

Introduction

Lynda Taylor

Consultant to Cambridge ESOL

Cyril J Weir

University of Bedfordshire

The IELTS Joint-funded Research Program

To support the ongoing development of the International English Language Testing System (IELTS), the IELTS partners co-ordinate an ongoing and wide-ranging research and validation programme. A major component of this programme for the past 15 years has been the grant-funded research.

The origins of this research date back to the mid-1990s when the IELTS Australia Board first set aside grant funding for research purposes associated with the IELTS test and invited external researchers to submit IELTS-related proposals for consideration and selection. The Board believed that such external research would complement internal research and validation activities being conducted by the IELTS partnership to provide valuable additional insights and information on a range of issues relating to the quality and standing of IELTS. Another reason for offering funding grants for external research studies was to help IELTS stakeholders (including English language professionals and teachers) to develop a greater knowledge and understanding of the test, and thus contribute to improved assessment literacy in the public domain. In 1998 the British Council joined IDP: IELTS Australia in setting aside annual funding for research grants and since that time the programme has been jointly funded by these two IELTS partners. Cambridge ESOL, the third IELTS partner, supports the Joint-funded Research Program through the provision of data, materials, advice and various other types of assistance to approved researchers.

The first round of funded studies was conducted in 1995 and selected reports resulting from these projects were edited and published jointly by English Australia English Language Intensive Courses for Overseas Students (ELICOS) and IELTS Australia as *IELTS Research Reports 1998*, Volume 1. Since 1998, 10 more volumes of *IELTS Research Reports* have been published. From 2006 onwards (*IELTS Research Reports 6*), most of the volumes were published jointly with the British Council and they contain a useful introduction by Lynda Taylor discussing and evaluating the impact of

IELTS Collected Papers 2

findings from the funded studies on the ongoing development of IELTS. In 2011 all the *IELTS Research Reports* were made freely available in electronic format via the IELTS website, in addition to being available for purchase in hard copy.

The annual IELTS Joint-funded Research Program is widely publicised via print and electronic media. The call for proposals is issued in April each year and it aims to reflect current concerns and issues relating to IELTS as a major international English language proficiency test with high-stakes value. A joint research committee, comprising representatives of the three IELTS partners, agrees on the year's research priorities and oversees the tendering process. Research proposals are reviewed and evaluated according to the following criteria:

- relevance and benefit of outcomes to IELTS
- clarity and coherence of the proposal's rationale, objectives and methodology
- feasibility of outcomes, timelines and budget
- qualifications and experience of proposed project staff
- potential of the project to be reported in a form which would be both useful to IELTS and of interest to an international audience.

In determining the quality of the proposals and the research to be carried out, the Committee consults with a panel of external reviewers and with members of Cambridge ESOL's Research and Validation Group, according to their specialist areas of expertise. Research projects are currently funded up to a maximum of £22,000 or AU\$36,000, though from 2012 this figure is expected to be significantly increased. The Committee also oversees the publication and/or presentation of research findings.

Since 1995 the outcomes of the funded research programme have made a significant contribution to the monitoring, evaluation and ongoing development of IELTS, with particular reference to:

- the assessment of speaking in IELTS, e.g. issues of task design, candidate discourse, assessment criteria, test bias, examiner/rater behaviour, examiner/rater training and monitoring
- the assessment of writing in IELTS, e.g. issues of task design, construct validity, features of writing performance, rater training and monitoring, approaches to assessment
- the assessment of reading in IELTS, e.g. issues relating to the construct of academic reading, cognitive processing in reading assessment, approaches to developing reading test materials
- the assessment of listening in IELTS, e.g. issues relating to the construct of academic listening, test-taker strategies, assessing listening-into-speaking ability

Introduction

- the impact of IELTS in education and society, e.g. stakeholder attitudes, use of test scores, score gains, impact on courses and preparation materials, with key user groups
- computer-based assessment and IELTS.

One of the most valuable outcomes of reports from joint-funded projects is the surveys of recent literature they provide. These help the IELTS test developers stay up to date with theoretical and empirical work in a wide range of fields (including some that are only indirectly linked to language testing) allowing them to take account of these in their work.

Since 1995, more than 90 research studies and over 130 individual researchers have received grants under the IELTS Joint-funded Research Program (up to and including Round 16). Over the years, the funded research programme has become a key component within the larger research and validation agenda in support of IELTS and it reflects the IELTS partners' well-established commitment to the continuing improvement of the test.

The background to this volume

As the body of research from the Joint-funded Research Program steadily increased in the early 2000s, Cambridge ESOL considered publishing a selection of the funded project reports as a single volume, based around a common theme. In this way it was hoped they might become available to a wider audience and illustrate the value of this work within the larger research and validation agenda underpinning IELTS.

Many of the funded research studies conducted between 1995 and 2005 focused specifically on the IELTS Speaking and Writing components, both of which were receiving considerable attention at that time from the test developers. Findings from these external studies complemented internal validation and research studies conducted or commissioned by the IELTS partnership, especially those undertaken by Cambridge ESOL. Taken together, research findings directly informed major revision projects for the productive components of IELTS: the IELTS Speaking Revision Project took place between 1998 and 2001, and the IELTS Writing Revision Project was carried out between 2001 and 2005. Ten studies on IELTS speaking and writing assessment were therefore selected for a volume in the *Studies in Language Testing* series, produced jointly by Cambridge ESOL and Cambridge University Press. The volume was entitled *IELTS Collected Papers: Research in speaking and writing assessment*. It was jointly edited by Lynda Taylor and Peter Falvey and published in 2007. At that time it was hoped that a companion volume might follow at some point with a focus on IELTS reading and listening.

This volume – *IELTS Collected Papers 2: Research in reading and listening*

IELTS Collected Papers 2

assessment, co-edited by Lynda Taylor and Cyril J Weir – fulfils that aspiration by bringing together eight joint-funded studies on IELTS reading and listening assessment, all of which have been completed in recent years. The eight studies have direct relevance to validation claims made for the Academic Reading and Listening components of IELTS. As such, this title should be of immediate interest to test stakeholders and all who are directly involved with IELTS in some way, e.g. test takers, test score users and teachers preparing students for the test. However, the volume is also likely to be of interest to anyone concerned with the assessment of reading and listening proficiency in English, especially as it relates to language use for academic and professional purposes.

The four studies in Part 1 (Chapters 1–4) focus on the IELTS Academic Reading test. Findings from these studies provided the IELTS partners with valuable insights into the construct validity of the Academic Reading test, as well as into the nature and effectiveness of the test writing process. The four studies in Part 2 (Chapters 6–9) focus on the assessment of listening in IELTS. Findings from these studies offered rich insights into the construct validity of the Listening test, as well as into the nature of test takers' listening performance and the relationship between listening and speaking skills. The findings from the eight studies offer important evidence in support of claims about test usefulness, while at the same time helping to highlight specific aspects needing closer review and possible attention in future test revision projects. In combination with outcomes from other commissioned studies and internal validation investigations, they feed into the ongoing process of development and validation for the IELTS Reading and Listening tests.

The publication of *IELTS Collected Papers 2* is seen by the IELTS partners as part of their continuing contribution to the field of language testing and assessment in a number of ways. It allows more of the IELTS-related funded research conducted in recent years to be shared with a wider audience, not just among IELTS stakeholders but within the broader language testing and assessment community. One rationale for the IELTS Joint-funded Research Program is to promote and support research activity among test stakeholders which not only contributes to the ongoing validation and development of IELTS, but also helps to build greater knowledge and understanding of the strengths and limitations of the test. With this in mind and following the pattern of the earlier *IELTS Collected Papers* in 2007, two additional chapters have been included alongside the eight research reports in this volume. Chapters 5 and 10 review and evaluate the outcomes of the studies presented in this volume in terms of their specific implications for the validation and ongoing development of the IELTS Academic Reading and Listening components. In particular, these sections discuss the value of the research findings and explain why some recommendations made in the studies may not be straightforward to implement.

The development of reading and listening assessment in IELTS

The direct assessment of reading and listening proficiency is a long-established feature of the IELTS test. Its origins can be traced back more than 30 years to the development and introduction of ELTS (English Language Testing Service) – the test which preceded IELTS. To contextualise the more recent developments in the IELTS assessment of reading and listening proficiency, this Introduction will review how L2 reading and listening ability were tested in the past in IELTS and its predecessors. We summarise the history of the Reading and Listening components to provide readers with a brief chronological overview of the steady evolution of the test. For a more detailed and comprehensive account of the development of ELTS/IELTS, and its role within the broader context of English language proficiency assessment for academic purposes over more than half a century, the reader is recommended to consult *Assessing Academic English: Testing English proficiency 1950–1989 – the IELTS solution* by Davies (2008), published as Volume 23 in the *Studies in Language Testing* series.

The English Proficiency Test Battery (1965–80)

From 1965 until 1980 the British Council relied on an English language proficiency measure called the English Proficiency Test Battery (EPTB) as part of its procedures for recruiting overseas students into higher education in Britain. The EPTB was a traditional set of standardised tests in a multiple-choice format, focusing on the receptive skills of reading and listening together with a knowledge of grammar and pronunciation. (Facsimile test material for the original EPTB can be found in Appendices 2.1–2.3 of Davies (2008:120–135).)

Although the EPTB developers readily acknowledged the importance of writing and speaking skills, the practical problems of testing these skills (e.g. the requirement for skilled examiners), combined with the British Council's need for a test which could be taken in a short period of time, meant that tests of speaking and writing could not realistically be included in the EPTB. Thus it was tests of reading and listening comprehension which constituted the primary measures of students' English language proficiency via their implied relationship to students' ability to follow a university course and to pass the associated examinations at the end of the course. Davies (2008) explains the dilemma that faced the original EPTB designers:

In other words, this was very much a pragmatic approach; could success on an English-medium academic course be predicted on the basis of tests

IELTS Collected Papers 2

of reading and listening alone? There was no assumption that speaking or writing were in any way less important than reading and listening, rather, that if the test proved to be satisfactory, then it could be claimed that the language skills relevant to academic success were accessible through reading and listening. They could, of course, be equally accessible through writing and/or speaking if appropriate procedures could be found to test these skills (Davies 2008:14).

Davies records how the Reading component of EPTB included subtests of *grammatical structure*, *reading speed* and *comprehension* (both science and non-science texts), while the Listening component comprised subtests of *phonemic discrimination* (in isolation and in context), *intonation and stress* (in conversation), and *comprehension* (again, both science and non-science texts) (see Davies 2008:16–20 for more detail on the individual EPTB subtests). With regard to the linguistic features of the Listening subtests, it is interesting to note that the test developers decided to use only a Standard British English dialect, since this was ‘the variety most in evidence among the educated’, along with modified Received Pronunciation (RP), on the grounds that this was ‘the most favoured accent and probably the one best described’ (2008:14). The issue of how to address linguistic variation in test development has in recent years become the subject of much debate among applied linguists and test designers (see Taylor 2008, 2009). Though the policy decision made in the 1960s may now appear to us somewhat conservative, and some might see it as ‘post-colonial’, it is interesting to note that even in the 1960s the test developers clearly saw dialect and accent as a matter for consideration and decision.

In the mid-1970s a project was established to develop a replacement for the EPTB which would address some of the problems the test was facing (e.g. limitations on the number of parallel versions), and which could also take account of the significant changes that took place in the 1960s and 1970s in approaches to language learning and teaching. The new communicative competence paradigm brought with it a much greater emphasis on the use of language skills in context. For the testing of listening and reading this was to mean a move away from assessing linguistic knowledge and comprehension via discrete-point test items derived from the behavioural and structuralist paradigm of the 1960s/1970s, represented in tasks such as cloze-elide, C-test, phonemic or grammatical triplets and conversational adjacency pairs. Instead, there was a shift towards a much more contextualised, language-in-use oriented approach to testing comprehension of spoken and written language (see below). Not surprisingly, the decision to replace EPTB also inspired a fresh discussion of whether the new test could/should now include components to assess writing and speaking skills.

The testing of reading and listening in ELTS (1980–89)

The replacement for EPTB was a brand new test, developed jointly by the University of Cambridge Local Examinations Syndicate (UCLES) and the British Council, entitled the English Language Testing Service (ELTS). It was introduced in 1980 after a four-year period of development. The test's overall design reflected the new paradigm of communicative language teaching and testing, with its emphasis on authenticity and relevance and its concern to assess ability to use language rather than simply test knowledge about language (see Chapter 2 in Davies 2008 for a full account of the development of ELTS).

The new on-demand test also took account of the growing interest at that time in English or Language for Specific Purposes (ESP/LSP). Test tasks in ELTS were based on a careful analysis of the ways in which language was actually used in academic contexts, i.e. as part of study on university courses, and they were intended to reflect the use of language in the 'real world'. A strong emphasis on needs analysis and on communicative language demands in the study/work context meant that, alongside the Reading and Listening components, subtests of writing and speaking ability were now allocated a place within the new test – in the form of the Writing subtest and the Individual Interview.

As far as reading and listening were concerned, there were two 'General Tests' adopting a multiple-choice format: G1 tested reading and G2 tested listening (see summary details of both below). These General Tests were taken by all test takers, regardless of their academic speciality. In addition to G1 and G2, a subject-specific 'Modular Test' (M1) was offered to assess reading comprehension within a specific academic domain: M1 (Study Skills). This Study Skills component was linked to one of six academic 'domains' or areas of study (Life Sciences, Social Studies, Physical Sciences, Technology, Medicine, General Academic; the latter – General Academic – was designed for those whose areas of interest did not fit into any of the preceding domains). In addition to M1 for reading, there was also a writing test (M2) and an oral interview (M3) both of which were available across the six domains (see the Introduction in Taylor and Falvey 2007 for more details of the M2 and M3 modules).

For the three Modular tests – M1, M2 and M3 – each ELTS candidate received a source booklet relevant to their chosen discipline from the six domains available. The source booklet contained extracts from appropriate academic texts, including bibliography and index, and it formed the basis for not only the multiple-choice reading comprehension tasks in M1, but also the writing tasks in M2 and the main discussion in the M3 Interview.

Although undoubtedly innovative in its design and implementation

IELTS Collected Papers 2

when compared with EPTB and similar tests available at that time (e.g. the Educational Testing Service Test of English as a Foreign Language (TOEFL), the new ELTS test nevertheless presented a number of administrative and practical challenges from the outset. Davies (2008:38–40) reflects that ELTS was a much longer test than EPTB and was more complicated to administer given the multiple and ‘tailored’ components involved. The direct Writing and Speaking subtests required trained and standardised markers and examiners, thus posing an additional administrative burden on British Council test centres and their staff in terms of recruiting and managing suitably qualified personnel. Furthermore, it was often difficult to match the prospective test candidate to the most appropriate subject domain. Finally, generating sufficient comparable test forms across multiple testing domains posed significant challenges for test production and sustainability. (See also Criper and Davies 1988.)

The assessment of reading and listening in ELTS between 1980 and 1989 can be summarised as follows:

Reading test (G1)

- 40 multiple-choice test items, divided into three sections
- presented in a single question booklet together with the texts on which they are based
- Section 1: sentence-length texts
- Section 2: paragraph-length texts (Multiple Choice Question (MCQ) gap-filling)
- Section 3: three related newspaper articles, with some test items on each text independently and some on the texts as a group
- length = 40 minutes
- assessed on a nine-band scale
- clerically marked according to the MCQ key and using a template.

Listening test (G2)

- a tape and a question booklet with 35 multiple-choice test items in four sections
- Section 1: choosing from diagrams
- Section 2: listening to an interview
- Section 3: replying to questions
- Section 4: listening to a seminar
- length = approx. 35 minutes
- assessed on a nine-band scale
- clerically marked according to the MCQ key and using a template.

Study Skills test (M1)

- linked to one of six academic domains (Life Sciences, Social Studies, Physical Sciences, Technology, Medicine, General Academic)
- based on a source booklet containing 5–6 textual extracts for input – taken from books, articles, reports etc. related to the specific subject area plus additional contents pages, bibliographies, appendices and indices
- an accompanying question booklet with 40 multiple-choice test items
- length = 55 minutes
- clerically marked according to the MCQ key and using a template.

Facsimile Reading and Listening test papers (all versions) for the original ELTS introduced in 1980 can be found in Appendix 6.2 of Davies (2008:203–206).

The ELTS Revision Project (1986–89)

Shortly after its introduction in 1980, the British Council and UCLES commissioned the Institute for Applied Language Studies at the University of Edinburgh to undertake a detailed validation study of the test. The ELTS Validation Project (Criper and Davies 1988) explored aspects of the practicality, validity and reliability of the existing English Language Testing Service (ELTS). Work on the five-year validation project was completed in 1986. In addition, valuable research was conducted by applied linguists elsewhere in the UK during the early 1980s which shed light on the English for Academic Purposes (EAP) language and literacy needs of overseas students at British universities (e.g. Geoghehan 1983, Hawkey 1982, Weir 1983).

By 1986 the producers of ELTS determined that the test was once again due for formal review and possible revision. The report of the ELTS Validation Project provided a convenient starting point for the ELTS Revision Project, a three-year project (1986–89) set up under the direction of Professor Charles Alderson of Lancaster University. British Council management support came from a team headed by Dr Peter Hargreaves, who was at that time with the British Council and from 1988 head of the UCLES EFL Division. An Australian perspective was provided by Professor David Ingram of Griffith University, seconded to the revision project in Lancaster from 1987 with support from the International Development Program (IDP) of Australian Universities and Colleges; IDP later became one of three IELTS partners to produce the test from 1989 onwards.

A large-scale, questionnaire-based consultation exercise was conducted in the mid-1980s with various ELTS user groups (receiving institutions, British Council staff, overseas administrators, EAP teachers, language testers and applied linguists) in order to determine the perceived strengths and weaknesses of the existing test and the desirable characteristics of a revised test (see Alderson and Clapham 1992). User views were also gathered via focus group meetings.

IELTS Collected Papers 2

In terms of the practicality and validity of ELTS, those responsible for administering the test at centre level expressed major concerns about the test's length and its logistical complexity, in particular the difficulty of selecting appropriate subject-specific modules for candidates. Specific concerns from test takers about the Reading and Listening components (G1, G2 and M1) touched upon the following issues: the varying ease or difficulty of the reading texts and tasks across and within the General and Modular tests (the M1 subject-specific component was seen as the most difficult); the proximity (or otherwise) of the M1 reading material to the test taker's background knowledge and focus of academic study; the quality of the G2 listening tape and poor acoustics at test venues; and pressure of time in the listening test. An exercise to reconstruct the needs analysis specification by aligning the ELTS Reading and Listening test items to Munby's list of specifications in his *Communicative Syllabus Design* (1978) highlighted the problems inherent in trying to match test items to specified language skills, as well as the challenge of adequate sampling of the relevant reading and listening skills in the tests and comparable coverage of these across the different domain-specific modules (Criper and Davies 1988:89–97).

The ELTS Validation Consultative Conference, held in July 1987, brought together language testing researchers from Britain, Australia, Canada and the USA to review the outcomes of the consultation exercise and to discuss possible options for the future of the test (Hughes, Porter and Weir 1988). It was generally agreed that the test needed shortening, its administration needed simplifying and its reliability needed to be improved.

As far as testing reading skills was concerned, the overlap between the G1 and M1 Reading components was felt to be such that one of these could be dropped without any great loss in order to reduce test length. The consultative committee considered that it made more sense to retain reading in M1 (rather than G1) since this linked reading directly to the M2 Writing component, creating a reading-into-writing proficiency measure which closely reflected academic literacy demands. Reading and writing would remain integrated so that, to some extent at least, candidates' written output depended on the reading input in the Reading subtest, though separate scores would be reported for the two skills.

With regard to testing listening skills, most members of the committee felt that it would be better to convert the Listening test (G2) into an M component, thus making it domain-specific and modular alongside the M1 reading (see Hughes, Porter and Weir 1988:101, Alderson and Clapham 1992:16–17). An integrative test was envisaged, perhaps involving candidates listening to a lecture, making notes and then carrying out a writing task. Unfortunately, practical and logistical constraints were to make such an approach impossible. Most test centres arranged for all candidates to take the existing ELTS G2 Listening test in one room. Offering multiple modular listening tests