

INTRODUCTION

The British Council/ IELTS Australia joint-funded research programme makes a significant contribution to the ongoing development of IELTS. External studies funded by these two IELTS partners complement internal validation and research studies conducted or commissioned by the third IELTS partner, Cambridge ESOL. The funded studies form an integral part of the process of IELTS monitoring, validation and evaluation.

This volume brings together a number of important empirical studies focusing on the IELTS Speaking Test. A major review of the IELTS Speaking Test took place towards the late 1990s and a formal project to revise the Speaking module was conducted between 1998 and 2001. The revision project concentrated on several key areas with the aim of achieving greater standardisation of test conduct and improving the reliability of assessment; this included:

- developing a clearer specification of tasks, eg in terms of input and expected candidate output, and the revision of the tasks themselves for some phases of the Test
- introducing an examiner frame to guide examiner language and behaviour, and so increase standardisation of test management
- re-developing the assessment criteria and rating scale to ensure that the rating descriptors matched more closely the output from candidates in relation to the specified tasks
- re-training and re-standardising a community of around 1500 IELTS examiners worldwide using a face-to-face approach, and introducing ongoing quality assurance procedures for this global examiner cadre.

The revised IELTS Speaking Test was introduced in July 2001 and since that time the joint-funded programme has invited research proposals for empirical studies which explore various aspects of the revised test module along the dimensions listed above. Such studies are considered essential to confirm that the revised test is functioning as intended, to identify any issues that may need addressing, and to contribute to the body of evidence in support of the validity arguments underpinning use of the test.

The first study reported in this volume, by **Gillian Wigglesworth and Catherine Elder**, investigated the relationship between three variables in the IELTS Speaking Test – planning, proficiency and task. Their study aimed to increase our understanding of how these variables interact with one another and how they impact on test-taker performance. The specific focus was the role and use of the one minute of planning time afforded to candidates in Part 2 of the Speaking Test. Part 2 is a long turn task with in-built pre-task planning time. The task design reflects the fact that some speech – especially in academic and professional contexts – is more formal in nature and is often planned prior to delivery (though, as the researchers acknowledge, it is clearly difficult to replicate this condition within the limited time-frame of a speaking test). Early Second Language Acquisition (SLA) research into the effect of pre-task planning, including work by Wigglesworth, suggested that planning time impacted positively on both content and quality of L2 oral performance; later research findings, however, proved less conclusive. As this was an innovative feature of the revised

IELTS Speaking Test introduced in 2001, the test developers were keen to investigate the effectiveness and validity of the planning time and how test-takers make use of it.

Interestingly, Wigglesworth and Elder's experimental study found no evidence that the availability of planning time advantages or disadvantages candidates' performance, either in terms of the discourse they produce or the scores they receive. Despite this finding, the researchers recommend that one minute of pre-task planning should continue to be included on Part 2 in the interests of fairness and for face validity purposes. An important dimension of this study was that it canvassed the candidates' own perceptions of the planning time available to them; feedback from test-takers suggests they perceive the one minute as adequate and useful. This study therefore offers positive support for the decision by the IELTS test developers to include a small amount of planning time in the revised Speaking Test; it also confirms that there would be no value in increasing it to two minutes since this would be unlikely to produce any measurable gain. Another useful outcome from this study is the feedback from both researchers and test-takers on possible task factors relating to topic; this type of information is valuable for informing the test writing process.

Volume 6 includes two studies by **Annie Brown** who has a long association with the IELTS Speaking Test dating back to the early 1990s. Findings from Brown's studies of the Test as it was then (some of which formed the basis of her doctoral research) were instrumental in shaping the revised Speaking Test introduced in 2001. The first of the two studies in this volume examined the validity of the analytic rating scales used by IELTS examiners to assess test-takers' performance. When the Speaking Test was revised, a major change was the move from a single global assessment scale to a set of four analytic scales that all IELTS examiners worldwide are trained and standardised to administer. The IELTS partners were therefore keen to investigate how examiners are interpreting and applying the new criteria and scales, partly to confirm that they are functioning as intended, and also to highlight any issues that might need addressing in the future.

Brown's study used verbal report methodology to analyse examiners' cognitive processes when applying the scales to performance samples, together with a questionnaire probing the rating process further. The study's findings provided encouraging evidence that the revised assessment approach is a significant improvement over the pre-revision Speaking Test in which examiners were relatively unconstrained in their language and behaviour and used a single, holistic scale. Firstly, the revised test format has clearly reduced the extent to which an interviewer's language and behaviour is implicated in a test-taker's performance. Secondly, examiners in this study generally found the scales easy to interpret and apply, and they adhered closely to the descriptors when rating; they reported a high degree of 'comfort' in terms of both managing the interaction and awarding scores. The study was instructive in highlighting several aspects that may need further attention, including some potential overlap between certain analytic scales and some difficulty in differentiating across levels; however, Brown suggests that these can relatively easily be addressed through minor revisions to the descriptors and through examiner training.

Brown's second study in this volume is a partner to the first. It too sought empirical evidence to validate the new Speaking Test scale descriptors but through a discourse analytic study of test-taker performance rather than a focus on examiner attitudes and behaviour. Overall, the study findings confirmed that all the measures relating to each analytical criterion contribute in some way to the assessment on that scale and that no single measure appears to dominate the rating

process. As we would wish, a range of performance features contribute to the overall impression of a candidate's proficiency and the results of this study are therefore encouraging for the IELTS team who developed the revised scales and band descriptors.

This study also highlights the complexities that are involved in assessing speaking proficiency across a broad ability continuum (as is the case in IELTS). Specific aspects of performance may be more or less relevant at certain levels, and so contribute differentially to the scores awarded. Furthermore, even though two candidates may be assessed at the same level on a scale, their respective performances may display subtle differences on different dimensions of that trait. This reminds us that, at the level of the individual, the nature of spoken language performance and what it indicates about their proficiency level can be highly idiosyncratic in nature.

Barry O'Sullivan and Yang Lu set out to analyse the way in which the examiner script (or Interlocutor Frame) used in the IELTS Speaking Test impacted on the test-taker's performance, specifically in cases where an examiner deviates from the scripted guide provided. An Interlocutor Frame was introduced in the 2001 revision on grounds of fairness – to increase standardisation of the test and to reduce the risk of rater variability; since then, the functioning of the Interlocutor Frame has been the focus of ongoing research and validation work. The study reported here forms part of that research agenda, and aimed to locate specific sources of deviation, the nature of the deviations and their effect on the language of the candidates. Taking a discourse analytic approach, the researchers analysed transcription extracts from over 60 recordings of live speaking tests to investigate the nature and impact of examiner deviations from the interlocutor frame.

Findings from their study suggest that in Parts 1 and 2 of the Speaking Test, the examiners adhere closely to the Frame; any deviations are relatively rare and they occur at natural interactional boundaries with an essentially negligible effect on the language of candidates. Part 3 shows a different pattern of behaviour, with considerable variation across examiners in the paraphrased questions, though even here little impact on candidate language could be detected. It is important to note, however, that some variation is to be expected in this third part of the Test as it is specifically designed to offer the examiner flexibility in choosing and phrasing their questions, matching them to the level of the test-taker within the context of a test measuring across a broad proficiency continuum. Once again, findings from this study confirm that the revised Speaking Test is functioning largely as the developers originally intended. The study also provides useful insights which will inform procedures for IELTS examiner training and standardisation, and will shape future changes to the Frame.

Cyril Weir, Barry O'Sullivan and Tomoko Horai investigated how the difficulty of speaking tasks can be affected if changes are made to three key task variables: amount of planning time offered; length of response expected; and amount of content scaffolding provided. Their study explored these variables in relation to Part 2 (individual long turn task) of the IELTS Speaking Test and it therefore complements the Wigglesworth and Elder study which focused on the planning time variable in isolation. Using an experimental design, the researchers collected performance and score data for analysis. They supplemented this with an analysis of questionnaire responses related to test-takers' cognitive processing based upon a socio-cognitive framework for test validation. Once again, the findings are encouraging for the IELTS test developers. There is welcome empirical support for the current design of the Part 2 task used in the operational IELTS, both in terms of the quality of candidate performance and the scores awarded, and also in relation to candidate perceptions of task difficulty. Task equivalence is an important issue for IELTS given the large number of tasks which are needed for the operational test, and this study provides useful insights into some of the variables which can affect task difficulty, especially for test candidates at different ability levels.

Paul Seedhouse and Maria Egbert explored the interactional organisation of the IELTS Speaking Test in terms of turn-taking, sequence and repair, drawing their sample for analysis from the large corpus of audio-recordings held by Cambridge ESOL. Since 2002 several thousand recordings of live IELTS Speaking Tests have been collected and these now form a valuable spoken language corpus used by researchers at Cambridge ESOL to investigate various aspects of the Speaking Test. By applying Conversation Analysis (CA) methodology to 137 complete speaking tests, Seedhouse and Egbert were able to highlight key features of the spoken interaction.

Like O'Sullivan and Yang, they observed that examiners adhere closely to the scripted guide they are given to ensure standardisation of the test event. Although spoken interaction in the IELTS Speaking Test is somewhat different to ordinary conversation due to the institutional nature of the test event, the researchers confirm that it does share similarities with interactions in teaching and academic contexts. In addition, the three parts of the Test allow for a variety of task types and patterns of interaction. Seedhouse and Egbert make a number of useful recommendations which will inform aspects of test design as well as examiner training, particularly in relation to the rounding-off questions at the end of Part 2.

In the final report in Volume 6, **John Read and Paul Nation** investigated the lexical dimension of the IELTS Speaking Test. Allocation of grant funding for this study once again reflected the IELTS partners' concern to undertake validation work following introduction of the revised Speaking Test in 2001. When the holistic or global scale for speaking was replaced with four analytic criteria and scales in July 2001, one of these four was *Lexical Resource*; this requires examiners to attend to the accuracy and range of a candidate's vocabulary use as one basis for judging their performance. The Read and Nation study therefore set out to measure lexical output, variation and sophistication, as well as the use of formulaic language by candidates. As the researchers point out in their literature review, there was a strong motivation to explore speaking assessment measures from a lexical perspective given the relative lack of previous research on spoken (rather than written) vocabulary and the growing recognition of the importance of lexis in second language learning.

For this study the researchers created a small corpus of texts derived from transcriptions of Speaking Tests recorded under operational conditions at IELTS tests centres worldwide. As for the Seedhouse and Egbert study, they were given access to the corpus of IELTS Speaking Test recordings at Cambridge ESOL from which they selected a subset of 88 performances for transcription and analysis.

The study's findings are broadly encouraging for the IELTS test developers, confirming that the *Lexical Resource* scale does indeed differentiate between higher and lower proficiency candidates. At the same time, however, the study highlights the complexity of this aspect of spoken performance and the extent to which candidates who receive the same band score sometimes display markedly different qualities in their individual performance. The study also provides useful insights into how different topics in Parts 2 and 3 influence the nature and extent of lexical variation. Such insights can feed back into the test writing process; they can also inform the training of IELTS examiners to direct their attention to salient distinguishing features of the different bands and so assist them in being able to reliably rate vocabulary performance as a separate component from the other three rating criteria. The researchers suggest that in the longer term, and following additional research into how this scale operates, there may be a case for some further revision of the rating descriptors.

Revision of the IELTS Speaking Test in 2001 made it possible to address a number of issues relating to the quality and fairness of the Test. Each of the research studies reported in Volume 6 offers important empirical evidence to support claims about the usefulness of the current IELTS Speaking Test as a measure of L2 spoken language proficiency. In addition, they all provide valuable insights which can inform the ongoing development process (task design, examiner training, etc) as well as future revision cycles. All the reports from the funded projects in this volume highlight avenues for further research and researchers wishing to apply for future grants with British Council and IELTS Australia joint-funded programme may like to take note of some of the suggestions made.

Dr Lynda Taylor
Assistant Director – Research and Validation Group
University of Cambridge ESOL Examinations
September 2006