

ResearchNotes

Contents

| | |
|---|----|
| Editorial Notes | 1 |
| IELTS, Cambridge ESOL examinations and the Common European Framework | 2 |
| Computer-based IELTS and paper-based versions of IELTS | 3 |
| IELTS Impact: a study on the accessibility of IELTS GT Modules to 16–17 year old candidates | 6 |
| IELTS Writing: revising assessment criteria and scales (Phase 4) | 8 |
| Set Texts in CPE Writing | 12 |
| IELTS – some frequently asked questions | 14 |
| IELTS test performance data 2003 | 18 |
| The IELTS joint-funded program celebrates a decade of research | 20 |
| Conference Reports | 21 |

Editorial Notes

Welcome to issue 18 of *Research Notes*, our quarterly publication reporting on matters relating to research, test development and validation within Cambridge ESOL.

The theme of this issue is the International English Language Testing System (IELTS). IELTS is the examination provided by the three IELTS partners, Cambridge ESOL, British Council and IDP: IELTS Australia and is used for a variety of high-stakes purposes in Academic and General Training contexts.

This issue covers a range of topics relating to IELTS including its position in Cambridge ESOL's own and European frameworks, the comparability of alternative formats, the impact of IELTS on stakeholder groups (candidates, teachers and examiners) and revisions to the rating of this exam. We begin with general issues concerning IELTS before focusing in on specific components and uses of IELTS, with reference to a range of research projects.

In the opening article Lynda Taylor explores the links between IELTS, Cambridge ESOL's other exam suites and two frameworks: the Common European Framework (described in Issue 17 of *Research Notes* which focused on language testing in Europe), and the UK National Qualifications Framework. Lynda describes a series of research studies and presents tables which provide indicative links between IELTS band scores and other examinations.

Tony Green and Louise Maycock describe a number of studies which investigate the comparability of computer-based and paper-based versions of IELTS in terms of candidates' scores and examiners' rating of both versions, in advance of the launch of a computer-based version of IELTS in 2005. Jan Smith reports on an Australian-based study commissioned by Cambridge ESOL to assess the accessibility of IELTS test materials and the teaching materials used to prepare senior school pupils aged 16–17 for the General Training module. These articles show how both the nature and candidature of IELTS are changing over time, issues which will be explored in greater detail in a future *Research Notes*.

The following two articles focus on the Writing component of two high level examinations. Firstly, Graeme Bridges and Stuart Shaw report on the implementation phase of the IELTS Writing: Revising Assessment Criteria and Scales study which consists of training and certificating examiners and introducing a Professional Support Network for IELTS. The next article, by Diana Fried-Booth, explores the rationale and history behind the set texts option in the CPE Writing paper which has been a distinguishing feature of this examination since 1913.

Returning to IELTS, the next article contains a list of frequently asked questions for IELTS covering its format, scoring and rating and other areas. This is followed by some performance data for IELTS including band scores for the whole candidate population and reliabilities of the test materials for 2003. We then review the first ten years of the IELTS Funded Research Program before ending this issue with conference reports focusing on Chinese learners in Higher Education, pronunciation and learner independence and a recent staff seminar given by Vivian Cook on multi-competence and language teaching.

The URL for reading/downloading single articles or issues of *Research Notes* is:
www.CambridgeESOL.org/rs_notes

The URL for subscribing to *Research Notes* is:
www.CambridgeESOL.org/rs_notes/inform.cfm

IELTS, Cambridge ESOL examinations and the Common European Framework

LYNDA TAYLOR, RESEARCH AND VALIDATION GROUP

Test users frequently ask how IELTS scores ‘map’ onto the Main Suite and other examinations produced by Cambridge ESOL, as well as onto the Common European Framework of Reference (CEFR) published by the Council of Europe (2001).

A *Research Notes* article earlier this year on test comparability (Taylor 2004) explained how the different design, purpose and format of the examinations make it very difficult to give exact comparisons across tests and test scores. Candidates’ aptitude and preparation for a particular type of test will also vary from individual to individual (or group to group), and some candidates are more likely to perform better in certain tests than in others.

Cambridge ESOL has been working since the mid-1990s to gain a better understanding of the relationship between its different assessment products, in both conceptual and empirical terms. The conceptual framework presented in *Research Notes 15* (page 5) showed strong links between our suites of level-based tests, i.e. Main Suite, BEC, CELS and YLE. These links derive from the fact that tests within these suites are targeted at similar ability levels as defined by a common measurement scale (based on latent trait methods); many are also similar in terms of test content and design (multiple skills components, similar task/item-types, etc). Work completed under the ALTE Can Do Project also established a coherent link between the ALTE/Cambridge Levels and the Common European Framework (see Jones & Hirtzel 2001).

The relationship of IELTS with the other Cambridge ESOL tests and with the Common European Framework of Reference is rather more complex; IELTS is not a level-based test (like FCE or CPE) but is designed to stretch across a much broader proficiency continuum. So when seeking to compare IELTS band scores with scores on other tests, it is important to bear in mind the differences in purpose, measurement scale, test format and test-taker populations for which IELTS was originally designed. Figure 1 in the *Research Notes 15* article acknowledged this complex relationship by maintaining a distance between the IELTS scale (on the far right) and the other tests and levels located within the conceptual framework.

Since the late 1990s, Cambridge ESOL has conducted a number of research projects to explore how IELTS band scores align with the Common European Framework levels. In 1998 and 1999 internal studies examined the relationship between IELTS and the Cambridge Main Suite Examinations, specifically CAE (C1 level) and FCE (B2 level). Under test conditions, candidates took experimental reading tests containing both IELTS and CAE or FCE tasks. Although the studies were limited in scope, results indicated that a candidate who achieves a Band 6.5 in IELTS would be likely to achieve a passing grade at CAE (C1 level).

Further research was conducted in 2000 as part of the ALTE Can

Do Project in which Can Do responses by IELTS candidates were collected over the year and matched to grades; this enabled Can Do self-ratings of IELTS and Main Suite candidates to be compared. The results, in terms of mean Can Do self-ratings, supported placing IELTS Band 6.5 at the C1 level of the CEFR alongside CAE.

More recently, attention has focused on comparing IELTS candidates’ writing performance with that of Main Suite, BEC and CELS candidates. This work forms part of Cambridge ESOL’s *Common Scale for Writing Project* – a long-term research project which has been in progress since the mid-1990s (see Hawkey and Barker 2004). Results confirm that, when different proficiency levels and different domains are taken into account, a strong Band 6 performance in IELTS Writing (IELTS Speaking and Writing do not currently report half bands) corresponds broadly to a passing performance at CAE (C1 level).

Additional evidence for the alignment of IELTS with other Cambridge ESOL examinations and with the CEFR comes from the comparable use made of IELTS, CPE, CAE and BEC Higher test

Figure 1: Alignment of IELTS, Main Suite, BEC and CELS examinations with UK and European frameworks

| IELTS | Main Suite | BEC | CELS | NQF | CEFR |
|-------|------------|-------|--------|---------|------|
| 9.0 | | | | | |
| 8.0 | | | | 3 | C2 |
| 7.0 | CPE | | | | |
| 6.0 | CAE | BEC H | CELS H | 2 | C1 |
| 5.0 | FCE | BEC V | CELS V | 1 | B2 |
| 4.0 | PET | BEC P | CELS P | Entry 3 | B1 |
| 3.0 | KET | | | Entry 2 | A2 |
| | | | | Entry 1 | A1 |

Key:

- IELTS: International English Language Testing System
- KET: Key English Test
- PET: Preliminary English Test
- FCE: First Certificate in English
- CAE: Certificate in Advanced English
- CPE: Certificate of Proficiency in English
- BEC: Business English Certificates: H-Higher, V-Vantage, P-Preliminary
- CELS: Certificates in English Language Skills: H-Higher, V-Vantage, P-Preliminary
- NQF: National Qualifications Framework
- CEFR: Common European Framework of Reference

Figure 2: Indicative IELTS band scores at CEFR and NQF levels

| Corresponding NQF Level | Corresponding CEFR Level | IELTS approximate band score |
|-------------------------|--------------------------|------------------------------|
| Level 3 | C2 | 7.5+ |
| Level 2 | C1 | 6.5/7.0 |
| Level 1 | B2 | 5.0/5.5/6.0 |
| Entry 3 | B1 | 3.5/4.0/4.5 |
| Entry 2 | A2 | 3.0 |

scores by educational and other institutions (for more details see www.CambridgeESOL.org/recognition).

The accumulated evidence – both logical and empirical – means that the conceptual framework presented in early 2004 has now been revised to accommodate IELTS more closely within its frame of reference. Figure 1 illustrates how the IELTS band scores, Cambridge Main Suite, BEC and CELS examinations align with one another and with the levels of the Common European Framework and the UK National Qualifications Framework. Note that the IELTS band scores referred to in both figures are the overall scores, not the individual module scores.

Figure 2 indicates the IELTS band scores we would expect to be achieved at a particular CEFR or NQF level.

It is important to recognise that the purpose of Figures 1 and 2

is to communicate relationships between tests and levels in *broad* terms within a common frame of reference; they should *not* be interpreted as reflecting strong claims about exact equivalence between assessment products or the scores they generate, for the reasons explained in *Research Notes* 15.

The current alignment is based upon a growing body of internal research, combined with long established experience of test use within education and society, as well as feedback from a range of test stakeholders regarding the uses of test results for particular purposes. As we grow in our understanding of the relationship between IELTS, other Cambridge ESOL examinations and the CEFR levels, so the frame of reference may need to be revised accordingly.

References and further reading

- Council of Europe (2001) *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*, Cambridge: CUP.
- Hawkey, R and Barker, F (2004) Developing a common scale for the assessment of writing, *Assessing Writing*, 9 (2), 122–159.
- Jones, N and Hirtzel, M (2001) Appendix D: The ALTE Can Do Statements, in the *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*, Council of Europe, Cambridge: Cambridge University Press.
- Morrow, K (2004) (Ed.) *Insights from the Common European Framework*, Oxford: Oxford University Press.
- Taylor, L (2004) Issues of test comparability, *Research Notes* 15, 2–5.

Computer-based IELTS and paper-based versions of IELTS

TONY GREEN AND LOUISE MAYCOCK, RESEARCH AND VALIDATION GROUP

Introduction

A linear computer-based (CB) version of the IELTS test is due for launch in 2005. The CB test will, in the context of growing computer use, increase the options available to candidates and allow them every opportunity to demonstrate their language ability in a familiar medium. As the interpretation of computer-based IELTS scores must be comparable to that of paper-based (PB) test scores, it is essential that, as far as is possible, candidates obtain the same scores regardless of which version they take.

Since 2001, the Research and Validation Group has conducted a series of studies into the comparability of IELTS tests delivered by computer and on paper. Early research indicated that we could be confident that the two modes of administration do not affect levels of performance to any meaningful extent. However, the findings were muddled by a motivational effect, with candidates performing better on official than trial tests. To encourage candidates to take trial forms of the CB test, these had been offered as practice material to those preparing for a live examination. However, candidates tended not to perform as well on these trial versions (whether computer- or paper-based) as they did on the live PB versions that provided their official scores.

This report relates to the findings of the first of two large scale trials, referred to as Trial A, conducted in 2003–2004. In these studies, to overcome any effect for motivation, candidates for the official IELTS test were invited to take two test versions at a reduced price – a computer-based version and a paper-based version – but were not informed which score would be awarded as their official IELTS result.

Previous studies of CB and PB comparability

When multiple versions or ‘forms’ of a test are used, two competing considerations come into play. It could be argued that any two test forms should be as similar as possible in order to provide directly comparable evidence of candidates’ abilities and to ensure that the scores obtained on one form are precisely comparable to the scores obtained on another. On the other hand, if the forms are to be used over a period of time, it could be argued that they should be as dissimilar as possible (within the constraints imposed by our definition of the skill being tested) so that test items do not become predictable and learners are not encouraged to focus on a narrow range of knowledge. On this

basis, Hughes (1989) argues that we should 'sample widely and unpredictably' from the domain of skills we are testing to avoid the harmful backwash that might result if teachers and learners can easily predict the content of the test in advance. Indeed, this would pose a threat to the interpretability of the test scores as these might come to reflect prior knowledge of the test rather than ability in the skills being tested.

Different forms of the IELTS test are constructed with these two considerations in mind. All test tasks are pre-tested and forms are constructed to be of equal difficulty (see Beeston 2000 for a description of the ESOL pretesting and item banking process). The test forms follow the same basic design template with equal numbers of texts and items on each form. However, the content of the texts involved, question types and targeted abilities may be sampled differently on each form. The introduction of a CB test raises additional questions about the comparability of test forms: Does the use of a different format affect the difficulty of test tasks? Do candidates engage the same processes when responding to CB tests as they do when responding to PB tests?

Earlier studies of IELTS PB and CB equivalence have involved investigations of the receptive skills (Listening and Reading) and Writing components. The Speaking test follows the same face-to-face format for both the CB and PB test formats and so is not affected by the CB format.

Shaw et al (2001) and Thighe et al (2001) investigated the equivalence of PB and CB forms of the Listening and Reading IELTS components. Shaw et al's study (ibid.) involved 192 candidates taking a trial version of CBIELTS shortly before a different live PB version of the test which was used as the basis for their official scores. The CB tests were found to be reliable and item difficulty was highly correlated between PB and CB versions ($r = 0.99$ for Listening, 0.90 for Reading). In other words, test format had little effect on the order of item difficulty. Correlations (corrected for attenuation) of 0.83 and 0.90 were found between scores on the CB and PB versions of Listening and Reading forms respectively, satisfying Popham's (1988) criterion of 0.8 and suggesting that format had a minimal effect on the scores awarded. However, Shaw et al (ibid.) called for further investigation of the comparability of PB test forms as a point of comparison.

The Thighe et al (2001) study addressed this need. Candidates were divided into two groups: *Live* candidates comprised 231 learners preparing to take an official IELTS test at eight centres worldwide who took a trial form of either the Reading or Listening component of PB IELTS two weeks before their official 'live' test, which was then used as a point of comparison; *Preparatory* candidates were 262 students at 13 centres who were each administered two different trial forms of either the Reading or Listening PB component with a two week interval between tests. Table 1 shows rates of agreement – the percentage of candidates obtaining identical scores, measured in half bands, on both versions of the test – between the different test forms. Half band scores used in reporting performance on the Reading and Listening components of IELTS typically represent three or four raw score points out of the 40 available for each test. For the Live candidates, who more closely represented the global IELTS candidature, there was absolute agreement (candidates obtaining identical band

scores on both test forms) in 30% of cases for Reading and 27% of cases for Listening. 89% of scores fell within one band on both test occasions. The rates of agreement found between PB test versions would serve as a useful benchmark in evaluating those observed in the current study.

For IELTS Writing, the difference between the CB and PB formats is mainly in the nature of the candidate's response. On the PB test, candidates write their responses by hand. For CB they have the option either of word-processing or hand-writing their responses. Brown (2003) investigated differences between handwritten and word-processed versions of the same IELTS Task Two essays. Legibility, judged by examiners on a five-point scale, was found to have a significant, but small, impact on scores. Handwritten versions of the same script tended to be awarded higher scores than the word-processed versions, with examiners apparently compensating for poor handwriting when making their judgements. Shaw (2003) obtained similar findings for First Certificate (FCE) scripts.

A study by Whitehead (2003) reported in *Research Notes 10* investigated differences in the assessment of writing scripts across formats. A sample of 50 candidates' scripts was collected from six centres which had been involved in a CBIELTS trial. Candidates had taken a trial CB version of IELTS followed soon afterwards by their live pen-and-paper IELTS; thus for each candidate a handwritten and a computer-generated writing script was available for analysis. For Whitehead's study, six trained and certificated IELTS examiners were recruited to mark approximately 60 scripts each; these consisted of handwritten scripts, computer-based scripts and some handwritten scripts typed up to resemble computer-based scripts. The examiners involved also completed a questionnaire addressing the assessment process and their experiences of, and attitudes to, assessing handwritten and typed scripts. Whitehead found no significant differences between scores awarded to handwritten and typed scripts. Although CB scripts yielded slightly lower scores and higher variance, Whitehead suggests that these differences could be attributable to the motivation effect described above.

Although response format seemed to have little impact on scores, Brown (2003), Shaw (2003) and Whitehead (2003) all identified differences in the way that examiners approach typed and handwritten scripts. IELTS examiners identified spelling errors, typographical errors and judgements of text length in addition to issues of legibility as areas where they would have liked further guidance when encountering typed responses. One response to this feedback from examiners has been to include a word count with all typed scripts, an innovation that was included in the current study.

CBIELTS Trial A 2003–2004

627 candidates representing the global IELTS test-taking population took one CBIELTS Listening form and one CBIELTS Academic Reading form, alongside one of three CB Writing versions. Each candidate took the computer-based test within a week of taking a live paper-based test (involving 18 different forms of the PB test). Half of the candidates were administered the CB test first, the other half took the PB test first. Candidates could choose whether to type

Table 1: Agreement rates of live and preparatory candidates for Reading and Listening (Thighe et al 2001)

| | Live candidates | | Preparatory candidates | |
|------------------------------------|-----------------|-----------|------------------------|-----------|
| | Reading | Listening | Reading | Listening |
| % agreement | 30% | 27% | 27% | 25% |
| % agreement to within half a band | 68% | 62% | 61% | 68% |
| % agreement to within a whole band | 89% | 89% | 85% | 91% |

Table 2: Agreement rates for Reading, Listening, Writing, and Overall scores in Trial A

| | Reading | Listening | Writing | Overall† |
|------------------------------------|---------|-----------|---------|----------|
| % agreement | 26% | 22% | 53% | 49% |
| % agreement to within half a band | 72% | 62% | * | 95% |
| % agreement to within a whole band | 91% | 85% | 92% | 100% |

* Scores for Writing tests are awarded in whole band increments † Note that overall scores for the two tests (CB and PB) include a common Speaking component

their answers to the CB Writing tasks or to hand-write them. All candidates took only one Speaking test, since this is the same for both the PB and CB tests. The candidates (and Writing examiners) were not aware of which form would be used to generate official scores and so can be assumed to have treated both tests as live. Candidates were also asked to complete a questionnaire covering their ability, experience and confidence in using computers as well as their attitudes towards CBIELTS. The questionnaire was administered after the second of the two tests and results will be reported in a future issue of *Research Notes*.

Of the 627 candidates who took part in the trial, 423 provided a complete data set, including responses to two test forms and the questionnaire. Despite a slightly higher proportion of Chinese candidates in the sample compared with the live population, the sample represented a range of first languages, reasons for taking IELTS, level of education completed, gender and age groups.

Findings

Table 2 shows rates of agreement between the band scores awarded on the CB versions with band scores awarded on the PB versions. The figures for absolute agreement are similar to, albeit slightly lower than those obtained in the earlier trials comparing PB test forms, while agreement to within half a band is slightly higher. The similarity of the results suggests that the use of a different test format (CB or PB) has very little effect on rates of agreement across forms with nearly 50% of candidates obtaining an identical band score for the test on both occasions and a further 45% obtaining a score that differed by just half a band on the nine band IELTS scale (see Overall column).

Although the results suggested that format has a minimal effect on results, some areas were identified for further investigation (Maycock 2004). Among these it was noted that, for Writing, candidates performed marginally better on the paper-based test than on the computer-based test. It was suggested that this could be due to differences in task content between versions, the effects

of typing the answers, or differences in the scoring of typed and handwritten scripts.

To respond to this concern, a follow-up study was implemented to identify sources of variation in the scoring of writing scripts. The study involved 75 candidates selected to represent variety in L1 background and IELTS band score (Green 2004). Their scripts included responses to both computer- and paper-based versions of the test. All handwritten responses (all of the PB scripts and 25 of the 75 CB scripts) were transcribed into typewritten form so that differences in the quality of responses to the two exam formats could be separated from differences attributable to presentation or to response mode. Multi-faceted Rasch analysis was used to estimate the effects of test format (CB/PB) response format (handwritten/typed) and examiner harshness/leniency on test scores. The evidence from the study indicated that there was no measurable effect of response type and that the effect of test format, although significant, was minimal at 0.1 of a band.

Conclusion

A further trial (Trial B) of additional forms of CBIELTS has just been completed and analysis is underway. The evidence gathered to date suggests that CBIELTS can be used interchangeably with PB IELTS and that candidates, given adequate computer familiarity, will perform equally well on either version of the test. However, Trial A has raised issues of scoring and the treatment of errors that will need to be addressed through examiner training and guidance. The marking process and how examiners are affected by scoring typed rather than handwritten scripts will be a continuing area of interest and will be explored further in Trial B. Initial analysis of questionnaire data suggests that candidates are generally satisfied with the CB version of the test and regard it as comparable to the PB version.

Additional questions remain regarding the processes that candidates engage in and the nature of the language elicited when taking tests with different formats. To address this, work has been

commissioned by Cambridge ESOL to investigate candidate test taking processes on CB and PB tests and this is currently being undertaken by Cyril Weir, Barry O'Sullivan and colleagues at the Centre for Research in Testing, Evaluation and Curriculum in ELT at Roehampton University.

References and further reading

- Beeston, S (2000) The UCLES EFL Item Banking System, *Research Notes*, 2, 8–10.
- Brown, A (2003) Legibility and the Rating of Second Language Writing: An Investigation of the Rating of Handwritten and Word-processed IELTS Task Two Essays, in *IELTS Research Reports Volume 4*, Canberra: IDP: IELTS Australia.
- Green, A (2004) *Comparison of Computer and Paper Based Versions of IELTS Writing: A further investigation of Trial A data*, Cambridge: Cambridge ESOL internal Validation report 585.
- Hughes, A (1989) *Testing for Language Teachers*, Cambridge, Cambridge University Press.
- Maycock, L (2004) *CBIELTS: A Report on the Findings of Trial A (Live Trial 2003/04)*, Cambridge: Cambridge ESOL internal Validation report 558.
- Popham, W J (1988) *Educational Evaluation* (2nd edition), Prentice Hall: New Jersey.
- Shaw, S (2003) Legibility and the rating of second language writing: the effect on examiners when assessing handwritten and word-processed scripts, *Research Notes* 11, 7–10.
- Shaw, S, Jones, N and Flux, T (2001) *CBIELTS – A comparison of computer-based and paper versions*, Cambridge: Cambridge ESOL internal Validation report 216.
- Thighe, D, Jones, N and Geranpayeh, A (2001) *IELTS PB and CB Equivalence: A Comparison of Equated Versions of the Reading and Listening Components of PB IELTS in relation to CB IELTS*, Cambridge: Cambridge ESOL internal Validation report 288.

IELTS Impact: a study on the accessibility of IELTS GT Modules to 16–17 year old candidates

JAN SMITH, CONSULTANT FOR CAMBRIDGE ESOL

Introduction

During the past 10 years Australia has experienced a huge growth in its international education sector as a result of underlying economic, political and educational change in the region. What was initially a rise in demand for tertiary places, first at post-graduate then at undergraduate level, soon began to manifest itself in the secondary sector. Students wishing to undertake their university education in Australia began to see advantage in taking the upper secondary mainstream examinations in order to gain their places through the same pathway as local students.

IELTS was already successfully established as an English language proficiency examination for prospective students applying for undergraduate and vocational courses. Government institutions in Australia found IELTS to be a useful language proficiency measure in educational and social contexts and this soon led to its consideration by Australian secondary schools as a possible means of setting English language proficiency requirements for international students, aged 16–17, entering upper secondary. The IELTS General Training Modules (GT) were perceived to be appropriate since these were originally developed to suit the needs of those wishing to undertake further study/training of a non-academic nature, or as a bridge between school and university.

There remained, however, some questions as to whether the General Training Modules with their perceived adult education content would be accessible to upper secondary students. The necessity to answer these questions became apparent when the Australian government proposed the use of scores gained off-shore

in the IELTS General Training Modules as a pre-requisite for entry to Australia to take up secondary school places.

In 2003 an IELTS impact study set out to investigate whether the General Training Modules of IELTS would be accessible to candidates in this age group and would, as a result, be effective in providing the information required by secondary schools when recruiting prospective students. Since the purpose of the study was to assess the accessibility of the testing materials the project brief required a comparison between materials used in classes preparing students for entry to upper secondary and those used in the IELTS General Training Modules.

Choice of data sources

The study brief required that two sources of examination data relating directly to the IELTS General Training Modules be comprehensively analysed. Since the items and tasks in the examination were to form the basis for comparison, the specimen materials obtainable from IDP (*IELTS Specimen Materials 2003*) would certainly provide an acceptable overview of the test. These materials have been designed for the purpose of familiarising prospective candidates with the requirements of the test. A second test data source was sought and a decision was made to use an IELTS practice test book published by Cambridge University Press (*Cambridge IELTS 3*). Testing material in books such as this is subjected to the same quality assurance procedures as live test materials.

The selection of teaching material for analysis alongside this was

a little more difficult. Attempts were made to determine which materials were being used in English language preparation courses for students entering upper secondary both in Australia and overseas. The three major providers of these courses in Australia – private language schools, public sector language schools (mostly located within Technical and Further Education – or TAFE – Colleges) and the Migrant English Service (Intensive English Centres) – were all approached. This contact was informal and consisted of talking to Directors of Studies about the size of their secondary preparation courses and the teaching materials used.

A number of major private providers nominated *Cutting Edge* (Cunningham and Moor 2002) as the text used as core material for their courses. In addition to this textbook, all centres contacted used texts from mainstream secondary subject areas to develop language tasks and teaching activities. The largest private institution teaching secondary age students also prepares for University Foundation Studies. This is an alternative study pathway to that offered by the various year 12 assessment models in the different States, and is for university applicants who have already completed year 11 in their own country. This college also used *Insight into IELTS* (Jakeman and McDowell 1999) as a course textbook because the majority of their students were required to take an IELTS test before taking up places in their secondary or Foundation Studies course. It was decided that *Cutting Edge Upper-intermediate* should be analysed as this is the textbook and level which the largest number of students reach prior to entering their mainstream program.

Contact with TAFE providers, however, revealed a significantly different and interesting situation. Their approach was developed in conjunction with the public secondary schools using mainstream materials only and no recognised English language textbook. However, it was not possible to use any of this material for analysis as both the approach and materials developed by teachers were still being documented.

Every attempt was made to find suitable materials for analysis from Australia's major overseas markets for secondary education but this was not possible within the project scope. It is hoped that analysing materials from TAFE and overseas sources may form the basis for a further study.

In the light of the information provided through this brief survey, texts selected for analysis on the basis of their use in the Australian secondary school English preparation context were:

- *IELTS Specimen Materials* – GT modules only
- *Cambridge IELTS 3* – Examination Papers
- *Cutting Edge Upper-intermediate Student's Book*
- *Insight into IELTS*

Methodology and results

The analysis of the four data sources was conducted using the *Instrument for Analysis of Textbook Materials* (IATM). The IATM is one of a series of instruments for impact investigation which have been developed and validated by Cambridge ESOL in collaboration with Dr Roger Hawkey and building on earlier work commissioned from the University of Lancaster (UK). The IATM

allows for classification of textbooks according to their type and organisation by providing a framework for objective analytic responses using listings and tick boxes for items included and a further opportunity for comment. It enables the researcher to determine the content through analysis of such aspects as language features, enabling skills, text types and topic areas. (For a fuller description see Saville & Hawkey 2004.)

The four data sources were subjected to comprehensive analysis using the IATM so meaningful comparisons could be made between the requirements of the test and the materials being used to prepare students for entry into upper secondary classes. The analysis was very detailed and provided considerable insights into the potential for the chosen content and approach to prepare a candidate for the IELTS General Training Modules.

As would be expected, there was a very close match between the two examples of actual testing material (*IELTS Specimen Materials* and *Cambridge IELTS 3*) and the test preparation book *Insight into IELTS*. This book looks at each skill area – Reading, Writing, Listening and Speaking – relating it to the examination sub-test and offering advice and enabling skill exercises. The authors have set out to de-mystify the test and in doing so relate their teaching material directly to it.

The match between the testing materials and *Cutting Edge Upper-intermediate* – the textbook in use in many of the preparation programs – was also very strong. However, as a general English textbook *Cutting Edge* has broader objectives than mere test preparation, as stated by the publishers themselves on the cover of the book

“...It combines rich international content, comprehensive grammar, and real life functional grammar.”

In fact, *Cutting Edge* was found to contain far more wide-ranging material than any other of the data sources. The range of topics and their exploitation, text and task types significantly exceeded the requirements of the test in the skill areas of reading, listening and speaking. In only one area, that of essay writing, was it found to be inadequate at preparing students for the requirements of IELTS GT Writing Task 2. Although students using this textbook would have been exposed to many of the ideas and concepts which regularly appear in GT Writing Task 2, *Cutting Edge Upper-intermediate* does not offer specific preparation exercises for this type of writing.

Conclusions and implications

Analysis of the four sources of material indicated a significant match between the requirements of the IELTS GT Modules, as represented by *Cambridge IELTS 3* and *IELTS Specimen Materials 2003*, and typical English Preparation for Secondary School Programs in Australia for students aged 16/17, provided by study of *Cutting Edge Upper-intermediate* and *Insight into IELTS*.

More specifically, students who have had the opportunity to use the full range of materials from both of these textbooks would be comprehensively prepared for the IELTS examination. They would have had exposure to the language, question types and topics which commonly appear in IELTS General Training Modules and

would have been able to practise examination skills and techniques. However, programs preparing students for secondary school admission would do well to supplement their teaching of essay writing skills if they are to depend on *Cutting Edge Upper-intermediate* for their teaching material.

There are, however, some wider implications to the findings from this study. From the IATM analysis it can be concluded that large numbers of students preparing for secondary study on-shore in Australia will find the IELTS General Training Modules accessible; this provides a response to criticism that IELTS is inappropriate for use in this context on grounds of unfamiliarity or inaccessibility. Indeed, the inclusion of *Insight into IELTS* on preparation courses may suggest there has already been some washback from secondary schools requesting an IELTS score from applicants.

Further investigation is desirable, however, to determine whether 16–17 year old students who are being prepared off-shore for upper secondary study in Australia are receiving the same sort of preparation; a complementary study would provide insights into typical materials used in secondary preparation off-shore and their match to the IELTS General Training Modules.

This study forms an important part of the extensive program of

Impact Studies being conducted by the IELTS partners to investigate the impact of IELTS from various perspectives (see previous articles in *Research Notes* 6 and 15); this program of research is designed to confirm the ‘usefulness’ (i.e. the appropriate and effective use) of IELTS in specific contexts and for specific purposes.

References and further reading

- Cambridge ESOL (2003) *IELTS Specimen Materials*, Cambridge: Cambridge ESOL.
- Cunningham, S and Moor, P (2002) *Cutting Edge Upper-intermediate Student's Book*, London: Longman.
- Jakeman, V and McDowell, C (1999) *Insight into IELTS: The Cambridge IELTS Course*, Cambridge: Cambridge University Press.
- Saville, N and Hawkey, R (2004) The IELTS Impact Study: Investigating washback on teaching materials, in Cheng, L, Watanabe, Y and Curtis, A (Eds) *Washback in Language Testing: research contexts and methods*, London: Lawrence Erlbaum, 73–96.
- UCLES (2002) *Cambridge IELTS 3 Student's Book with Answers: Examination Papers from the University of Cambridge Local Examinations Syndicate*, Cambridge: Cambridge University Press.

IELTS Writing: revising assessment criteria and scales (Phase 4)

GRAEME BRIDGES, EXAMINATIONS AND ASSESSMENT GROUP
STUART D SHAW, RESEARCH AND VALIDATION GROUP

Introduction

The project to revise the assessment criteria and rating scale descriptors for the IELTS Writing Test began in 2001 with three main objectives:

1. the development of revised rating scales, including description of assessment criteria and band descriptors
2. the development of materials for training trainers and examiners
3. the development of new certification/re-certification sets for examiners.

It was envisaged that the phases of the Writing Revision Project would closely parallel those of the Speaking Revision Project and that the project would be divided into the following five phases with a tentative schedule as shown in Table 1.

This schedule proposed that development work would begin in January 2002 for full operational implementation in January 2005.

Phase 1 of the project – *Consultation, Initial Planning and Design* – involved consultation with a range of stakeholders and was completed in December 2001. In addition to a review of routinely collected score and performance data for the operational Writing Test, Phase 1 involved a survey of commissioned and non-commissioned studies relating to IELTS Writing (including studies funded under the IELTS funded research program), as well as a

Table 1: Revision Project Plan

| | | |
|---------|---|---------------------------|
| Phase 1 | Consultation, Initial Planning and Design | June 2001 – December 2001 |
| Phase 2 | Development | January 2002 – May 2003 |
| Phase 3 | Validation* | June 2003 – May 2004 |
| Phase 4 | Implementation (incl. examiner re-training) | June 2004 – December 2004 |
| Phase 5 | Operation | January 2005 onwards |

* The Validation Phase included examiner marking trials, investigations into proposed changes to writing rubrics and the preparation of training materials in readiness for the Implementation Phase.

review of the literature on holistic and analytic approaches to writing assessment. Another key component of Phase 1 was a stakeholder survey to investigate IELTS rater attitudes and behaviour; the purpose of this was to highlight theoretical and practical factors which might inform redevelopment of the writing assessment criteria and scales. The phase revealed several important issues from the perspective of the Examiner, more particularly, individual approaches and attitudes to IELTS Writing assessment, differing domains (Academic and General Training) and differing task genres (Task 1 and Task 2) – all of which provided a valuable focus for the subsequent re-development of existing rating scale criteria.

Following completion of Phase 1 in Dec 2001 (reported in *Research Notes 9*), Phase 2 – the *Development Phase* – focused on redeveloping the assessment criteria and reconstructing the band descriptors for the rating scales. The combined use of quantitative methodologies (application of draft criteria and scales to sample language performance) and qualitative methodologies (insightful and intuitive judgements derived from ‘expert’ participants) informed the re-construction of assessment criteria and scales for the IELTS Writing Test.

A number of key revision areas were identified during the *Development Phase* including the assessment approach (the benefits of analytical assessment in relation to the IELTS examination and the consequent move towards compulsory profile marking of all tasks in July 2003); assessment criteria (the five revised criteria for both Modules and both Tasks); rating scale descriptors (evolved through a succession of iterative drafts and fine tunings); examiner training (the implementation of new training systems to re-train all writing examiners subsequent to any revision and the requirement for all examiners to re-certificate). Phase 2 of the project was completed in May 2003. Work accomplished during this phase was reported in *Research Notes 10* and 15.

The third Phase of the project – the *Validation Phase* – began in June 2003 and was completed in October of the same year (see *Research Notes 16*). A detailed and thorough validation programme – employing both qualitative and quantitative methods for the establishment of validity, reliability, impact and practicality – engendered confidence in the revised assessment criteria and band level descriptors. The quantitative and qualitative dimension involved the collection and subsequent analysis of data, in the form of examiner scores and questionnaire responses, from a rating trial comprising senior IELTS examiners in the UK and Australia. Studies confirmed that the revised criteria and scales were functioning in order to confirm their suitability for release into operational mode. A secondary consideration was to identify any issues which might need to feed into the design of the revised rater training program implemented during Phase 4 of the project.

Phase 4 – The Implementation Phase

This article will describe the *Implementation Phase* of the project (Phase 4). We begin by describing the contents of the new Training Materials, followed by a brief overview of examiner training. Details of Certification will also be given. Aspects of the *Professional Support Network (PSN)* for IELTS test centres and examiners are highlighted in the next section and the article concludes with a consideration of training-related issues such as assessment method, revised rubrics and script legibility.

Phase 4 began in June 2004 with the first training session of Senior Trainers in Sydney, Australia followed by a subsequent session in the UK. In a similar manner to the examiner re-training for the Speaking revision in 2001, training is cascaded from Senior Trainers to Examiner Trainers who in turn train the Examiners. Phase 4 will be completed in December 2004 in time for Phase 5 – the *Operation Phase* from January 2005.

Contents of the Training Materials

The *Examiner Training Materials 2004* (Writing) folder as well as the *Instructions to IELTS Examiners Booklet* (Writing Module) provide trainers with detailed information on the revised assessment criteria and the band level descriptors. The Training Materials include a PowerPoint presentation containing the revised criteria in the form of a super-ordinate question that asks what this criterion is fundamentally addressing, followed by three or four supplementary questions that ask more specific questions about the focus of each criterion. Core terminology (key indicators) of each criterion are provided in order to clarify the meaning of all the terms through reference to example scripts.

The five revised criteria for both Modules and both Tasks are:

- Task Achievement (Task 1) / Task Response (Task 2)
- Coherence and Cohesion (Task 1 and 2)
- Lexical Resource (Task 1 and Task 2)
- Grammatical Range and Accuracy (Task 1 and Task 2).

An example definition for one of the five criteria – *Lexical Resource* (Task 1 and Task 2) follows:

Lexical Resource refers to the range of vocabulary that the candidate shows an ability to use, and the precision with which words are used to express meanings and attitudes. Lexical Resource also covers aspects of mechanical accuracy including spelling and appropriacy of vocabulary, key indicators of which include the use of vocabulary of an appropriate register; collocational patterns; accuracy of word choice, and controlled use of word relations such as synonymy and antonymy.

Descriptions have been developed for each criterion at each of the nine band levels. A descriptor for one of the higher bands on the *Lexical Resource* rating subscale is given below as an illustrative example of a band level description:

A wide resource is fluently and flexibly used to convey precise meanings.

There is skilful use of uncommon and/or idiomatic items when appropriate, despite occasional inaccuracies in word choice and collocation.

Errors in spelling and/or word formation are rare and do not reduce the communicative effect.

Training and Certification

Training for new IELTS examiners covers a period of 1.5 days for the Writing Module. Certification is a process that occurs after examiner training and consists of a rating exercise. It is designed to ensure that new examiners rate to standard. Ideally, Certification is done on the afternoon of Day Two, but may be scheduled by the test centre administrator to be done on-site, under supervision, at another time within two weeks of the initial training for logistical reasons. All examiners worldwide must re-certificate every two years for the modules for which they are trained and certificated.

All examiners will have signed a Code of Practice/Code of Conduct document demonstrating their awareness of the standards of professional behaviour required of IELTS examiners; examiners may also be required by test centre administrators to sign a contract of employment that reflects the local situation. All new examiners will receive support in their first three test administrations (ideally, the first scheduled test at the centre after training). Support will include test day induction, mentoring, standardisation, and double marking. A maximum of six Writing scripts will be rated in one day by new examiners, for these three sessions.

Training is organised so that trainees are exposed to exemplar scripts highlighting key features in the revised band descriptors as well as common problems such as scripts being underlength, off-topic or memorised. On the first day trainees are asked to rate Task 2 (Academic and GT) and GT Task 1 scripts. Initially the trainer uses a Standardisation Script whereby trainees are asked to identify the features that constitute a certain profile. Several scripts are then group rated and discussed until trainees have the confidence to rate scripts alone.

A Homework Pack is distributed at the end of Day One which not only reinforces the knowledge and skills acquired but also prepares trainees for Day Two. The second day initially consists of rating Academic Task 1s. An Examiner Quiz follows which runs through some of the key issues. Before filling in a Feedback Questionnaire the trainer presents a short presentation on the *Professional Support Network*.

All certification and re-certification is conducted on-site, under supervision. Administrators have clear guidelines on the appropriate method of conducting certification and re-certification, including:

- for reasons of security, materials are not allowed to be taken off-site
- examiners must rate 12 writing scripts for certification and re-certification
- examiners may have two attempts at certification and re-certification.

Professional Support Network

The IELTS PSN is a global system that integrates all aspects of IELTS Examiner recruitment, training, certification, standardisation, monitoring and conduct. The PSN operates on a worldwide basis, to cover all aspects of the work of IELTS examiners. It has been developed and is supported by the three IELTS partners: IDP: IELTS Australia, British Council, Cambridge ESOL Examinations. It seeks to ensure that:

- all stages of the processes of IELTS examiner management are transparent, ethical, thorough, and well supported by appropriate documentation;
- the roles of Examiners, Examiner Trainers (ET) and Examination Support Co-ordinators (ESC) are defined in the documentation to provide adequate professional support for examiners and support for the efficient functioning of test centres.

Professional support for examiners is crucial in the maintenance of reliability and standardisation among IELTS examiners.

The differing support needs of new and continuing examiners are also recognised in the PSN. The PSN is designed to keep examiners informed about the IELTS test and about their role as examiners. Effective performance is acknowledged through feedback, as are any problem areas.

Monitoring and Standardisation

Cambridge ESOL currently conducts routine sample monitoring worldwide for the purpose of calculating rater reliability. Targeted monitoring occurs in response to jagged profiles and Enquiries on Results. From 2005, centres are to conduct scheduled in-centre monitoring of each examiner every 2 years. In-centre monitoring of examiner performance will occur in the alternate year from that in which the examiner is due to re-certificate. Thus, every examiner will be monitored once every two years. Useful and constructive feedback will be provided to the examiners. Feedback will acknowledge effective, standardised performance as well as providing support in problem areas as necessary.

From 2005, all examiners are required to participate in a standardisation session before taking re-certification. Standardisation sessions require one half day for the Writing module, i.e., four hours, and will be face-to-face where possible; in centres where an ET cannot be present, other modes of delivery may be used, for example, video conferencing, or Internet delivery. In addition, examiners will have access on-site to *Self-Access Standardisation* materials for Speaking and for Writing (which replace the current *Writing Assessment Guidelines – WAG*), which they may use for extra standardisation in response to feedback after monitoring.

Training-related Issues

During the Writing Revision Project the issues of assessment approach, revised rubrics, guidelines for word counts and script legibility were raised and discussed. These issues had implications for Phase 4 training, training materials and related documentation.

Assessment Approach

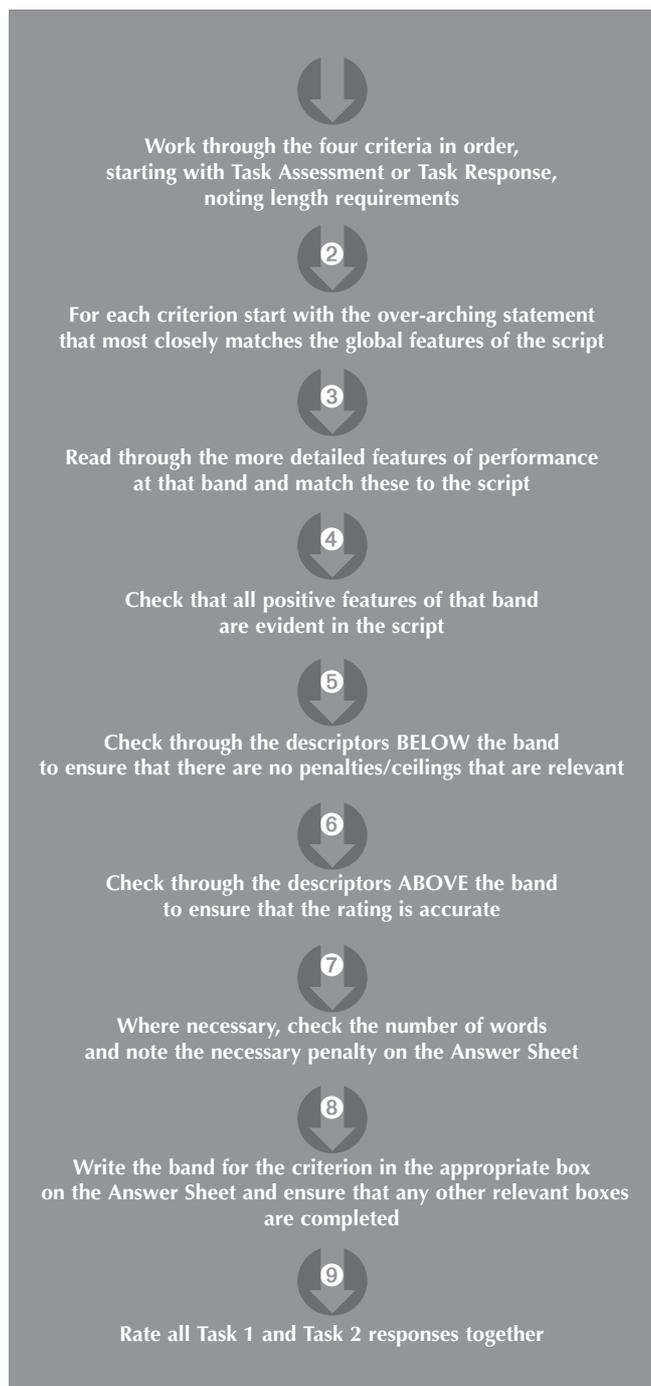
The revised method of assessment can be represented pictorially and is summarised in Figure 1.

Revised Rubrics

Language test instructions play an important role in test takers' performances; their overall performance depends upon how well they understand the test conditions under which the test will be administered, the actual test procedures to be followed and the nature of the tasks they are to undertake.

From a validation perspective, consultative exercises with stakeholders (examiners, teachers and candidates) elicit a greater understanding of the nature of the impact of the changed rubrics on the IELTS assessment community as well as engender confidence in the new rubrics. Qualitative findings from two consultation exercises with IELTS preparation candidates and experienced IELTS examiners at the Anglia Polytechnic University (APU) in March and April 2004 constituted an attempt to establish confidence – on the part of the stakeholder community – in the new rubrics.

Figure 1: Revised method of assessment for IELTS Writing



Trialling the rubric changes with IELTS preparation candidates and examiners has shown that the changes have ensured that instructions are clearer for both groups. Candidates and examiners alike have stated that the revised rubrics appear to be less ambiguous than their current counterparts. The proposed changes to the IELTS Academic Writing rubrics satisfy – at least from the stakeholder perspective – Bachman and Palmer’s (1996) three essential guidelines for the writing of instructions. The revised rubrics have been shown to be:

- simple enough for candidates and examiners to comprehend – having greater clarity than their current counterparts
- short enough so as not to take up too much of the test administration time and, more importantly

- sufficiently more explicit than their current counterparts allowing candidates to know exactly what is expected of them.

The principal rationale for the rubric changes has been both to clarify the test instruction in response to expressed concerns and perceived confusion among the stakeholders about the test, and to reflect current trends in EAP pedagogy. In the case of Academic Writing Task 1 and Task 2, EAP and tertiary preparation courses have developed significantly since the current rubrics were first formulated. The revised rubrics will be introduced in January 2006 and full information will be provided to stakeholders prior to their introduction.

Guidelines for word counts

It was agreed that guidelines for examiners were required regarding word counts for underlength scripts. Fully copied rubrics, or sentences that copy the rubrics with minimal word order or lexical alterations, it was felt, should be discounted from the word count. Words that are normally written as one word but that have been written by the candidate as two words (*some times, can not*) should count as one word because this is due to candidate error. Moreover, compounds that are normally written as two words (*rail station*) but have been written by the candidate as one word, should count as one word (again this acts as a penalty for error). Words that can be written as one or two words, depending on the dictionary used or on general usage, should be counted as they are written on the answer sheet by the candidate (*life style* = 2 words/ *lifestyle* = 1 word). Other guidelines include:

- hyphenated words and contractions should count as one word
- in General Training Task 1, the salutation and name at the end of the letter should not form part of the word count
- numbers (currency, percentage, temperature, distance, weight) count as one word if they are written using numbers and symbols (*15,000,000, 15m, £20, 20%*) and two or more words if they are written using numbers and words (*15 million, 20 pounds, 20 percent*)
- symbols/abbreviations (& or e.g.) count as one word
- dates count as one (*3/5/2004*), two (*June 1995; 23rd December*) or three (*23rd of December*) words
- titles and headings are not included in the word count.

Underlength responses to all IELTS Writing Tasks continue to be penalised using the revised scales. A sliding scale system – communicated to examiners as an explicit statement of quantification and located outside the band descriptors – has been imposed where a fixed penalty is applied to a response comprising a word length falling within a specific range. Examiners will now be trained to locate a response within the band descriptors and then invoke a penalty from a list of specific penalties for varying degrees of ‘underlengthness’.

Script legibility

Cambridge ESOL’s position on illegible scripts is that it is rare for a script to be totally illegible i.e. impossible to read or understand. If an examiner considers an IELTS script to be totally illegible they

should award Band 1 for all four sub-scores. Such scores would be flagged up and treated as a jagged profile. The Administrator should then pass the script to two other examiners for further scrutiny. Partially legible scripts can only be rated against retrievable language and self-penalise under *Coherence and Cohesion*. Unreadable sections will not be discounted under word count.

Conclusion

Comments from trainers at both the initial sessions in July 2004 and subsequent training sessions reveal a very favourable response to the revised rating scale and the *Examiner Training Materials*. The new band descriptors are seen as being clearer and much more user-friendly. The new rating scale alongside the improved training, certification, monitoring and standardisation procedures under the auspices of the *Professional Support Network* should

ensure greater reliability of rating when the new revised scale is operational from January 2005.

Reference

- Bachman, L and Palmer, A (1996) *Language Testing in Practice: Designing and Developing Useful Language Tests*, Oxford: Oxford University Press.
- Shaw, S (2002a) IELTS Writing: revised assessment criteria and scales (Phase 1), *Research Notes* 9, 16–18.
- (2002b) IELTS Writing: revised assessment criteria and scales (Phase 2), *Research Notes* 10, 10–13.
- (2004a) IELTS Writing: revised assessment criteria and scales (concluding Phase 2), *Research Notes* 15, 9–11.
- (2004b) IELTS Writing: revised assessment criteria and scales (Phase 3), *Research Notes* 16, 3–7.

Set Texts in CPE Writing

DIANA FRIED-BOOTH, CONSULTANT FOR CAMBRIDGE ESOL

Background

The Certificate of Proficiency in English (CPE) is the highest level examination within the Cambridge ESOL Main Suite of examinations. It is linked to the Common European Framework and corresponds to C2 level within this Framework (see table on page 2). At CPE level, a learner is approaching the linguistic competence of an educated native speaker; success at this level is regarded as evidence that the learner has the ability to cope with high-level academic work and so CPE, similarly to the IELTS Academic Module, is recognised by many organisations for university entrance or professional membership (see recognition tables on Cambridge ESOL's website).

One of the distinguishing, long-established features of CPE Paper 2, the Writing paper, is the inclusion of set texts in Part 2 of the paper; candidates choose one question from four in this Part, one of which offers three set text options. This option is intended to encourage extended reading as a basis for enriching language study, at the same time as enabling students and teachers to develop a shared love of language through literature.

Choice of texts

The choice of set texts on the CPE syllabus must take into account the fact that the candidature represents about 90 countries and the choice of texts must be sensitive to a range of international cultures. At this level it would be a gross disservice not to acknowledge the wealth of literature written in English by writers of many different nationalities. Moreover, it can no longer be assumed that students learning the language know anything about, or have any interest in, British culture; if set texts assume a common cultural heritage, certain cultures are bound to be alienated by these assumptions. Inevitably there are constraints in

selecting suitable texts for an international candidature, and factors which govern the selection must constantly be borne in mind:

- Is the text enjoyable and worth reading?
- Is the length of text appropriate and the quality of writing acceptable?
- Is the text suitable for the age range? (almost 75% of candidates are 25 years of age or under)
- Are there any issues which could offend or distress? (bearing in mind cultural differences and sensitivities)
- Is the text available and affordable?
- What is the status of the writer – how is he or she regarded internationally?
- Is the genre, for example, crime, science fiction etc. likely to enhance the choice when included with other set texts?
- Will the text generate a range of potential questions? (a text remains on the syllabus for 2 years)
- Is related material available in the form of film, audiobook, support notes etc?

It is unlikely that any one single text can meet all of the above criteria, but a combination of three different texts is always available on each Paper.

Procedure for selecting texts

When a shortlist of suggested texts has been drawn up, relevant people such as item writers for the Paper and CPE teachers are asked to read and report on a text. In addition to the criteria which underpin the selection, they will also consider other factors:

- The suitability of the content and theme(s) in the text
- The level of language difficulty

- The likely appeal of the text
- Any undue culture bias in the text.

In conclusion, readers can strongly recommend a text, indicate that they consider it to be reasonably suited or reject it as being unsuitable. Once the comments from readers' reports have been collated, the final decision to include a text is made by the Subject Officer in conjunction with others working on the Writing Paper.

Set texts from 1913 onwards

It is interesting to look at the shift in the selection of set texts which has taken place over the last century and continues to take place.

CPE was introduced in 1913 and during the following decades, not surprisingly, an anglocentric choice of authors predominated: for example, Matthew Arnold, Jane Austen, Charles Dickens, William Shakespeare. In 1934 'questions were framed to give candidates an opportunity of showing whether they can understand and interpret representative works of literature' (Weir 2003:4).

The last twenty years have seen a gradual change to include a far greater range of authors: for example, Ruth Praver Jhabvala, Timothy Mo, Alan Paton, Kazuo Ishiguro, Anne Tyler, Brian Moore. The aim is still to give candidates the opportunity to read a range of genres written in English and to show in their writing that they have appreciated themes, characters and relationships within that work. Questions do not focus on literary or stylistic analysis, although candidates may allude to stylistic devices in their answers.

In the last few years the change in types of set texts chosen has accelerated. The choice of the three set texts on each Paper, however, always aims to achieve a balance between what may be regarded as a classic set text, such as LP Hartley's *The Go-Between* and that of a classic contemporary writer such as Chinua Achebe's *Things Fall Apart*. The selection for 2006 illustrates this approach and includes: *An Inspector Calls* by JB Priestley, *Bel Canto* by Ann Patchett and *Clear Light of Day* by Anita Desai.

Assessment of set text answers

Set text questions require a candidate to respond by writing between 300–350 words in a specified form: one of an article, essay, letter, review or report. Some teachers and candidates perceive a tension in the assessment of a candidate's response to literature in an examination which focuses on a candidate's ability to communicate. It is recognised that although this reservation may not prevent the study of a set text in class, it may inhibit a candidate opting to answer a set text question in the examination.

Examiners make reference to both General and Task Specific Markschemes in assessing a task. A Task Specific Markscheme will clearly relate to the content of a specific question and focus on how effectively the candidate has realised the task.

It is taken for granted that a candidate who chooses to answer a set text question will provide clear evidence of having read that text. In developing a response to a question, a candidate is expected to support their writing with reference to the text,

providing relevant examples to illustrate their argument. At CPE level, a set text question will never merely expect a candidate to retell the plot or narrative. A candidate therefore needs to recognise the importance of answering a question appropriately, by marshalling their ideas and understanding of what they have read in order to produce a relevant response. Assessment is based on language output, not on literary analysis skills, and an examiner's judgement is based on control of language demonstrated in the way in which a candidate responds to the points raised by a specific question. The criteria which underpin the assessment of a set text response are those which underpin all tasks on the Paper:

- Content
- Organisation and cohesion
- Accuracy in the use of structures and vocabulary
- Register of structures and vocabulary
- Register and format
- Awareness of role of writer
- Awareness of role of reader.

Advantages to be gained from studying set texts at CPE level

Students and teachers alike testify to the benefits gained by studying a set text, even if that option is not taken up in the actual examination. Whether in the classroom or by studying privately, there are many advantages, both tangible and intangible, which result from choosing to pursue a set text as part of a CPE course, including:

- To provide a source of pleasure and stimulation
- To encourage a lifelong interest in reading
- To widen cultural, linguistic and literary horizons
- To promote contact with other media – film, TV, radio, theatre
- To engage the reader on a subconscious level in terms of structure, vocabulary, collocation which may also have an influence on the reader's own writing skills
- To create the potential to exploit literature on more than one level by studying vocabulary, collocation, structure etc. in an extended context
- To give rise to classroom activities such as reading a play aloud.

The future of set texts in CPE

The very nature of this part of the CPE syllabus means that there is on-going consideration of ways forward in terms of text selection for the future. Other genres are under discussion including screenplays, comic books or graphic novels and writers in translation. In the fast-moving world of electronic communication it is vital that a set text syllabus can remain relevant and attractive to the CPE candidature. Cambridge ESOL strives to maintain the relevance of the format and content of all of its language tests for its stakeholders, whilst embracing new technologies and trying to maintain the historical traditions of our language tests, of which CPE is the oldest.

Suggestions for suitable set texts for CPE are very welcome and teachers and anyone else interested are encouraged to send in their ideas either by visiting the Cambridge ESOL website www.CambridgeESOL.org or by contacting the ESOL Helpdesk on ESOL@ucles.org.uk

References and further reading

Weir, C (2003) *Continuity and Innovation: Revising the Cambridge Proficiency in English Examination 1913–2002* (Studies in Language Testing, volume 15), Cambridge: UCLES/Cambridge University Press.

UCLES (2003) *Certificate of Proficiency in English Handbook*, Cambridge: UCLES.

IELTS – some frequently asked questions

IELTS test users often have specific questions about how IELTS is designed or the way it works; answers to the most common questions can be found in the current IELTS Handbook and on the IELTS website www.ielts.org/candidates/faq. Here are some additional questions which are sometimes raised by stakeholders.

? Why do candidates have to wait 3 months before being able to resit IELTS?

Current policy means that candidates who have not achieved the band score(s) they need must wait 90 days before retaking the IELTS test though this policy is under review. A major reason for policy to date has been to avoid candidates retaking the test to no good purpose. Recent research into score gains in the UK, Australia, New Zealand and elsewhere, much of which has been carried out with grant funding from the IELTS partners, indicates that candidates are unlikely to show any real improvement in language ability and test performance without further study; three months is the minimum period likely to yield substantive improvement. Furthermore, studies show evidence of considerable individual variation in rate of gain due to the complex interaction of multiple factors relating to the learner and context. Any gains that are made are more likely to be achieved among lower proficiency level candidates (i.e. in the band 4–5 range) than those higher up the proficiency continuum (i.e. in the band 6–7 range). (See Anthony Green's PhD, 2003; also his review article in *Research Notes* 16.)

? Why does an IELTS Test Report Form have a recommended 2-year validity period?

The IELTS Handbook recommends that a Test Report Form which is more than two years old should only be accepted as evidence of present level of language ability if it is accompanied by proof that a candidate has actively maintained or tried to improve their English language proficiency. This recommendation is based upon what we know about the phenomenon of second language loss or 'attrition', a topic which is well-researched and documented in the literature.

The level of second language competence gained and the extent of opportunity for subsequent practice both affect how much language is retained or lost over a period of time. Research points to two types of attrition. At lower proficiency levels, rapid

language loss occurs soon after the end of language training/exposure (for approximately two years) and then levels off leaving a residual competency (Bahrck 1984; Weltens 1989); at higher proficiency levels the reverse pattern can be observed (Weltens and Cohen 1989) – a few years of non-attrition (an 'initial plateau') followed by steady loss. It appears that a critical period exists after disuse; although the nature of this may differ for high and low proficiency users, a two-year limit has been selected as a reasonable 'safe period'.

The two-year period also parallels ETS recommendations for the use of TOEFL scores (used in a similar way to IELTS): ETS suggests that non-native speakers who have taken the TOEFL test within the past two years and who have successfully pursued academic work in an English-speaking country for a specified minimum period of time (generally two years) with English as the language of instruction may be exempted from providing TOEFL test scores.

? Why can't the IELTS modules be taken as separate tests?

IELTS is designed to assess a candidate's *overall* English language proficiency within a specified time-frame. This is achieved by asking candidates to provide evidence of their reading, listening, writing and speaking abilities at a certain point in time: the Listening, Reading and Writing modules are administered on the same day; for logistical reasons the Speaking module can be administered up to 7 days before or after the other components. The four component modules are not offered as separate tests to be taken at different times; in this sense IELTS is not a modular test.

Performance in the four skill areas is combined to provide a maximally reliable composite assessment of a candidate's overall language proficiency at a given point in time. Scores on the four component modules are computed to provide an overall band score; the four component scores are also reported separately for their diagnostic value, to indicate a candidate's relative strengths and weaknesses.

? In what ways can the IELTS test be described as 'integrated'?

The term 'integrated' is sometimes used to refer to different features or qualities of testing procedures or test tasks, e.g. cloze tasks have been described as 'integrative' as opposed to 'discrete-point'.

A more common approach today is to talk about testing 'integrated skills'; this usually means that completion of a test task involves using more than one macro-skill, e.g. a speaking or writing task depends upon the test-taker processing some associated reading and/or listening input. The term 'integrated' may also be used to suggest that test tasks bear a close resemblance to 'real-life' language activities, i.e. the content is based on authentic language (however defined), and the task mirrors features of everyday 'communicative' language use which the test-taker would carry out in a non-test context. An extension of this idea is that because such tasks are 'integrated', they can provide a realistic and useful measure of how well people will communicate in a particular setting (e.g. workplace, academic); a further claim is sometimes made that a test which reflects an 'integrated approach' will help test-takers prepare appropriately for future success in that particular setting – though predictive validity studies have shown that 'future success' can depend on many different factors in addition to language proficiency.

IELTS (and ELTS before it) has always been a test to which the term 'integrated' could be applied on various counts. For example, IELTS has always included testing of the four skills – Listening, Reading, Writing and Speaking; profile scores on the four modules are reported separately and also contribute equally to an overall band score. Furthermore, although each module focuses on a particular skill, test tasks often entail the use of other skills and are thus 'integrated' to some degree. This is most apparent in the Writing and Speaking modules where information which is read or heard helps shape the candidate's own production.

For example, Task 1 of the Academic Writing Module gives candidates some visual input (a diagram or table) and asks them to present the information in their own words. Task 1 of the General Training module involves reading a short prompt about a particular problem and using the information it contains to write an appropriate letter of response. Task 2 for both modules presents a point of view, argument or problem which candidates must read and respond to in their writing. All tasks contain some indication of audience and purpose for writing.

The face-to-face Speaking module clearly involves listening skills as well as speaking ability; the examiner frame constrains the listening input to make it fair and accessible for all candidates. In Part 2 candidates are given a prompt to read on a card; they are also given one minute of preparation time and invited to make written notes if they wish. All these task features reflect a degree of 'integratedness'.

Tasks in the Writing and Speaking modules are designed to achieve a careful balance between two factors: on one hand, providing adequate support for the test-taker in terms of task content and level of language needed to access the task (level of input is constrained at the test-writing stage); and on the other hand, the opportunity for the test-taker to 'engage' with the task by drawing on their personal experience, opinions, creativity, etc. in demonstrating their language proficiency. This is another way of defining the notion of 'integratedness'.

Tasks in the Reading and Listening modules can involve note-taking, labelling, classification, and table/flowchart completion. What is important is that any task (or test items) should be

consistent with a likely focus for reading/listening to the text(s) and should encourage test-takers to engage in appropriate cognitive processes. Once again, we could argue that such tasks are 'integrated' in terms of the relationship between the input and the cognitive processes they elicit. Validation studies help to confirm the match between task input, cognitive processing, and task output.

While IELTS tasks are designed to reflect certain features of university-level tasks, they do not set out to 'simulate' tasks which students will need to do in their university studies. Constraints of time are one reason for this: an IELTS reading test lasts only 1 hour – a typical university task normally takes much longer. More importantly, IELTS assumes *readiness to enter* a particular domain; it does not assume that mastery of study skills has already been achieved (see further discussion below). Test tasks are designed to balance the requirements of validity, reliability, impact and practicality, the four essential qualities which underpin the Cambridge ESOL approach.

As Davies et al (1999) point out, 'integration' can have its limitations; scores derived from tasks which combine different aspects of ability may be difficult to interpret – does this task measure *writing*, *reading*, or something called *reading/writing*? Some of the problems associated with a strongly integrated-skills approach are discussed in the next section.

Why isn't there a link between the Reading and Writing modules?

Until 1995, a strong thematic link existed between the Reading and Writing modules (for both Academic and General Training). This link was removed in the 1995 IELTS Revision Project because it increased the potential for confusing assessment of writing ability with assessment of reading ability (Charge and Taylor 1997). Monitoring of candidates' writing performance suggested that the extent to which they exploited the reading input varied considerably. Some candidates drew heavily on the written content of the reading texts, apparently treating the writing task as a measure of their reading ability; as a result many risked masking their actual writing ability. Other candidates chose to articulate their own ideas on the topic, making very little reference to the reading input or forging artificial connections for the sake of the task. In some cases, cultural background meant that candidates were confused about whether to articulate their personal point of view on a topic or to reflect the more 'authoritative' view expressed in the reading text(s).

Such variation in response to the linked reading/writing task made the achievement of fair assessment at the marking stage very difficult. Removal of the link between the IELTS Reading and Writing modules resulted in a more equitable form of task design. It also made it easier to control comparability of task difficulty across the many different test versions which need to be produced each year to meet the demands of candidature volume and security. (An earlier link between the field-specific Reading/Writing modules and the Speaking module had been removed as part of the ELTS/IELTS Revision Project in 1989 for reasons explained in Alderson and Clapham 1993, Clapham and Alderson 1997).

? Why aren't the IELTS Academic Reading and Writing tasks more like university-level tasks?

IELTS is designed to test *readiness to enter* the world of university-level study in the English language and the ability to cope with the demands of that context immediately after entry. It does not assume that test-takers have already mastered (or even partially acquired) the range of university-level reading or writing skills which they are likely to need; in fact, they will probably need to develop many of these skills during their course of study, often in ways that are specific to a particular academic domain. The implication of this is that IELTS Academic Reading and Writing tasks cannot simulate the sort of university-level tasks which test-takers will encounter in their studies. It would be unreasonable to define the 'authenticity' of IELTS Academic Reading and Writing tasks purely in terms of 'simulated university-level tasks' and then to judge them against that criterion.

Instead, tasks are designed to be accessible to a wide range of test-takers (irrespective of their academic discipline) and to reflect features of writing activities that are already familiar to candidates from previous study experience as well as some general features of writing they may encounter in subsequent study. An essay format is used for Writing Task 2 precisely because it is a written genre widely used in both secondary and higher education contexts. Moore and Morton (1999) describe the essay as the predominant written genre used in university study and their study demonstrated empirically that IELTS Academic Writing Task 2 does share features in common with this format.

? Is IELTS culturally biased?

All the texts and tasks in the IELTS modules are designed to be widely accessible and to accommodate as far as possible candidates' prior linguistic, cultural and educational experience irrespective of nationality or first language. Removal in 1995 of the thematic link between the Reading and Writing modules was in part for this reason (see above). Topics or contexts of language use which might introduce a bias against any group of candidates of a particular background (e.g. due to gender, ethnic origin) are avoided at the materials writing/editing stage. Pre-testing procedures prior to live test construction monitor feedback on texts and topics and so provide another safeguard in this regard. An external study by Mayor, Hewings, North and Swann (2000) which investigated the written performance of different L1 groups found no evidence of significant cultural bias due to task.

? Why aren't Speaking and Writing scores reported using half bands, like Reading and Listening?

In their original design, the IELTS/IELTS Speaking and Writing modules were rated using a holistic 9-band scale. Examiners were trained and standardised to match a candidate's writing or speaking performance to one of 9 'band descriptors', i.e. described levels of performance. Examiners were not trained to differentiate quality of written/spoken performance within a given band level using half bands. The ability of examiners to differentiate a larger number of performance levels, and to do so

reliably, is partly constrained by the nature of the judgement process in assessment.

The introduction of revised, analytical (rather than holistic) scales for Speaking in July 2001 and for Writing in January 2005 would make it possible to *report* scores on these two modules in terms of half bands (though examiners will continue to *rate* using whole bands for each of the analytical subscales).

A proposal to report half bands for Speaking and Writing in future is currently under consideration; several internal studies have already been carried out to model the effects of a move to half band reporting and to evaluate the impact on mean band scores for Speaking, for Writing, and for the test overall. The main benefits of changing to half band reporting for Speaking and Writing would be two-fold:

- the final Writing or Speaking band score could more sensitively reflect the quality of performance captured through the analytical rating process;
- the reporting of half bands for Writing and Speaking (in line with existing practice for Reading and Listening) could be helpful to test users, especially if they use Speaking or Writing profile band scores (as well as the overall band score) for decision-making purposes.

The earliest such an approach could be implemented is January 2007. This takes into account the need to complete ongoing validation and technical studies; it also allows time to prepare the stakeholder community for a change in the way IELTS scores are reported.

? Is there a risk of gender bias in the IELTS Speaking Test?

O'Loughlin (2000) used the IELTS Oral Interview to investigate the potential impact of gender in oral proficiency assessment and found no evidence that the test was a strongly gender differentiated event. He concluded that IELTS interviewers and candidates 'generally adopted a more collaborative, co-operative and supportive communicative style irrespective of their gender or the gender of their interlocutor' (p 20). Furthermore, he found no empirical evidence of significant bias due to rater/candidate gender with regard to the rating process and the scores awarded. The introduction of the revised IELTS Speaking Test in 2001 was partly to minimise even further any potential for examiner language or behaviour during the test to be a source of bias.

? Is IELTS suitable for younger students below the age of 18?

IELTS/IELTS was originally designed as an English language proficiency test for students who had already completed their secondary education and who wished to undertake further academic study in an English-speaking country, at first degree or post-graduate level. In this sense it was targeted at adults, i.e. those in their late teens or above. This is particularly true for the Academic modules (Reading and Writing) which tend to assume a level of cognitive maturity normally not achieved until early adulthood.

The cognitive demands of the Academic Reading and Writing tasks were demonstrated during a series of native speaker trialling studies conducted in 1993/1994 as part of the 1995 IELTS Revision Project. One study involved administering IELTS subtests to 148 English native-speaker students at sixth-form colleges, universities and technical colleges in the UK and Australia; this sample population included both 16–17 year olds (pre-university) and 18–21 year olds (undergraduate). Results showed that the tests were able to discriminate effectively within the native speaker population: the Listening subtest attracted generally high raw scores – with a mean of Band 8/9; however, the spread of scores for the Academic Reading and Writing modules showed that native speakers responded with varying degrees of success, depending in part on their age and experience.

The IELTS General Training (GT) modules, however, were developed to suit the needs of a slightly different population – those wishing to undertake further study/training of a non-academic, vocational nature, or as a bridge between school and university. Post-test analysis shows that significant numbers of candidates in the younger (16–18 year old) age group take IELTS GT each year; no significant problems have been noted in terms of content or difficulty and analysis of live test performance by age indicates that 16 and 17 year olds perform better than some other age groups (e.g. candidates aged between 18 and 21). This also appears true for Academic candidates and is a phenomenon observed with other Cambridge ESOL exams. One possible explanation is that 16 and 17 years olds are still in full time education so are well used to the demands of studying and test-taking.

A study under the IELTS grant-funded program investigated the performance and attitudes of a specific group of 15–17 year old candidates on IELTS General Training. Merrylees (2003) found that most students in the study coped reasonably well with the demands of the sub-tests in terms of mean scores achieved; students reported finding Listening sections 3 and 4, and Reading section 3 most challenging. An impact study conducted in Australia in 2003 (see Jan Smith's article in this issue) confirmed the accessibility of the IELTS General Training module content to the 16–17 year old population.

The available evidence suggests that IELTS – particularly General Training – is suitable for use with students below 18.



How well does IELTS predict academic success?

Findings from predictive validity studies (which seek to measure the relationship between language proficiency test scores and academic outcomes) are often very mixed, suggesting that the relationship between English language proficiency and subsequent academic success is an extremely complex one. (See the IELTS website for details of IELTS-related studies.) Correlations are often relatively weak, mainly because academic performance is affected by so many other factors, e.g. academic ability/knowledge, the amount of in-session English language tuition received, motivation, cultural adjustment, and circumstances relating to welfare.

It is vital for users of IELTS test scores to set responsible admissions criteria and to have a clear understanding of the

contribution that IELTS scores can make in determining an applicant's suitability for entry, including the relative importance of scores in the four modules for particular academic courses. The IELTS partners are working to help University admissions departments and other test users improve their understanding of the relationship between students' English language proficiency and subsequent performance; this includes building awareness of key influences on academic outcomes and of other factors which need to be taken into consideration, e.g. provision of ongoing language and study skills support for international students, as well as academic and acculturation programs, including appropriate pastoral care.

References and further reading

- Alderson, J C and Clapham, C (1993) (Eds) *Examining the ELTS Test: An Account of the First Stage of the ELTS Revision Project*, *IELTS Research Report 2*, Cambridge: UCLES, British Council and IDP.
- Bahrck, H P (1984) Fifty years of second language attrition: Implications for programmatic research, *Modern Language Journal*, 68, 105–118.
- Charge, N and Taylor, L (1997) Recent developments in IELTS, *ELT Journal*, 51/4, Oxford: Oxford University Press, 374–380.
- Clapham, C and Alderson, J C (1997) (Eds) *Constructing and Trialling the IELTS Test*, *IELTS Research Report 3*, Cambridge: UCLES, British Council and IDP.
- Davies, A, Brown, A, Elder, C, Hill, K, Lumley, T and McNamara, T (1999) *Dictionary of language testing*, Cambridge: UCLES/Cambridge University Press.
- Green, A (2003) *Test Impact and English for Academic Purposes: A Comparative Study in Backwash between IELTS Preparation and University Pre-sessional Courses*, unpublished PhD dissertation, University of Roehampton, Surrey (UK).
- (2004) Making the grade: score gains on the IELTS Writing Test, *Research Notes* 16, 9–13.
- Mayor, B, Hewings, A, North, S and Swann, J (2000) *A linguistic analysis of Chinese and Greek L1 scripts for IELTS Academic Writing Task 2*. Report from study conducted under the IELTS joint-funded research program.
- Merrylees, B (2003) An Impact Study of Two IELTS User Groups: Candidates who sit the Test for Immigration Purposes and Candidates who sit the Test for Secondary Education Purposes, in *IELTS Research Reports Volume 4*, Canberra: IDP: IELTS Australia.
- Moore, T and Morton, J (1999) Authenticity in the IELTS Academic Module Writing Test: A Comparative Study of Task 2 Items and University Assignments, in *IELTS Research Reports Volume 2*, Canberra: IELTS Australia
- O'Loughlin, K (2000) The Impact of Gender in the IELTS Oral Interview, in *IELTS Research Reports Volume 3*, Canberra: IELTS Australia
- Weltens, B (1989) *The attrition of French as a Foreign Language*, Dordrecht/Providence: Foris Publications.
- Weltens, B and Cohen, A D (1989) Language attrition research. An introduction, *Studies in Second Language Acquisition*, 11/2 127–133.
- Website of IELTS FAQs:
<http://www.IELTS.org/candidates/FAQ/>

IELTS test performance data 2003

ANDREW BLACKHURST, RESEARCH AND VALIDATION GROUP

Candidate performance on IELTS and the reliability of IELTS test materials are routinely monitored at Cambridge ESOL and the most recent figures are presented here for 2003 test takers and test materials.

The number of candidates taking IELTS in 2003 was well over 475,000 – a rise of 34% on the previous year's candidature.

IELTS band score information

IELTS is assessed on a nine-band scale. The Overall Band Scores for Academic and General Training candidates in 2003 are reported below for all candidates and by gender (Tables 1 and 2), together with Mean Band Scores for the individual modules, for the most frequent first languages (Tables 3 and 4) and nationalities (Tables 5 and 6). All Mean Band Scores are given in descending frequency of first languages or nationalities. These band scores are in line with expected parameters of performance.

In 2003, just over three-quarters of candidates (76%) took the Academic Reading and Writing modules of IELTS and just under a quarter (24%) took the General Training Reading and Writing modules. Overall, the IELTS candidature during the year was 45% female and 55% male. Of candidates taking the Academic modules 47% were Female and 53% male, while of candidates taking the General Training modules 39% were female and 61% were male.

General Training candidates show greater competence in their Writing and Speaking skills relative to their skills in Reading and Listening. Compared to 2002 data mean band scores for General Training candidates have increased, indicating changes in the composition of the candidature. On average, Academic candidates show less variation across the skills, but find the Writing module most challenging.

Reliability of test material

Each year, new versions of each of the six IELTS modules are produced for each test administration¹. The reliability of listening and reading tests is reported using Cronbach's alpha, a reliability estimate which measures the internal consistency of a test. The

following Listening and Reading material used from September 2003 has sufficient candidate responses to estimate and report meaningful reliability values as shown in Table 7 overleaf.

The figures reported in Table 7 for Listening and Reading modules indicate the expected levels of reliability for tests containing 40 items. The reliability of the Writing and Speaking modules cannot be reported in the same manner because they are not item-based: task responses are marked by trained examiners according to detailed descriptive criteria. The performance of materials in the Writing and Speaking modules is routinely analysed to check on the comparability of different test versions.

Mean band scores for the Academic Writing versions used from September 2003, and for which a sufficient sample size has been obtained, ranged from 5.54 to 6.01. Mean band scores for the General Training Writing versions used from September 2003 ranged from 5.62 to 6.05. Mean band scores for Speaking versions used from September 2003 ranged from 5.94 to 6.27. The analysis for both Writing and Speaking materials shows a consistent pattern across different test versions over time.

Reliability of assessment

Reliability of rating is assured through the face-to-face training and certification of examiners and all examiners must undergo a re-training and re-certification process every two years. Continuous monitoring of the reliability of IELTS Writing and Speaking assessment is achieved through a sample monitoring process. Selected centres worldwide are required to provide a representative sample of examiners' marked tapes and scripts such that all examiners working at a centre over a given period are represented. Tapes and scripts are then second-marked by a team of IELTS Senior Examiners. Senior Examiners monitor for quality of both test conduct and rating, and feedback is returned to each centre. Analysis of the paired Examiner-Senior Examiner ratings from the sample monitoring data in 2003 produced an average correlation of .91 for the Writing module and .91 for the Speaking module.

The data presented here will shortly be posted on the IELTS website (www.ielts.org) along with other performance data for IELTS².

1. Academic Reading, Academic Writing, General Training Reading, General Training Writing, Listening and Speaking modules.
2. Previous performance data can be found in the relevant annual review also available on the IELTS website.

Table 1: Mean band scores for whole population

| Module | Listening | Reading | Writing | Speaking | Overall |
|------------------|-----------|---------|---------|----------|---------|
| Academic | 5.93 | 5.83 | 5.78 | 6.04 | 5.96 |
| General Training | 5.70 | 5.66 | 5.91 | 6.11 | 5.91 |

Table 2: Mean band scores by gender

| Gender/Module | Listening | Reading | Writing | Speaking | Overall |
|------------------|-----------|---------|---------|----------|---------|
| FEMALE | | | | | |
| Academic | 5.98 | 5.86 | 5.81 | 6.09 | 6.00 |
| General Training | 5.73 | 5.62 | 5.96 | 6.11 | 5.92 |
| MALE | | | | | |
| Academic | 5.89 | 5.80 | 5.76 | 5.99 | 5.92 |
| General Training | 5.68 | 5.68 | 5.87 | 6.12 | 5.90 |

Table 3: Mean band scores for most frequent first languages (Academic)

| | Listening | Reading | Writing | Speaking | Overall |
|------------|-----------|---------|---------|----------|---------|
| Chinese | 5.49 | 5.57 | 5.32 | 5.49 | 5.53 |
| Urdu | 6.25 | 5.77 | 6.12 | 6.24 | 6.16 |
| Bengali | 5.43 | 5.36 | 5.51 | 5.86 | 5.60 |
| Tagalog | 6.28 | 5.99 | 6.21 | 6.50 | 6.31 |
| Arabic | 5.96 | 5.61 | 5.70 | 6.39 | 5.98 |
| Hindi | 6.78 | 6.38 | 6.62 | 6.86 | 6.73 |
| Telugu | 6.25 | 5.91 | 6.33 | 6.38 | 6.28 |
| Korean | 5.70 | 5.59 | 5.25 | 5.63 | 5.61 |
| Malayalam | 6.31 | 6.13 | 6.49 | 6.52 | 6.43 |
| Thai | 5.53 | 5.53 | 5.16 | 5.58 | 5.51 |
| Punjabi | 6.20 | 5.76 | 6.15 | 6.21 | 6.15 |
| Japanese | 5.67 | 5.55 | 5.38 | 5.72 | 5.64 |
| Indonesian | 6.01 | 5.91 | 5.37 | 5.81 | 5.84 |
| Tamil | 6.61 | 6.22 | 6.48 | 6.72 | 6.57 |
| Spanish | 6.27 | 6.42 | 6.08 | 6.64 | 6.41 |
| Farsi | 5.68 | 5.63 | 5.92 | 6.28 | 5.94 |
| Gujurati | 6.20 | 5.78 | 6.12 | 6.21 | 6.14 |
| Vietnamese | 5.50 | 5.78 | 5.64 | 5.66 | 5.71 |
| Malay | 6.54 | 6.31 | 6.02 | 6.32 | 6.36 |
| Russian | 6.35 | 6.13 | 6.11 | 6.69 | 6.38 |

Table 4: Mean band scores for most frequent first languages (General Training)

| | Listening | Reading | Writing | Speaking | Overall |
|------------|-----------|---------|---------|----------|---------|
| Chinese | 5.40 | 5.69 | 5.50 | 5.49 | 5.58 |
| Tagalog | 6.00 | 5.77 | 6.13 | 6.29 | 6.11 |
| Hindi | 6.26 | 6.04 | 6.42 | 6.71 | 6.42 |
| Punjabi | 5.77 | 5.61 | 6.09 | 6.19 | 5.98 |
| Gujurati | 5.41 | 5.33 | 5.84 | 5.92 | 5.69 |
| Arabic | 5.27 | 5.04 | 5.33 | 5.98 | 5.47 |
| Malayalam | 5.52 | 5.41 | 6.02 | 5.99 | 5.80 |
| Urdu | 5.76 | 5.47 | 5.99 | 6.21 | 5.92 |
| Russian | 5.34 | 5.60 | 5.63 | 5.84 | 5.66 |
| Korean | 5.19 | 5.33 | 5.19 | 5.33 | 5.32 |
| Farsi | 5.37 | 5.36 | 5.94 | 6.23 | 5.79 |
| Spanish | 5.67 | 6.06 | 5.84 | 6.23 | 6.02 |
| Bengali | 5.52 | 5.43 | 5.85 | 6.26 | 5.83 |
| Tamil | 6.02 | 5.83 | 6.22 | 6.45 | 6.20 |
| Japanese | 5.35 | 5.36 | 5.27 | 5.54 | 5.44 |
| Telugu | 5.84 | 5.68 | 6.21 | 6.45 | 6.11 |
| Singhalese | 5.70 | 5.44 | 5.80 | 6.02 | 5.80 |
| Marathi | 6.29 | 6.03 | 6.61 | 6.76 | 6.48 |
| Indonesian | 6.08 | 6.05 | 5.66 | 5.90 | 5.99 |
| German | 6.62 | 6.57 | 6.55 | 6.94 | 6.73 |

Table 5: Mean band scores for most frequent nationalities (Academic)

| | Listening | Reading | Writing | Speaking | Overall |
|-----------------------|-----------|---------|---------|----------|---------|
| China | 5.33 | 5.41 | 5.20 | 5.36 | 5.39 |
| India | 6.51 | 6.15 | 6.48 | 6.62 | 6.51 |
| China (Hong Kong SAR) | 6.37 | 6.50 | 5.97 | 6.11 | 6.30 |
| Pakistan | 6.18 | 5.73 | 6.09 | 6.18 | 6.11 |
| Philippines | 6.29 | 6.00 | 6.22 | 6.51 | 6.31 |
| Bangladesh | 5.26 | 5.22 | 5.39 | 5.74 | 5.47 |
| Korea, South | 5.70 | 5.60 | 5.25 | 5.63 | 5.61 |
| Malaysia | 6.62 | 6.48 | 6.14 | 6.37 | 6.47 |
| Thailand | 5.53 | 5.53 | 5.16 | 5.58 | 5.52 |
| Taiwan | 5.37 | 5.44 | 5.25 | 5.63 | 5.49 |
| Japan | 5.67 | 5.55 | 5.38 | 5.72 | 5.64 |
| Indonesia | 6.02 | 5.91 | 5.38 | 5.82 | 5.85 |
| Iran | 5.68 | 5.64 | 5.92 | 6.28 | 5.94 |
| South Africa | 7.87 | 7.47 | 7.76 | 8.17 | 7.88 |
| Vietnam | 5.51 | 5.79 | 5.65 | 5.67 | 5.72 |
| Sri Lanka | 6.35 | 5.94 | 6.12 | 6.51 | 6.29 |
| Germany | 7.24 | 6.88 | 6.82 | 7.28 | 7.12 |
| Nigeria | 6.41 | 6.31 | 7.25 | 7.57 | 6.95 |
| Greece | 6.61 | 6.28 | 5.97 | 6.30 | 6.36 |
| Nepal | 5.86 | 5.58 | 5.66 | 5.92 | 5.82 |

Table 6: Mean band scores for most frequent nationalities (General Training)

| | Listening | Reading | Writing | Speaking | Overall |
|-----------------------|-----------|---------|---------|----------|---------|
| India | 5.92 | 5.74 | 6.24 | 6.40 | 6.14 |
| China | 5.35 | 5.67 | 5.48 | 5.42 | 5.54 |
| Philippines | 6.01 | 5.78 | 6.14 | 6.31 | 6.13 |
| Pakistan | 5.72 | 5.43 | 5.94 | 6.14 | 5.87 |
| Korea, South | 5.19 | 5.33 | 5.19 | 5.33 | 5.32 |
| Iran | 5.38 | 5.37 | 5.95 | 6.24 | 5.80 |
| Russia | 5.41 | 5.72 | 5.65 | 5.91 | 5.74 |
| Sri Lanka | 5.72 | 5.46 | 5.81 | 6.05 | 5.83 |
| Japan | 5.35 | 5.37 | 5.27 | 5.55 | 5.45 |
| Bangladesh | 5.14 | 5.08 | 5.59 | 6.02 | 5.52 |
| Egypt | 5.76 | 5.77 | 5.77 | 6.16 | 5.93 |
| United Arab Emirates | 4.44 | 3.77 | 4.55 | 5.38 | 4.61 |
| China (Hong Kong SAR) | 5.71 | 5.90 | 5.68 | 5.88 | 5.86 |
| Indonesia | 6.07 | 6.04 | 5.66 | 5.90 | 5.99 |
| Ukraine | 5.35 | 5.51 | 5.75 | 5.93 | 5.71 |
| Romania | 5.69 | 5.99 | 5.93 | 6.16 | 6.01 |
| Jordan | 5.62 | 5.46 | 5.57 | 6.13 | 5.75 |
| Iraq | 5.34 | 5.57 | 5.37 | 5.71 | 5.57 |
| Colombia | 5.06 | 5.45 | 5.30 | 5.79 | 5.47 |
| Malaysia | 6.46 | 6.23 | 6.42 | 6.76 | 6.53 |

Table 7: Reliability for Listening and Reading Modules

| Module | Version | Alpha | Module | Version | Alpha | Module | Version | Alpha |
|-----------|---------|-------|------------------|---------|-------|--------------------------|---------|-------|
| Listening | 151 | 0.89 | Academic Reading | 151 | 0.90 | General Training Reading | 151 | 0.87 |
| | 152 | 0.90 | | 152 | 0.87 | | 152 | 0.87 |
| | 153 | 0.90 | | 153 | 0.87 | | 153 | 0.87 |
| | 154 | 0.88 | | 154 | 0.87 | | 154 | 0.84 |
| | 155 | 0.88 | | 155 | 0.90 | | 155 | 0.90 |
| | 156 | 0.88 | | 156 | 0.85 | | 156 | 0.85 |
| | 157 | 0.89 | | 157 | 0.87 | | 157 | 0.86 |
| | 158 | 0.87 | | 158 | 0.89 | | 158 | 0.88 |
| | 159 | 0.90 | | 159 | 0.92 | | 159 | 0.87 |
| | 160 | 0.86 | | 160 | 0.89 | | 160 | 0.85 |
| | 161 | 0.86 | | 161 | 0.91 | | 161 | 0.88 |
| | 162 | 0.91 | | 162 | 0.89 | | 162 | 0.89 |
| | 163 | 0.86 | | 163 | 0.87 | | 163 | 0.88 |
| | 164 | 0.90 | | 164 | 0.88 | | 164 | 0.87 |
| | 165 | 0.91 | | 165 | 0.88 | | 165 | 0.83 |
| | 166 | 0.90 | | 166 | 0.86 | | 166 | 0.83 |

The IELTS joint-funded program celebrates a decade of research

All IELTS research activities are co-ordinated as part of a coherent framework for research and validation of the test. A major component of this framework is the funded research program sponsored jointly by the British Council and IDP: IELTS Australia with support from Cambridge ESOL.

The origins of this program date back to 1995 when the IELTS Australia Board first set aside grant funding and invited external researchers to submit IELTS-related proposals. They believed that such research would provide valuable information on a range of issues relating to the quality and standing of IELTS; it would also help IELTS stakeholders (including English language professionals and teachers) to develop a greater knowledge and understanding of the test.

The first round of funded studies was conducted in 1995 and a selection of these were edited and published in 1998 in *IELTS Research Reports*, Volume 1. Since then IDP has published four more edited volumes of selected reports from the period 1996–2001. The volumes can be ordered via the IELTS website (www.ielts.org).

In 1998 the British Council joined IELTS Australia in setting aside annual funds for research grants and since that time the program has been jointly funded by these two IELTS partners. Cambridge ESOL, the third IELTS partner, supports the program by supplying data, materials, advice and other types of assistance to approved researchers.

The annual call for research proposals is widely publicised and aims to reflect current concerns and issues relating to IELTS as a major international English language proficiency test. A Joint

Research Committee, comprising representatives of the three IELTS partners, agrees on 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 (including ability to keep to deadlines)
- 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 carried out, the Committee may call on a panel of external reviewers. The Committee also oversees the publication and/or presentation of research findings.

Over the past 10 years the results of the funded research program have made a significant contribution to the monitoring, evaluation and development process of IELTS, particularly in the following areas:

- The IELTS Writing test: issues of task design, construct validity, features of writing performance, examiner training and monitoring, approaches to assessment;
- The IELTS Speaking test: issues of task design, candidate discourse, assessment criteria, test bias, examiner/rater behaviour, examiner training/monitoring;

- The impact of IELTS: stakeholder attitudes, use of test scores, score gains, impact on courses and preparation materials, with key user groups;
- Computer-based IELTS: approaches to rating, issues of candidate processing.

Research into the Speaking Test directly informed the project to revise that module in 2001, and research into the Writing Test has contributed to the current Writing Revision Project (both these projects have been extensively reported in *Research Notes*). A selection of key reports relating to the IELTS Speaking and Writing Modules is currently in preparation as a volume in the *Studies in Language Testing* series published jointly by UCLES and Cambridge University Press. Issues of washback and impact have grown in importance in recent years and funded studies have

contributed to our understanding of the key role IELTS now plays in international education and society. More recently, work to develop a computer-based version of IELTS has benefited from funded studies in this area.

Since 1995, 55 research studies and nearly 70 separate researchers have received grants under the joint-funded program (see the list of all projects in rounds 1–7 in *Research Notes* 8, pages 23–34). A list of funded research projects can be found on the pages of the IELTS website. The next call for proposals (for round 11, 2005–6) will be published in *Research Notes* and elsewhere in April/May/June 2005.

Ten years on, the joint-funded research program now constitutes a key component in the IELTS partners' commitment to the validation and research agenda in support of IELTS and we look forward to the next ten years of collaborative research.

Conference Reports

In the first report below Helen Spillett and David Booth report on an event held at the University of Portsmouth in July entitled *Responding to the needs of the Chinese Learner in Higher Education* which is of particular relevance to IELTS as Chinese learners make up a considerable proportion of all IELTS test takers. The second report concerns an IATEFL joint Pronunciation and Learner Independence Special Interest Groups Event called *Independently Speaking* which was attended by Glyn Hughes and Dittany Rose. The third report by Vivian Cook (Professor of Applied Linguistics, University of Newcastle) summarises the topic of a staff seminar he led in Cambridge in May on the topic of multi-competence and language teaching. At the time of his visit Vivian was Reader at Essex University and his research interests include the English writing system, the design of course materials and the multi-competence view of L2 acquisition.

Responding to the needs of the Chinese Learner in Higher Education

This conference on 17–18 July 2004 was organised by the School of Languages and Area Studies at the University of Portsmouth, in association with SIETAR, the Society of Intercultural Education, Training and Research in the UK. There were 134 participants and around 30 sessions. The largest group of participants were members of ELT and EAP departments from UK universities. Other participants included representatives from university business departments and a mix of linguists, experts on Chinese education and testing, professionals with an interest in cross-cultural issues, agents (who send students from China to universities in the UK) and several delegates from further afield including Hong Kong and New Zealand.

Many UK universities have seen a very large increase in the numbers of mainland Chinese students registering for undergraduate and postgraduate degrees (currently around 30,000

in the UK as a whole). The main conference themes were the language skills of Chinese students, their study skills, their process of adjustment (academic, cultural, social and psychological) to higher education in the UK and the adequacy of the provision and support they receive.

There were four plenary sessions. The first was a joint presentation by Martin Cortazzi (Brunel University) and Jin Lixian (De Montfort University) on 'Changing Practices in Chinese Cultures of Learning'. They outlined the increase in English language learning in China, the expansion of universities, changes in pedagogical practices in Chinese schools and universities and the characteristics of students coming to study in Chinese universities. They focused particularly on the blend of traditional and changing modes of study which has influenced Chinese students before their arrival in the UK when they are confronted with western approaches to study there. Martin Cortazzi and Jin Lixian concluded that there were many areas for further research into meeting the needs of these students.

The second plenary was delivered by Jin Yan, Director of the Administration Office, and Yang Huizhong, Chairman, of China's National College English Testing Committee. This was an impressively detailed description of CET 4 and CET 6, the two levels of China's national college exit test for English language proficiency which is held twice a year in 31 provinces of China. CET 4 and 6, established since 1987, test the skills of Listening, Reading and Writing. CET is norm-referenced against the scores of a group of over 10,000 students in China's six top universities. There were 9.15 million CET candidates in 2003. Jin Yan and Yang Huizhong also described SET, the newer, optional Speaking test (offered since May 1999 in 34 cities to candidates who achieve more than 80% in CET 4 or 75% in CET 6). SET is a face-to-face test with an interlocutor, an assessor and three or four students. It had been taken by around 112,000 students up to the end of 2003. Yang Huizhong mentioned the positive backwash effect of SET, in

line with the government's push for a greater emphasis on listening and speaking. He indicated that there may be plans to develop CET 8, an advanced proficiency test focusing specifically on academic skills.

Helen Spencer-Oatey, eChina Learning Programme Manager, based at the Language Centre, University of Cambridge, gave the third plenary on 'Chinese students' adjustments to Britain: How are they coping and how can we help?' She started by considering the framework proposed by Colleen Ward (Ward et al 2001) for studying intercultural adaptation processes and focused on two aspects: psychological and sociocultural adjustment. She used data from an ongoing research study by Xiong Zhaoning of 126 mainland Chinese students taking pre-degree English language training (Xiong, in progress). Interestingly, this research indicated that, although the mean score for psychological welfare for these students was within the normal range, a significant group of students were suffering from moderate to severe depression. Most respondents found it difficult to develop and maintain social contact with non-Chinese. Helen Spencer-Oatey referred to Bochner's work (Bochner et al 1997) on three social networks of overseas students (compatriots, significant host nationals and other friends and acquaintances) and stressed that findings about these need to be taken into account when planning university policy.

The fourth plenary was delivered by Ros Richards of the Centre for Applied Language Studies at the University of Reading who spoke on 'Meeting the needs of Chinese students during the academic acculturation process'. Ros Richards pointed out that students typically arrive in higher education expecting to depend on the study strategies which have succeeded in their previous learning environments and that, for growing numbers of Chinese students in the UK, the reality is very different. She referred to the Dearing Report's vision of a learning culture and the QAA descriptors for qualifications at BA, MA and PhD levels and asked how equipped overseas students are to succeed within this framework. She highlighted the requirements for success, e.g. the abilities to read critically and extensively, to paraphrase and summarise, to synthesise material, to organise and support information logically and coherently, to critically evaluate sources, to write extended essays or reports and she suggested that IELTS and TOEFL scores may not reflect a direct ability to meet these requirements. She then contrasted the features of the Chinese learning experience (e.g. courses based on one or two prescribed texts, teacher oriented learning, success attributed to hard work rather than ability, pressure to conform) and described the strengths and weaknesses of Chinese students: determination and a strong respect for education but little experience of reading or writing at length, critical analysis or independent learning. She argued that there should be more explicit strategy training, scaffolded activities and opportunities to develop and practise skills within pre-sessional and in-sessional English language and study skills programmes. Her conclusion stressed the responsibility of universities to become aware of other notions of learning and to provide support so that international students can benefit fully from their courses in the UK.

Many issues raised in the plenaries were discussed further in the other sessions. A recurrent theme was the notion of 'the Chinese

learner'. Martin Cortazzi and Jin Lixian had cautioned against stereotyping, emphasised the diversity and complexity of Chinese cultures of learning and argued that the Chinese focus on memorisation reflected a cognitively deep process of understanding rather than superficial rote learning. A number of presentations (e.g. Elizabeth Hauge and Simon Le Tocq, Aiqun Li, Joan O'Grady and Lijing Shi) were based on small-scale research projects into the strategies, beliefs and preferences of Chinese learners. Lijing Shi (Open University) gave a presentation on 'The successors to Confucianism or a new generation?' She started with two views of Chinese students: the first as 'heavily influenced by Confucianism' (passive, lacking in critical thinking, obedient) and the second as 'similar to western model students' (active and reflective, open-minded, with a spirit of inquiry). Lijing Shi presented the findings of a questionnaire given to 400 12–17 year-olds in Shanghai. These students presented a very mixed picture: they preferred equality to hierarchy in relationships with teachers and were willing to criticise coursebooks and teachers e.g. but they also emphasised the importance of hard work and perseverance. Lijing Shi argued that there is no dichotomy between two cultures of learning but a continuum from the more traditional to the most westernised, affected by many other factors (regional, financial, personality etc.) in the case of any particular group of students.

Other talks focused on specific skill areas (e.g. Lynn Errey on reading, Linda Lin, Barbara Mayor and Dilys Thorp on writing). Linda Lin (Hong Kong Polytechnic University) talked about 'Language learning hurdles Hong Kong students are to cross in higher education'. She looked at features of L1 transference leading to common errors in writing. Most of these features are common to Mandarin and Cantonese so the errors highlighted are also typical of mainland Chinese students. Using data from the Hong Kong Polytechnic University corpus of learner English, she analysed typical errors based on topicalisation (redundant verbs, misuse of adjectives) semantic prosody, cultural background (inappropriate register in academic writing, rhetorical questions) and secondary school learning experience (overuse and misuse of connectives). Barbara Mayor (Open University) and Dilys Thorp (Exeter University) looked at the writing of Chinese IELTS candidates and noted a high proportion of interrogatives and imperatives and a tendency to express themselves with a high level of certainty. They suggested that the IELTS Academic Writing Part 2 test prompt¹, coupled with the lack of source material, may lead to a more dialogic and assertive style in all students. However, they argued that in order to assess fairly the potential of Chinese learners, it is important to recognise that some students who have performed well in the Chinese educational system but received less explicit IELTS coaching may import into their writing a range of hitherto valued features of Chinese writing. Lynn Errey of Oxford Brookes University described a very interesting study of reading as a cultural practice. She compared the ways in which she and a visiting Chinese EGP (English for General Purposes) lecturer from Harbin approached the teaching of reading. She also looked at the

1. *Present a written argument or case to an educated reader with no specialist knowledge of the following topic. ... You should use your own ideas, knowledge and experience and support your arguments with examples and relevant evidence (see online sample materials at <http://www.ielts.org/candidates>).*

ways in which a pre-sessional group of Chinese students' perceptions of reading strategies changed after moving to the UK university context. She stressed the need to look at what Chinese students were doing in China and to provide explicit meta-cognitive scaffolding for encouraging them to develop appropriate EAP skills.

The conference organiser, Tricia Coverdale Jones, summarised the main insights of the conference. There is clearly a need for university policy to adapt to the challenges posed by the dramatic increase in numbers of Chinese students. However, a simplistic idea about the nature of 'the Chinese learner' needs to be readdressed. Chinese students, like students from any other cultural background, are characterised by tendencies rather than absolutes.

The focus of the IELTS test, and the interpretation of IELTS scores by universities and EAP departments, was a theme touched on by a number of conference participants. There was widespread recognition that an IELTS score is only an indication of language proficiency and not of academic study skills. However, it is clear that universities, and Chinese students themselves perhaps, are inclined to assume that achieving a particular score in IELTS is sufficient to ensure the success of a Chinese learner in higher education in the UK. EAP professionals at the conference noted this and highlighted the need for EAP departments to raise their profile within their own universities so that international students can access in-sessional courses to support their learning throughout their university careers².

References

- Bochner, S, McLeod, B and Lin, A (1997) Friendship patterns of overseas students. A functional model, *International Journal of Psychology*, 12: 277–297.
- Ward, C, Bochner, S and Furnham, A (2001) *The Psychology of Culture Shock* (2nd ed.), Cambridge: Cambridge University Press.
- Xiong, Z (in progress) *Exploring the sojourn experience of overseas Chinese students on an international foundation course at a British University*, PhD research, University of Luton.

IATEFL Special Interest Groups Event – Independently Speaking

An IATEFL (*International Association of Teachers of English as a Foreign Language*) event was held at the University of Reading on Saturday 26th June 2004, focussing on the topics of pronunciation and learner independence.

John Field (University of Leeds) opened the event with a session outlining *Exemplar Theory* and its implications for the teaching of pronunciation and also the teaching and testing of listening. According to exemplar theory, the brain stores traces of every voice it hears, which the listener can later use as points of reference, enabling the listener to establish tolerances for voices with certain accents, pitch etc. This makes it easier for the learner to understand voices that contain those features. For Field, the

most important implication of this theory is that all learners have individual experiences in terms of exposure so will have individual strengths and weaknesses that are not best dealt with by group activities in class. Instead Field suggests training students in listening strategies so that they can deal with the varied input they are exposed to in the real world. Field also emphasised that learners may have different stores for language they understand from language they produce, thus explaining why students can sometimes hear the differences between sounds without being able to reproduce them. In terms of implications for language testers, Field felt that a 'fetish for regional accents' in exams can disadvantage candidates who have not built up traces of that accent and suggested that a range of different 'standard' accents from around the world would be more appropriate, as Cambridge ESOL seeks to do in its listening papers.

Piers Messum (PhD student, University College, London) questioned the benefit of learner autonomy at early stages of pronunciation teaching. He outlined aspects of his research into L1 acquisition in children which has led him to believe that pronunciation is not learnt through pure imitation. He argued that it is only possible to imitate sounds once you are aware of the mechanics of how to produce them. He also stated his belief that children develop many features of English speech, such as its rhythm, not due to imitation but because of physiological factors such as the inability to regulate breathing sufficiently as a young child. Sally McPherson (Trinity College, London) outlined ways in which pronunciation work could be integrated into the classroom and also into teacher-training courses.

The afternoon began with a workshop by Richard Cauldwell to demonstrate his Streaming Speech CD-ROM. This uses recordings of spontaneous speech from native speakers and isolates and replays parts of the text to help advanced learners of English with listening and pronunciation skills. The texts are also of interest to producers of listening material, including test designers, in that, when analysed, they illuminate certain features of spontaneous speech. Cauldwell gave the example of weak forms which in the past have often been presented to students as existing for only a limited range of words such as 'was' and 'were'. He noted that in fact all words have a range of potential weak forms, depending on their use within an utterance. Cauldwell is currently working on a version of Streaming Speech using North American voices and is interested in finding a way to do a version for less advanced students, possible around the level of FCE. The techniques used by Cauldwell have already been used to compare Cambridge ESOL listening material with authentic speech. In addition, his assertion that authentic speech often fails to fit into the preconceived patterns found in many EFL publications is useful for us to bear in mind, both when producing written material and when looking at how spoken performance is assessed.

There followed a number of short presentations. Melanie Bell of Anglia Polytechnic University spoke about 'Equipping learners to work on their Pronunciation'. She described a course that she teaches in which students are provided with specific in-depth knowledge about pronunciation with which to go on and analyse their own needs. Frances Hotimsky presented a method she uses in Italy to help students with poor stress placement. She uses place

2. Some papers from this conference may be included in a special China issue of *Language, Culture and Curriculum*, to appear in 2005.

names from the students' L1 to categorise other words in the target language (English). It was pointed out that this may not work in other languages where word stress was not variable. Dermot Campbell and a group from the Dublin Institute of Technology introduced their 'DIT CALL' project, which enables spontaneous speech to be slowed down without distortion. In addition to creating a CD-ROM of worksheets aimed at helping students to hear the syllables that are lost during fast speech, they explained that the technology would also be useful for researchers in analysing sounds produced in English. This software could be useful to Cambridge ESOL in many ways, for example as a means of pinpointing what aspects of spoken English are most difficult for learners (and test-takers) to process.

John Field's session was particularly relevant to Cambridge ESOL as research into how learners process the language they hear has, as mentioned above, a potential impact on the way that we test listening. The new technologies and techniques that were discussed during this event, could be of benefit in helping us to analyse both the speech used in our listening tests and also candidate performance in speaking tests.

Vivian Cook's Portrait of an L2 User: Multi-competence and Language Teaching

The starting point is the concept of multi-competence – the knowledge of two or more languages in one mind. This term was devised to encompass both the language systems present in the same mind – the L1 and the interlanguage. Hence multi-competence means all the language systems that an L2 user possesses, say L1 French plus L2 English at some stage of development.

This conceptualisation opened the door to looking at multi-competence as a distinctive state from monolingualism. It can be shown that:

- L2 users think differently from monolinguals, for example categorising colours differently
- they use language in different ways, for example in code-switching and translation
- they have an increased awareness of language itself and learn to read more rapidly in their first language
- their knowledge of their L1 is affected by the L2 they have acquired
- they are more efficient communicators in the L1.

Learning a second language is not just adding another language it is changing different aspects of the user's mind in subtle ways. L2 users have different kinds of mind from monolinguals.

One overall consequence for language teaching is the status of the native speaker. Language teaching success has almost always been seen explicitly or implicitly as closeness to a native norm; language syllabuses and tests are based on what it is believed a normalised ideal native speaker would do; students (and teachers) feel that differences from natives are failures. The thrust of multi-

competence is that native speakers lack many of the attributes of L2 users by definition. We need then to train students to be successful L2 users. Course-books should feature role models of powerful L2 users that students can strive to emulate, not the typical powerless L2 users they usually feature in the form of tourists and students. They should also rely more on situations where non-native speakers are involved as pure native-to-native conversation is the one form the students will never take part in. This implies a greater tolerance of forms of language produced by L2 users, not just native speakers – the only problem is that with rare exceptions we still have no proper description of what these might be.

A second major implication is the role of the first language in second language acquisition. Multi-competence sees the L1 as always and unavoidably present in the L2 user's mind. The fact that the class appears to be all in the target language disguises the fact that another language is invisibly active in the minds of all the students. The instinctive EFL reaction has always been to ban the L1 as much as possible. But, if this is just driving it underground, we should think of ways in which this force could be utilised in language teaching. Current suggestions include using the L1 to convey meaning, to form emotional rapport with the students and to demonstrate the use of code-switching, a real L2 use of language.

A third implication from multi-competence is the reawakening of goals for language teaching other than communication. For the last part of the twentieth century teaching concentrated on the external goal of getting the students to use the language in useful situations. Hence it tended to de-emphasise the traditional internal goal of general educational values, self-development, cognitive training, the development of language understanding itself, etc. all of which are supported by the multi-competence view of the L2 user as a distinctive type of person.

This seminar discussed what it means to be an 'L2 user' and raised important questions including: Is being an L2 user the same as being a 'deficient native speaker'? Does the L2 user have a status in his/her own right? Vivian helped Cambridge ESOL to consider these issues and the implications they have for assessment practice, where the 'native speaker' has traditionally provided the standard for comparison.

References and further reading

- Cook, V (Ed.) (2002) *Portraits of the L2 User*, Clevedon: Multilingual Matters.
- (Ed.) (2003a) *Effects of the Second Language on the First*, Clevedon: Multilingual Matters.
- (2003b) Materials for adult beginners from an L2 user perspective, in B Tomlinson (Ed.), *Developing Materials for Language Teaching*, London: Continuum, 275-290.
- (2004) The goals of ELT: Reproducing native-speakers or promoting multi-competence among second language users? in Davison, C and Cummins, J (Eds) *Handbook of English Language Teaching* (volume 2), Dordrecht: Kluwer Academic Publishers.