

Assessing Speaking

Professor Rebecca Hughes

Centre for English Language Education,
University of Nottingham, UK

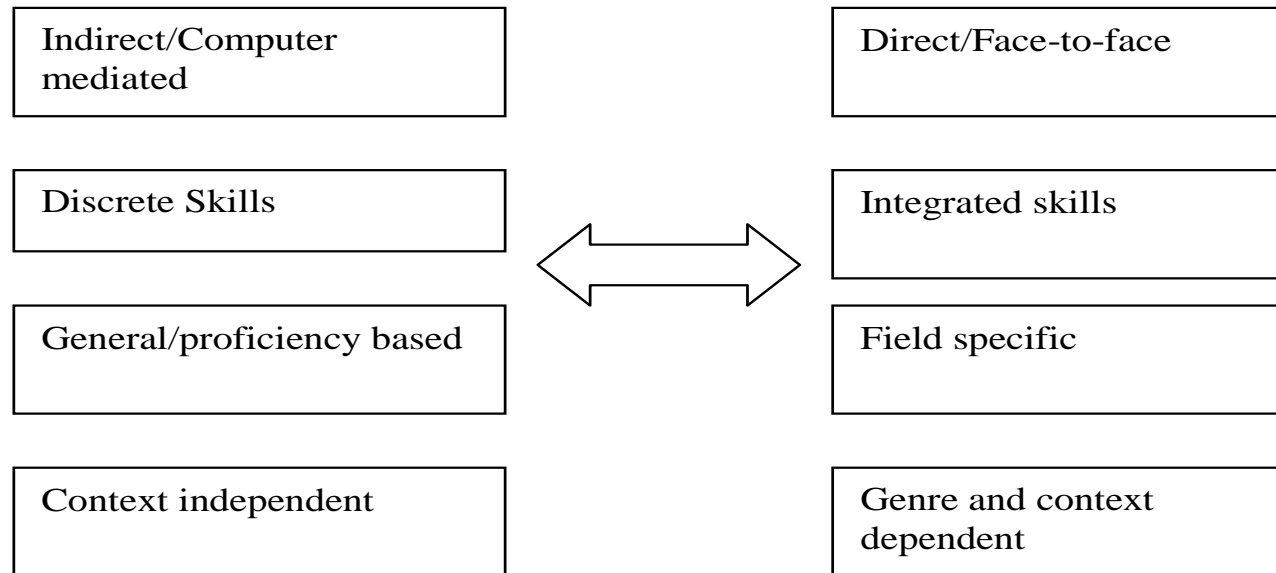
Overview and Introduction

- Preliminary comments on language testing
- Understanding the construct: The nature of speaking versus writing
- Test formats, task types, commonly used criteria/descriptors
- Issues to bear in mind and new directions

Preliminaries

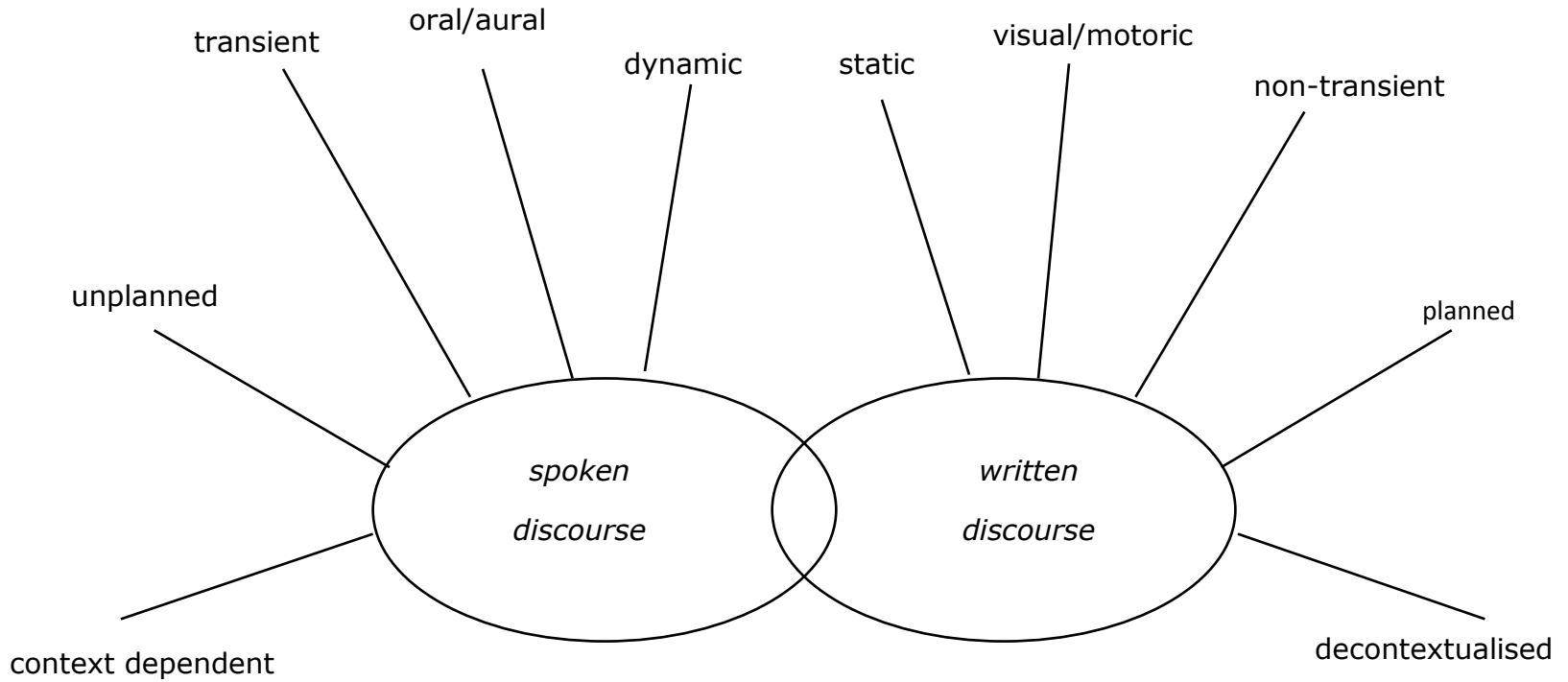
- Language assessment is not a perfect tool
- Language assessment is a socially embedded practice with repercussions for test takers
- All language tests are positioned in relation to the issue of score interpretation:
 - How does test performance relate to actual performance in a range of contexts?
 - What do the test designers claim for the test in this regard?

Contrasting approaches in oral testing



Hughes (forthcoming)

Aspects of Production: speech versus writing



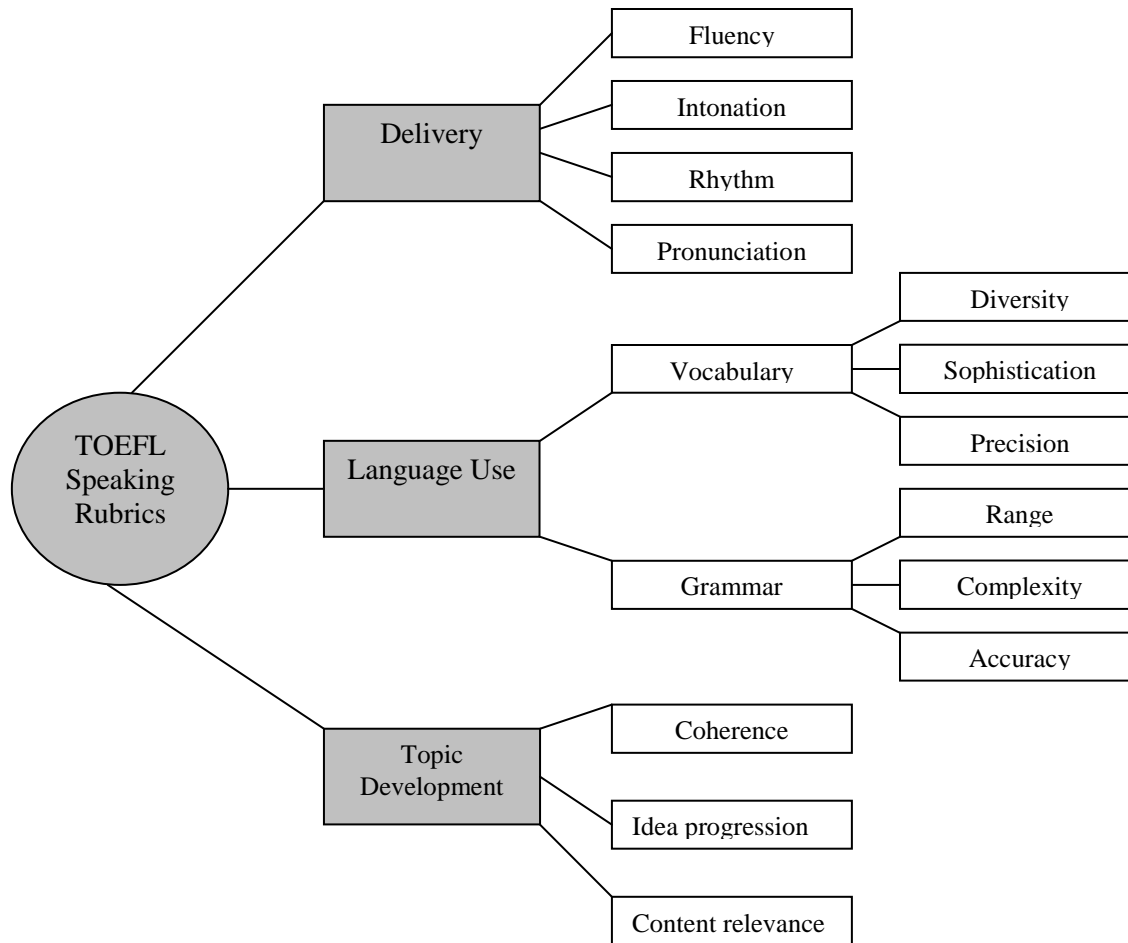
Assessing speaking objectively is a challenge!

- Attitudes to language proficiency – matching assessment criteria to norms of speech (or not)
- Pragmatic/cultural/conversational norms of first and target language – for instance, turn-taking, tolerance of silence
- Practical challenges of test delivery, examiner training, standardisation, and the examiner/candidate interface
- Maintaining objectivity and transparency, particularly in ‘high stakes’ testing – eg is the system robust enough to deal with candidate appeals?

How test developers (in ELT) have responded to these challenges

- Two tendencies:
 - Orienting towards whole performance
 - Explaining validity in terms of authentic interaction in context (more skills based)
 - Influence of test conditions taken into account
 - Orienting towards discrete facets
 - Explaining validity in terms of objectivity (more proficiency based)
 - Influence of test conditions abstracted away/minimised

Example of discrete facet approach



Xi et al (2008:29)

Test delivery formats for speaking (each have pros and cons)

- Non-present interlocutor – recorded performance, anonymous rater (eg Internet based TOEFL test, ‘iBT’)
- One-to-one interview – present interlocutor who is also examiner (eg IELTS)
- One-to-one interview – present interlocutor, and a separate examiner/rater (eg Studieprøven)
- Paired/small groups of candidates – generally with one interlocutor and one examiner/rater (eg Prøve i Dansk 1, 2, and 3)

Task types for oral assessment

- Repetition or reading aloud
- Guided narrative – visual prompts
- Guided description – visual prompts
- Guided functional language (eg giving directions) – mixed prompts, often visual
- Guided monologue (eg giving an opinion on...) – mixed prompts, often text/question based
- Structured/semi-structured interaction (eg interviewer led exchange of ideas)
- Open interaction (eg two candidates discuss a topic)

Commonly found criteria in oral assessment frameworks

- Negative (pronunciation/delivery facet)
 - Repetition
 - Hesitation
 - Unintelligible
 - Pauses (long)
 - Pauses (frequent)
 - Self-correction
 - ‘Finding words’
 - ‘Slips’
 - Lapses (in completeness)
 - Lapses (pronunciation)
 - Difficulties (pronunciation)

- Positive (pronunciation/delivery facet)
 - Fluent
 - Effortless to understand
 - Intelligible
 - Fluid
 - Well paced
 - Clear (production)
 - Automatic

Processing demands and performance: a UK Anglo-phone student speaks...

*The plan was always...[=pause]. The plan was never to let
go but*

*to er assimilate em the Africans into er sort of wider France
[inaudible] or whatever*

*was to er assimilate the Africans into a larger France, never
to sort of...like Britain.*

*The plan was, I mean they saw an end to the road,
they never said they would hold on to it.*

(Carter and McCarthy, 1997: 137)

- ...[O]ne of the central characteristics of naturally occurring conversation...is that language users are largely unaware of how conversation is typically structured and managed. When asked to articulate conversational practices, native-speaker pronouncements are often at odds with what speakers actually do (Wolfson, 1989). Much of how everyday conversation works is so deceptively familiar that people studying and testing language often overlook fundamental characteristics of conversation.

(Johnson and Tyler, 1998: 27)

Speaking under test conditions: Three main considerations

- Effects of cognitive load and speech processing
 - what is possible, what is normal?
 - does the test value performance that is really only possible in written mode/high planning contexts?
- Intercultural and linguistic background(s)
 - polite conversational behaviour in one language may transfer badly to another
- Inter-personal and interpretive factors
 - positive and negative effects of gender, cultural stereotypes, personality type and so on
 - effects of one individual on another (examiner-candidate, and candidate-candidate)

Emerging and new directions in assessing speaking: Skills/Performance based strands

- Socio-phonetic studies on interface between rater judgements and preconceptions of speakers
- Ethically oriented assessment (eg creation of 'indigenous' rather than 'expert/outsider' criteria/descriptors; focus on role of participants in assessment processes)
- Stronger influence of other applied linguistic disciplines (esp. Conversation Analysis/Corpus Linguistics/Pragmatics)
- Intelligibility versus accuracy debates will continue (particularly in English language testing alongside the English as a *Lingua Franca*/International Englishes debates)

Emerging and new directions in assessing speaking: Proficiency based strands

- Speech recognition systems and automated rating
- On-line delivery of speaking tests and greater reliance on software to assess performance
- Some convergence with eg interactional linguistics/CA
- Continued work on the relationships between task type, skills, performance and scoring

Conclusions

- The nature of speech provides a challenge for language testing
- The complex matrix of skills and knowledge that underlies speech is hard to distil into a single score, or even a set of scores
- The demands of objectivity and validity need to be balanced against authentic contextualised speech performance
- High-stakes assessment such as testing proficiency for residency needs to strike this balance particularly thoughtfully
- An interesting time in assessment of speaking as insights about the norms of speech feed into the test-development forum

References

- Carter, R. and McCarthy, M. (1997). *Exploring Spoken English*. Cambridge: Cambridge University Press.
- Fulcher, G. (2003). *Testing Second Language Speaking*. London: Pearson Longman.
- Hughes, R. (1996). *English in Speech and Writing: investigating language and literature*. London: Routledge.
- Hughes, R. (forthcoming). *Teaching and Researching Speaking* (2nd Ed.). London: Longman.
- Johnson, M. and Tyler, A. (1998). Re-analysing the OPI: how much does it look like natural conversation? In R. Young and A. Weyin He (Eds), *Talking and Testing: Discourse approaches to the assessment of oral proficiency*. Studies in Bilingualism. 14: 27–51. Amsterdam and Philadelphia: John Benjamins Publishing Company.
- Xi, X., Higgins, D., Zechner, K. and Williamson, D. M. (2008). Automated Scoring of Spontaneous Speech Using SpeechRaterSM v1.0. *Educational Testing Services Research Report* ETS RR-08-62.