15 minutes)
  - Phase 1 (1–2') – *Introduction*
  - Phase 2 (3–4') – *Extended Discourse*
  - Phase 3 (3–4') – *Elicitation*
  - Phase 4 (3–4') – *Speculation and Attitudes*
  - Phase 5 (1') – *Conclusion*

Other QA measures from 1989

- New global, holistic scale
- New examiner training package (face-to-face and self-access)
- Standardisation video
- Certification process for examiners
- Sample monitoring process (recording)
- A targeted research and validation agenda

IELTS 1989

- Designed to assess the language ability of candidates seeking to study in English-medium universities
- Widely recognised as an English language proficiency requirement for entry
- Testing ability in 4 skills – reading, writing, listening, speaking
- Major revision to Reading, Writing and Listening modules undertaken in 1995 (but not to the Speaking test)

Why not the Speaking Test?

- No clear consensus by 1995 on best approach – 1989 test still 'bedding down'
- Various research studies in progress – findings still awaited
- Large-scale revision of a face-to-face speaking test is complex
  - Test content and format
  - Retraining examiners
  - Re-engineering delivery systems
- Issues of resourcing and sustainability

IELTS Speaking Revision Project 1998–2001

- Sources for identifying key issues :
  - Review of routinely collected candidate score data and test performance data for operational speaking test
  - Review of theoretical and empirical studies on the IELTS speaking test conducted between 1992 and 1998
  - Review of other research into speaking assessment, esp work on the Cambridge speaking tests (e.g. Lazaraton)

Key issues

- Test format
- Task design
- Rating scales
- Examiner/rater behaviour
- Test monitoring
- Test validation

The 'continuity' imperative

- Face-to-face, one-on-one format, audio-recorded
- Retaining 9 performance bands
- Overall test length (max 15 mins)
- Multi-phase approach

The 'change' imperative (1)

- Developing a clearer specification of tasks, i.e. in terms of input and expected output (and revising tasks where necessary, esp for Phases 3 and 4)
- Redeveloping the assessment criteria and rating scale to ensure that descriptors matched more closely the output from the candidates in relation to the specified tasks

The 'change' imperative (2)

- Introducing an examiner frame to guide examiner language and behaviour and so increase standardisation of test management
- Retraining and restandardising a community of around 1,500 IELTS examiners worldwide, using a face-to-face approach
- Introducing sustainable quality assurance procedures for this global cadre

Test format and task design

- Underlying construct of spoken language proficiency retained (cognitive view of competence, e.g. Garman 1990, Levelt 1989)
- Resolving problems with Phases 2-4 – replacing the earlier 5 phases with 3
  - Part 1: Introduction and interview
  - Part 2: Individual long turn
  - Part 3: Two-way discussion
- Length: 11-14 mins

Assessment criteria and rating scales

- Global/holistic scales deconstructed into **analytical** criteria and scales: *Fluency and Coherence; Lexical Resource; Grammatical Range and Accuracy; Pronunciation*.
- A triangulated approach to developing criteria and rating scales to ensure theoretical relevance, discriminability and practicability

Examiner conduct and assessment

- Introduction of an interlocutor frame
- New induction and training programme
- New certification programme
- A cascade model for retraining/recertification
- Continuation of sample monitoring

Ongoing validation research...

- To explore/compare task functioning
- To understand examiner conduct and assessment
- To examine criteria/scale functioning
- The key role of both quantitative and qualitative methodologies, esp corpus analysis, discourse analysis, conversation analysis and VPA

Useful publications

- *IELTS Collected Papers* (ed. Taylor & Falvey 2007)
- *Assessing Academic English* (Davies 2008)
- *IELTS Research Reports, Vols 1-13* (available online via the IELTS website)
- Cambridge's *Research Notes* (since 2000)
- Various journal articles and book chapters

The burden of expectation laid on a test score

- What does a score on an IELTS Speaking test mean?
- What do we sometimes expect (or hope) a score from the IELTS Speaking test will tell us?
- How informed and realistic are our expectations about a test score?

The future?

- What might a different (better?) sort of Speaking Test used for academic purposes look like?
- Alternative solutions/approaches:
  - iBTOEFL
  - Pearson Test of English (Academic)

Considerations and constraints

- Task design
- Band levels
- Discrimination
- Security
- Timing
- Interview conduct
- Assessment procedures
- Administration
- Interviewer/rater training

The future?

- What might a different sort of Speaking Test designed for academic purposes look like?
- Answers on a postcard, please.....!