6

ResearchNotes

UNIVERSITY OF CAMBRIDGE LOCAL EXAMINATIONS SYNDICATE
ENGLISH AS A FOREIGN LANGUAGE (EFL)

NOVEMBER 2001

UNIVERSITY *of* CAMBRIDGE
Local Examinations Syndicate

Research**Notes**

Introduction

Welcome to Issue 6 of *Research Notes,* the UCLES EFL Newsletter which reports on our work on research, test development and validation issues.

The very first issue of *Research Notes* (March 2000) drew attention to the existence of the UCLES/CUP Learner Corpus (or Cambridge Learner Corpus). In this issue Fiona Ball provides an update on the CLC and describes the ways in which UCLES EFL is increasingly making use of both learner and native speaker corpora to develop and validate its EFL examinations.

The revised IELTS Speaking Test has now been fully operational since July 2001. Lynda Taylor concludes her recent series on the IELTS Speaking Test Revision Project with an article describing the examiner retraining programme which was carried out worldwide in readiness for the introduction of the revised test format. A further article in this issue focuses on some recent studies – both within UCLES and elsewhere – into the use of the paired format for assessing speaking ability.

In *Research Notes 2* Nick Saville traced the development of the IELTS Impact Study, a major long-term programme of research by UCLES EFL into the impact of a widely-used, high-stakes language test for those needing to study or train in English. In this issue Roger Hawkey reports on developments in this project over the past year, particularly the refinement of the data collection instruments in preparation for their implementation towards the end of 2001.

In the first of what will become a series of articles on assessing second language writing ability, Stuart Shaw outlines a range of key issues in this field, including the nature of the proficiency continuum, marker strategies, rater training/standardisation, and the impact of new technology; he highlights some of the ways in which UCLES EFL is addressing these key issues.

Finally, as the 2001 European Year of Languages draws to a close, Marianne Hirtzel reports on the success of the ALTE Conference held in July in Barcelona (Spain) to celebrate this special year. This issue of *Research Notes* also includes a brief review of various contributions made by visiting speakers to our UCLES EFL Seminar Programme during 2001.

Contents

Issues in the assessment of second language writing

Stuart D Shaw, Validation Officer, UCLES EFL

Introduction

Of particular interest to second language writing assessment research is the direct test of language performance in writing and its contribution to profiles of learner communicative competence (Hawkey, 1982; Weir, 1993; North, 2000). According to Jacobs et al. (1981:3),

> The direct testing of writing emphasizes the **communicative** purpose of writing … (it) utilizes the important intuitive, albeit subjective, resources of other participants in the communicative process – the **readers** of written discourse, who must be the ultimate judges of the success or failure of the writer's communicative efforts.

By "direct test" we mean only tests that test writing through the production of writing, in contrast to "indirect" or "objective" tests of writing that purport to test writing through error recognition, verbal reasoning and other measures that have been demonstrated to correlate reasonably highly with measured writing ability.

Hamp-Lyons (1991) suggests that a direct test of writing has at least the following five characteristics: First, an individual taking the test must physically produce at least one piece of continuous written text of 100 words or more – a length widely considered to be a minimum sample. Second, although the test taker is presented with a set of instructions and a text, picture or other "stimulus" material, he or she is given substantial freedom within which to create a response to the stimulus. Third, each written response is assessed by at least one, usually two, human raters (with a third in case of extreme disagreement), who have undertaken some form of standardisation or preparatory training for the essay evaluation process. Fourth, the judgements made by the rater are in some way associated with a common 'yardstick' which may be a description of expected performance at certain proficiency levels, a set of exemplar scripts or one of several rating scales. Fifth, the response of the rater can be expressed numerically, instead of or supplemental to, written and/or verbal comments.

Almost all the Cambridge EFL examinations include a direct writing test designed to assess a candidate's skill in writing above the sentence level. The quality of the output produced by a candidate in response to a writing stimulus is assessed by one or more examiners according to specified criteria.

The assessment of second language writing ability raises a number of important issues. The intention of this paper is to introduce some of the key issues in a broad sense such that they can be explored in follow-up articles in relation to current UCLES research. The issues fall into five categories:

1. Features which distinguish second language writing performance at different levels of proficiency i.e. a Common Scale for Writing;

2. Process by which writing examiners form judgements about scripts i.e. marker strategies;

3. Achievement of an acceptable level of reliability in written assessment i.e. the nature of examiner training and standardisation;

4. Effect on examiners when assessing handwritten or computer-generated scripts i.e. the relationship between computer-based (CB) and paper-and-pencil (P&P) tests of writing;

5. Impact of new technology on approaches to assessing writing i.e. Electronic Rating (E-Rating) and Electronic Script Management (ESM).

Second language writing issues

1 Common Scale for Writing

The Common Scale for Writing project comprises two phases. Phase 1 (Capel and Hamp-Lyons, 1995) concentrated solely on :

- attempting to discover the relationship between the levels of the Cambridge suite of exams on a hypothesised *common scale* of writing ability;

- seeking direction for the pursuit of a *harmonised approach* to the assessment of writing across the levels.

Existing writing assessment scales were used to derive a draft set of "pass-level" descriptors of the writing proficiencies of candidates from KET through to CPE and typical main suite candidate scripts from all five levels were studied in order to propose "can do", "can sometimes do", and "cannot do" statements.

The second phase of the project (Hawkey, Banks and Saville, 2001), a corpus-based study, is currently in progress (see *Research Notes 5*) and attempts to answer the following research questions :

- Using a text-based analysis, what distinguishing features in the writing performance of EFL/ESL learners or users taking the Cambridge English examinations can be identified across three proficiency levels (FCE, CAE, CPE) addressing a common task?

- How can these features be incorporated into a single scale of bands, that is, a common scale, describing different levels of L2 writing proficiency?

It is expected that the methods and findings of these studies will be co-ordinated with related corpus-analytical research currently being undertaken by Kennedy, Dudley-Evans and Thorp (University of Birmingham) which focuses on IELTS Academic and General Writing.

2 Marker Strategies

A variety of studies have found that "different markers respond to different facets of writing" (Diederich, French and Carlson, 1961). Vaughan (1991) notes that "despite their similar training, different markers focus on different essay elements and perhaps have individual approaches to reading essays".

Research on what determines marker judgements indicates that content and organisation are the writing features which have the most significant influences on those who mark, rate or evaluate writing texts. Freedman shows that content is the most significant single factor in the final judgement of an essay and finds a hierarchy of assessment criteria such that raters "value content first and then organisation" and that features like mechanics are less important (1979:161).

A second factor which influences the way in which raters assess writing is the characteristics of raters themselves. Results from previous studies show that marker behaviour varies in ways that can be partially attributed to variables such as gender, professional background, amount of training in the use of assessment tools, experiential background and amount of exposure to L2 writing (Hamp-Lyons, 1990; Vann, Lorenz and Meyer, 1991). Weigle (1994) demonstrates that rater training affects the marks given by inexperienced raters, such that inter-rater consistency is improved.

Over recent years L2 writing researchers have consistently recommended that investigating the processes raters go through in arriving at judgements is one way to reach a greater understanding of rater behaviour (Huot, 1990, 1990a; Hamp-Lyons, 1990; Tedick and Mathison, 1995; Brown, 1995 and Milanovic, Saville and Shuhong, 1996). Hamp-Lyons (1990:81) suggests that studying how the "experiential backgrounds" of raters may influence their responses has suffered as a consequence of a major preoccupation with testing and validation of scoring procedures; she suggests that the experiential background of a rater is not just a matter of

demographics and professional background, but also a whole range of aspects relating to the processes of rating including the decisions raters make, how raters use rating tools, the experience of reading, and the nature of external pressures and how they affect rater judgements.

Milanovic, Saville and Shuhong (1993/6) reported on the thought processes of examiners for Cambridge EFL compositions in terms of the approaches they employ while marking and the elements markers focus on (*Studies in Language Testing, Volume 3*, 1996). In further studies of the decision-making behaviour of composition markers, Milanovic and Saville (1994) used verbal protocols to investigate marker strategies and revealed four discernible approaches to composition marking. Such studies are designed to address certain fundamental questions:

- What is the raters' decision-making behaviour in terms of the approaches they employ while marking EFL compositions?
- What elements do the markers focus on while marking compositions?
- How does examiner marking behaviour compare in relation to inter-rater consistency?
- Do examiners adjust their marking behaviour according to the level of the script?

3 Nature of Training and Standardisation

The training of examiners is widely accepted, among specialists in both writing assessment and educational measurement, as being essential to reliability and validity in the testing of second language performance (Alderson, Clapham and Wall, 1995). Training may also have an important role to play in the professionalisation of language testing which has been called for by Bachman (2000).

However, in both the writing assessment and measurement literature some debate exists as to the efficacy and purpose of rater training. Rater variation is a potentially serious weakness in tests of language performance and rater-training has been widely recommended as a means of keeping this variation within acceptable limits.

Two broad fields of research concern are related to rater-training. The first, discussed above, is the need for a greater understanding of the processes by which a rater arrives at a rating: "lack of knowledge in this area makes it difficult to train markers to make valid and reliable assessments" (Milanovic, Saville and Shuhong, 1996:93). The second concern is with the

effects of training and, more significantly, asks whether training is capable in practice of engendering the required change in rater behaviour.

Studies in recent years have highlighted fundamental questions about the function of training and the effectiveness of existing programmes. Weigle (1998) notes that "little is known about what actually occurs during rater training and how it affects the raters themselves". She observes that training could help raters to be self-consistent, but that it is less successful at getting different raters to give the same scores.

On-going research at UCLES EFL is currently investigating the effect of standardisation training on rater judgement and inter-rater reliability for the revised CPE Writing Paper 2. This study focuses on the standardisation process as the variable most critical to improving the assessment of writing, and aims to find ways of improving inter-rater agreement. The study, which is goal-oriented, tests the hypothesis that a steady improvement in inter-rater correlation will take place with each successive iteration of the standardisation exercise. In particular the study addresses :

- What change is there during successive standardisation training in the scores given by raters?

- How can raters be trained to use a mark scale in order to achieve a satisfactory level of reliability?

- Do iterative standardisation exercises improve the inter-rater reliability of multiple raters rating the same set of scripts?

Results of this investigation will be reported in *Research Notes 7*.

4 Relationship between CB and P&P tests of writing

The introduction of computer administered tests, as a new medium for assessment, has raised fundamental considerations. Whilst computer-based tests are subject to the same requirements of reliability and validity expected of any other test, certain critical issues of equivalence of scores yielded from computer and paper-and-pencil administered test versions have emerged.

Clearly, in translating a test from one medium to another, new medium it is crucial to ascertain to what extent the new medium may alter the nature of the underlying test construct, or change the scale. Specific research questions raised by the new medium with regard to the assessment of second language writing are:

- What is the impact of composition medium on essay raters in second language writing assessment?

- Handwriting and general presentation is widely regarded as a contaminating factor in the assessment of writing ability (Brown 2000). Given that handwriting, neatness and layout contribute to the basis of legibility, how significant is the role of legibility in the assessment of writing on-screen and what is its impact?

- Do raters rate handwritten responses differently to computerised responses and do any differences interact with gender, ethnicity or socio-economic background?

- Does a candidate's response take longer to assess on-screen rather than on-paper?

- Does computerised assessment of writing contribute to test bias?

- Are raters affected by their own familiarity with computers and attitudes to the computerised delivery of tests?

Whilst computer-based testing is a comparatively new development in UCLES EFL it is in a number of ways still developing. A current consideration is the relationship between tests in parallel computer-based and paper-and-pencil formats – test equivalence.

5 E-Rating and ESM

UCLES is investigating the feasibility of automatically grading essays using a computer. In this context an essay is simply a textual response to a question typed into a computer including any number of words from a few sentences upward.

The objectives of these studies are to investigate a number of possible approaches to the automatic assessment of natural language essays. Human assigned grades represent an overall judgement of quality in language form, structure and content. The research question raised by such studies is whether automatic techniques can simulate human judgement.

ESM defines the process by which scripts are scanned and the images transmitted electronically to an image server at UCLES. These images are then distributed electronically and marked on screen by examiners. Marks are captured electronically, without manual intervention.

A number of ESM projects have been undertaken by UCLES. The projects have two major sets of objectives: to investigate practical possibilities and the impact on process quality and time, and to provide data for research to enable an evaluation of the impact on assessment reliability of handling documents on screen rather than on paper.

A range of issues arise from the potential to assess writing on screen rather than by the conventional approach:

- Should different approaches to current assessment methods (standardisation and double marking) be adopted in order to detect and correct deviant marking earlier in the process than is possible at present?

- Should techniques be employed to detect deviant examiners?

- Once deviation is detected, what should be the nature of any corrective action, how should it be applied and how should its efficacy be superintended?

Conclusion

This introductory overview of second language writing assessment has highlighted a wide range of key issues; as we continue to report on the UCLES EFL research programme through articles in *Research Notes* we plan to address many of these issues in more detail. Hopefully these contributions will provide a useful and coherent view of current thinking and practice for all those who are working in the field of second language writing research.

References and further reading

Alderson, J C, Clapham, C and Wall, D (1995): *Language Test Construction and Evaluation*, Cambridge University Press

Bachman, L F (2000): Modern language testing at the turn of the century: assuring that what we count counts, *Language Testing* 17/1, 1-42

Brown, A (1995): The effect of rater variables in the development of an occupation-specific language performance test. *Language Testing*, 12, 11-15

Brown, A (2000): Legibility and the rating of second language writing : an investigation of the rating of handwritten and word-processed IELTS Task Two essays. IELTS Research Projects 1999/2000. University of Melbourne

Diederich, P B, French, J W and Carlton, S T (1961): Factors in judgements of writing ability, *ETS Research Bulletin RB-61-15*, Princeton N J: Educational Testing Service

Freedman, S W (1979): Why do teachers give the grades they do? *College Composition*, 32, 365-387

Hamp-Lyons, L (1990): Second language writing: assessment issues. In B Kroll (Ed) *Second Language Writing: Research Insights for the Classroom*, Cambridge University Press

Hamp-Lyons, L (1991): Reconstructing 'academic writing proficiency'. In L Hamp-Lyons (Ed) *Assessing Second Language Writing in Academic Contexts*, Norwood, N J: Ablex Publishing Corporation

Hamp-Lyons, L and Capel, A (1994): A Common Scale Writing Project. UCLES EFL Internal Report

Hawkey, R (1995): An investigation of inter-relationships between cognitive/affective and social factors and language learning. Unpublished PhD thesis, Department of English for Speakers of Other Languages, Institute of Education, London University

Hawkey, R, Saville, N and Banks, C (2001): Common Scale for writing: toward a common performance description scale for non-native speaker writing. Presented as a research in progress paper at LTRC 2001, St Louis, USA

Huot, B (1990): Reliability, validity and holistic scoring: what we know and what we need to know, *College Composition and Communication*, 41/2, 201-213

Huot, B (1990a): The literature of direct writing assessment: major concerns and prevailing trends, *Review of Educational Research*, 60/2, 237-263

Jacobs, H, Zinkgraf, S, Wormuth, D, Hartfiel V F and Hughey, J (1981): *Testing ESL composition: a practical approach*, Rowley, Mass:Newbury House

Milanovic, M and Saville, N (1994): An investigation of marking strategies using verbal protocols, UCLES EFL Internal Report

Milanovic, M, Saville, N and Shuhong, S (1996): A study of the decision-making behaviour of composition markers. In M. Milanovic and N. Saville (Eds), Studies in Language Testing – *Performance Testing, Cognition and Assessment: Selected Papers from the 15th Language Testing Research Colloquium*. University of Cambridge Local Examinations Syndicate and Cambridge University Press

North, B (2000): Linking language assessments: an example in a low-stakes context. *System*, 28, 55-577

Tedick, D J and Mathison, M A (1995): Holistic scoring in ESL writing assessment: what does an analysis of rhetorical features reveal? In D Belcher and G Braine (Eds), *Academic Writing in a Second Language: Essays on Research and Pedagogy*, Norwood, NJ: Ablex Publishing Corporation

Vann, R J , Lorenz, F O and Meyer, D M (1991): Error gravity: faculty response to errors in the written discourse of nonnative speakers of English. In L Hamp-Lyons (Ed), *Assessing Second Language Writing in Academic Contexts*

Vaughan, C (1991): Holistic assessment: what goes on in the rater's mind? In L. Hamp-Lyons (Ed), *Assessing Second Language Writing in Academic Contexts*

Weigle, S C (1994): Effects of training on raters of ESL compositions, *Language Testing*, 11/2, 197-223

Weigle, S C (1998): Using FACETS to model rater training effects. *Language Testing*, 15/2, 263-287

Weir, C (1993): *Understanding and Developing Language Tests,* Hemel Hempstead: Prentice Hall

Using corpora in language testing

Fiona Ball, Validation Officer, UCLES EFL

Corpora are increasingly popular in pure and applied linguistic research communities, including the Language Testing community. The reasons for their growing popularity include their accessibility, cost and the range of research tools available for searching corpora. Corpora provide a ready-made collection of texts that can help to reveal many aspects of language use quickly and accurately, reducing the need for data collection and manual analysis. Corpora and Corpus Linguistics are increasingly used for UCLES' ongoing research projects alongside established methodologies. UCLES continues to use the intuition, knowledge and experience of its item writers, examiners and subject officers in developing suitable tasks and topics for EFL examinations.

A previous article in *Research Notes 1* (Boyle and Booth, 2000) described the Cambridge Learner Corpus (CLC) and outlined the current applications of this large collection of learners' written examination scripts on computer. This article returns to the issues of using corpora in language testing and describes the current corpus-based activities being undertaken at UCLES. Several new research activities have been started in 2001 and the profile of corpus linguistics and its applications for language testing have been raised considerably both within UCLES and at conferences. The corpus-related research profile of UCLES has therefore improved considerably since early 2000 and looks set to expand as more uses of corpora come to light.

Currently UCLES does two types of corpus-related activities: Corpus-building and Corpus-informed research.

Corpus-building at UCLES: Cambridge Learner Corpus

This aspect of UCLES' use of corpora has expanded significantly recently, with the ongoing development of the CLC and several new corpora being built. The Cambridge Learner Corpus now contains over 13.5 million words of Main Suite and Business English Certificate (BEC) examination scripts and has an annual growth rate of around 3 million words. The spread of texts within the corpus is due to expand with the inclusion of Academic English (IELTS) and modular general English examination (CELS) written scripts in the near future. Currently, the spread of words from General and Business English texts is uneven, likewise the spread between the five Cambridge levels. This is due to the difference in length of

examination tasks at various levels and the availability of examination scripts, amongst other factors. The ongoing development of the CLC is a high priority for UCLES and including a wider range of scripts in the corpus will increase the representativeness of the corpus. Ideally, this corpus would include an equal amount of words for each level and type of English, together with a proportional amount of words from each of many different first languages represented in the corpus. The searching options on the corpus have also been improved recently, to allow for searching by examination, Cambridge level, language, country as well as other demographic variables that were not available before. In collaboration with CUP, UCLES hopes to improve the user-friendliness of this corpus so that more UCLES and CUP staff can use it. The growth and significant developments in the applications of the CLC will be reported in future issues of *Research Notes*.

Cambridge Corpus of Spoken Learner English

As well as the CLC, UCLES has recently begun work on the Cambridge Corpus of Spoken Learner English, which is envisaged to include speaking tests for all Cambridge examinations, a formidable task. As well as being an archive of a sample of speaking performances from all over the world, this corpus will provide another means of researching various aspects of speech in relation to candidate information, and to grade and score data. The initial stage of this project involves building the first 'sub-corpus' to be included in the CCSLE, which will be of Young Learner speaking tests. The Cambridge Young Learners (YL) suite of examinations was chosen for this research because of the relatively short and straightforward nature of the speaking tests. These last between five and ten minutes and involve short answers on the part of the candidate together with an Interlocutor Frame that the examiner should follow throughout the test.

Speaking tests are being recorded in key centres world-wide and will be transcribed when they are returned to UCLES. Problematic areas of the transcription include deciding whether and how to represent gesture and other non-verbal communication, which are key parts of the YL tests, especially at the first level, Starters. The first goal of this corpus is to include approximately 100 speaking tests for the three levels of Starters, Movers and Flyers. Text and sound files of the speaking tests will be linked within an experimental corpus to provide two ways of working with the speaking test data. A future development might be to link the specific

materials used by the candidates to the sound and text files, in order to give the speaking test more context. Future articles will provide updates on the development of this small corpus, which is one of several ongoing research projects involving the Young Learners examinations.

Business English Texts Corpus

Another experimental corpus-building project currently underway is the web-based collection of business texts to inform the development of the BEC Preliminary wordlist. Wordlists are often used by item-writers to assist them in producing realistic examination tasks at specific levels. Previous approaches to validating new items for inclusion in the Lower Main Suite (KET, PET) and BEC wordlists included drawing on frequency data from the British National Corpus and from the CLC. This procedure is being reconsidered in the light of a more contextualised approach that will indicate the collocational patterns of certain words or phrases and suggest the different senses of words in real texts. The Internet will be the main source of general business related texts for this mini-corpus, with internal documentation providing an additional source of current English usage in a business context. This type of documentation is very similar to the tasks that Business English Certificate candidates are asked to write, for example, general letters, faxes and memos, whereas much Internet material would be too specific or complex to be considered relevant to the first two levels of BEC. This project will provide another way of investigating suitable vocabulary for inclusion on wordlists and its methodology is hoped to influence the development of other wordlists and other aspects of the question paper production process.

Corpus-informed research at UCLES

In addition to building general and project-specific corpora as outlined above, UCLES is involved in a wide range of research projects that use both existing corpora and those under development at UCLES. These projects are described as corpus-informed rather than corpus-based as other methodologies are used alongside corpus linguistic techniques, for example qualitative analyses. There are three main types of corpus-informed research activities at UCLES: developing examination materials, standardisation across examinations and comparative activities. The first group of activities is the most operationally significant and applies to specific examinations, with projects in the other two groups having

longer-term goals although being relevant across the different examination suites that UCLES offers.

Developing examination materials

Within this area corpora are being used to aid the development of examination materials. Native speaker corpora (e.g. the British National Corpus) are being used to investigate collocations, authentic stems and appropriate distractors which enable item writers to base their examination tasks on real texts. A second area is the continual revision process that all exams are subject to, for example revised CPE starting in December 2002 and CELS from June 2002.

Standardisation across examinations

Standardisation takes place both across and within the five ALTE/Cambridge levels and is at the heart of implementing valid and reliable examinations. Questions such as 'How does competence in one skill relate to competence in another?' can be answered with reference to corpus data, as the CLC, for example, provides an archive of previous performances on a range of examinations by candidates from many different backgrounds.

Different suites of exams also need to be standardised, especially those suites not testing general English such as IELTS, BEC or Young Learners. These suites need to be checked against Main Suite examinations, to ensure that their placing on a conceptual framework is appropriate and that an accurate picture of candidate performance at a particular level is produced.

Comparative activities

The third type of corpus-informed activity involves using small collections of scripts from UCLES' examinations to investigate the different types of English that our candidates are expected to produce. In relation to this, UCLES is involved in attempting to characterise learner language at different levels in both speaking and writing. The Common Scale for Speaking has already been developed in this regard, with the Common Scale for Writing under development following research by Roger Hawkey, Chris Kennedy and others (see Hawkey, 2001). One extremely useful

method of analysing texts has been the use of the WordSmith Tools software package (Stubbs, 2001). This package analyses any text or group of texts in statistical ways, providing frequency counts, wordlists and other user-friendly statistics that can indicate the best way to describe a group of texts or specify their commonality and differences.

In the future UCLES will seek to characterise Business and Academic English as sub-sets of General English using corpus data.

Conclusion

Corpus Linguistics is both a research methodology and a theory and therefore lends itself to the diverse demands that UCLES makes of it, in terms of the depth and breadth of research described here. Corpora are increasingly used in many academic and applied linguistic fields besides language testing and its breadth of application ensures that it continues to be one of the, if not the, fastest moving area of linguistics today. UCLES hopes to increase the applications of corpora in its testing and assessment activities and considers itself to be amongst the first to lead the way for the rest of the testing community in doing so.

References and further reading

Ball, F (2001): Validating EFL Examinations Using Corpora. Poster presented at British Association of Applied Linguistics Conference 2001, Reading, UK

Boyle, A and Booth, D (2000): The UCLES/CUP Learner Corpus, in *Research Notes 1*

Granger, S (Ed) (1998): *Learner English on Computer*, London: Longman

Halliday, M A K (1993): 'Quantitative Studies and Probabilities in Grammar'. In Hoey, M (ed) *Data, Description, Discourse. Papers on the English Language in Honour of John McH Sinclair*, London: Harper Collins

Hawkey, R (2001): Towards a common scale to describe L2 writing performance, in *Research Notes 5*

Oakes, M P (1998): *Statistics for Corpus Linguistics*, Edinburgh: Edinburgh University Press

Partington, A (1998): *Patterns and Meanings: Using Corpora for English Language Research and Teaching*, Amsterdam: John Benjamins

Sinclair, J (1991): *Corpus Concordance Collocation*, Oxford University Press

Tognini-Bonelli, E (2001): *Corpus Linguistics at Work*, Amsterdam: John Benjamins.

Selected Corpora

Cambridge Learner Corpus
    http://www.uk.cambridge.org/elt/reference/clc.htm

British National Corpus http://info.ox.ac.uk/bnc/

Cobuild Corpora http://titania.cobuild.collins.co.uk/

WordSmith Tools http://www.liv.ac.uk/~ms2928/wordsmith/

Revising the IELTS Speaking Test: retraining IELTS examiners worldwide

Lynda Taylor, Senior Research and Validation Co-ordinator, UCLES EFL

Issues 4 and 5 of *Research Notes* (February and July 2001) reported on the IELTS Speaking Test Revision Project, and in particular on the development of the assessment criteria, rating scales, test format and task design. This final article in the series reports on the worldwide examiner (re)training programme which was carried out in preparation for the introduction of the revised test format in July 2001. (A brief background to the International English Language Testing System was given in *Research Notes 4* so will not be repeated here; further information on IELTS is available via the website: www.ielts.org)

Background

Previous articles in *Research Notes* have described the methodology UCLES EFL employs for test development and revision (see Taylor in Issues 2 and 3, and Saville in Issue 4); one of those articles highlighted the special challenges posed by performance-based tests (i.e. speaking and writing), not only because of the many factors involved such as test format, task design, assessment criteria, and rating scales, but also because of the need to prepare examiners in readiness for the revised test becoming live; this means the revision project team must develop appropriate methods and materials for retraining and standardising existing examiners, often in large numbers and in many places.

Stuart Shaw's introductory article on issues in L2 writing assessment (see page 2) raises some key points which relate just as much to L2 speaking assessment. The importance of examiner training is well established in the literature, both in terms of its impact on the reliability and validity of performance testing (Alderson, Clapham and Wall, 1995) and also in terms of the call for increased professionalisation in language testing (Bachman, 2000). Several studies in recent years have highlighted key questions about the function and effectiveness of existing rater training programmes (Meiron and Schick, 2000; Weigle, 1998; Lumley and McNamara, 1995; Wigglesworth, 1993). However, as Weigle pointed out in 1994, there remains relatively little practical and empirical research to inform the development of an effective training programme; in addition, few test producers describe their procedures for examiner training in any degree of detail.

Planning and management

In the case of IELTS, the three partners – UCLES EFL, The British Council and IELTS Australia – traditionally share the responsibility for managing IELTS examiner training, including any retraining necessary because of test revision. When the plan for the IELTS Speaking Test Revision Project was first drawn up in 1998, it made provision for UCLES EFL to produce the examiner training materials and also to arrange for the first wave of training to train Senior Trainers at a regional level; The British Council and IELTS Australia routinely co-ordinate the IELTS examiner resource at centre level, so it was agreed they would arrange for examiner training to be cascaded to the local level via their worldwide test centre network and using their team of IELTS Trainers.

Production of examiner training materials

During the second half of 2000, a comprehensive set of examiner training materials for the revised IELTS speaking test was developed. These were prepared by the IELTS Chief Examiners and Senior Examiners in the UK and Australia in close consultation with UCLES EFL; all the personnel involved had extensive experience of working with the earlier training materials package and they had also been directly involved in developing the revised speaking test. The new set of materials included:

- an *IELTS Examiner Induction Pack* with accompanying video and worksheet;
- an *IELTS Examiner Training Pack*, with two accompanying videos and detailed Notes for Trainers.

Content and format

The content and format of the IELTS Induction and Training Packs draw upon previous practice in IELTS examiner training; but they have also been informed by the wealth of experience gained over recent years in inducting and training oral examiners worldwide for the various Cambridge EFL speaking tests. Both packs were designed to be suitable for immediate use in retraining existing examiners for July 2001, but also appropriate for training new IELTS examiners after July 2001.

IELTS examiner (re)training takes place during a face-to-face training session lasting a minimum of 7 hours. Before attending the training day,

trainees receive the IELTS Induction Pack to watch at home or in their local test centre; the induction video and worksheet help to familiarise them in general terms with the test format and procedures. The programme for the actual training day includes: a detailed focus on test format and procedures; peer-practice activities in handling the test materials; an explanation of the assessment criteria and rating scale descriptors; rating practice with volunteer candidates; viewing of video extracts for each test part as well as whole video performances. The training day ends with the trainees being asked to rate one or two video performances as a practice exercise; these ratings are then collected in and checked by the Trainer to monitor standards of performance in rating and identify any problem areas.

Delivering the retraining

Between January and March 2001, a small team of experienced IELTS Senior Trainers delivered examiner retraining to over 60 IELTS Trainers in fifteen regional locations around the world. During the early training sessions in February 2001 the Training Pack was 'trialled' with a small number of Trainers around the world; this meant that minor adjustments could be made to the final edition used from March onwards. Once Trainers had been retrained as IELTS examiners, they then delivered retraining to groups of IELTS examiners at a local level within their area. By the end of June 2001, over 1500 IELTS examiners had attended at least 150 face-to-face retraining sessions carried out in most of the 105 countries where IELTS is currently on offer.

From March 2001, queries and other comments began to feed back to the IELTS partners and were collated by the project team based at UCLES. This led to the development of an FAQ (Frequently Asked Questions) document which was circulated to all Trainers in May 2001 to provide helpful clarification and additional notes for guidance where necessary.

Feedback from IELTS trainers and examiners

The IELTS Examiner Training Pack included feedback questionnaires for Trainers and examiners inviting comments on their experience of using the materials. Completed forms were returned to UCLES EFL and were then analysed to help evaluate the usefulness of the training programme. By late September 2001 75 Trainer feedback forms had been returned for analysis and results showed that over 90% of Trainers considered the Training Pack

to be 'very good' or 'fairly good'; any concerns expressed related primarily to aspects of timings for the day, and to features of the training materials layout (e.g. size of print). Over 1000 examiner feedback forms were returned and analysed: 99% of examiners reported the training session to be 'very good' or 'fairly good' and 88% of examiners considered the guidelines in the Instructions to Examiners booklet to be 'very good' or 'fairly good'; 96% of examiners described the explanation of assessment procedures and criteria as 'very good' or 'fairly good', and similar figures reported finding the video profiles (96%) and the practice session with volunteer candidates (95%) either 'very helpful' or 'fairly helpful'. Examiners expressed some concern about the time available to cover everything in the training session.

On the whole, feedback from both Trainers and examiners was very positive and this is one measure of the success of the worldwide (re)training programme. Suggestions for improvement to the training materials will feed into the production of a second edition of the Examiner Training Pack sometime in 2002.

Some additional materials have been developed as part of the IELTS examiner training strategy. These include:

- two *IELTS Examiner Certification Sets* (to enable examiners to gain certificated status following attendance at a training session);
- an *IELTS Self-access Standardisation Pack*, with video and worksheets (a form of 'norming pack' to provide examiners with material for extra rating practice prior to live examining).

Certification of IELTS examiners

After IELTS examiners have attended a face-to-face training session they are routinely asked to rate a set of speaking test performances in order to demonstrate the accuracy of their assessment. An examiner must mark to an acceptable standard in order to receive certificated status and be qualified to examine for a period of 2 years. All examiners retrained so far this year will be required to complete a certification set by the end of October 2001. These certification ratings will be analysed by UCLES EFL to monitor examiner reliability and will also be compared with the ratings gathered at the end of the retraining event; they will also be used to investigate aspects of the criteria and scale functioning as part of the ongoing validation programme for the revised IELTS Speaking Test.

Other studies will focus on the reactions of test users (i.e. examiners, candidates, IELTS administrators) to the revised Speaking Test format. We hope to publish the results of these studies in future issues of *Research Notes*.

Conclusion

The exercise to retrain and standardise over 1500 IELTS examiners worldwide within a 4-5 month period has required considerable investment on the part of all three IELTS partners, not only in terms of professional input but also in terms of the logistical expertise and financial support needed. In the longer term the worldwide network of trainers and examiners established as a result of this retraining activity will be developed into a comprehensive professional support system for IELTS examiners; this system will include procedures for the regular co-ordination (i.e. standardisation) and monitoring activities already in place for the Cambridge speaking tests. In this way, we can ensure that the IELTS speaking test continues to be, in Alderson et al's terms, a 'quality instrument' for assessing L2 spoken language ability.

Acknowledgements

We would like to acknowledge the involvement of Kate Ingham, Val Sismey, Juliet Wilson and Nick Charge in the production of the IELTS training materials, as well as the contribution made by the UCLES Video Unit and the teams of Chief Examiners and Senior Examiners in the UK and Australia.

References and further reading

Alderson, J C, Clapham, C and Wall, D (1995): *Language Test Construction and Evaluation*, Cambridge University Press

Bachman, L F (2000): Modern language testing at the turn of the century: assuring that what we count counts, *Language Testing* 17/1, 1-42

Lumley, T and McNamara, T F (1995): Rater characteristics and rater bias: implications for training, *Language Testing*, 12/1, 54-71

Meiron, B and Schick, L (2000): Raters, ratings and test performance: an exploratory study. In Kunnan, A (Ed) *Fairness and validation in language assessment. Selected papers from the 19th Language Testing Research Colloquium, Orlando, Florida*. Studies in Language Testing 9, UCLES/CUP, 153-176.

Weigle, S C (1994): Effects of training on raters of ESL compositions, *Language Testing*, 11/2, 197-223

Weigle, S C (1998): Using FACETS to model rater training effects. *Language Testing*, 15/2, 263-287

Wigglesworth, G (1993): Exploring bias analysis as a tool for improving rater consistency in assessing oral interaction, *Language Testing*, 10/3, 305-335

The IELTS Impact Study:
development and implementation

Roger Hawkey, Consultant, UCLES EFL

As the consequential validity of tests (e.g. Messick, 1998) and the ethical responsibilities of testers (Spolsky, 1997; Hamp-Lyons, 1997; Shohamy, 1999) become a growing focus of attention, interest in the impact of language tests provided by major testing agencies and examination boards increases. In *Research Notes 2* (August 2000), Nick Saville's paper 'Investigating the impact of international language examinations' reported on the IELTS Impact Study (IIS), a major long-term programme of research by UCLES EFL into the impact of IELTS which is one of the most widely-used of language tests for those needing to study or train in the medium of English. This follow-up paper focuses on developments in the IELTS Impact Study over the past year.

In *Research Notes 2*, Nick Saville traced the development of the IIS from its inception in 1995 and the initial work undertaken in conjunction with Professor Charles Alderson of the Department of Linguistics and Modern English Language at Lancaster University and his research group (see, for example, reports to UCLES by: Alderson and Banerjee, 1996; Banerjee, 1996; Bonkowski, 1996; Herrington, 1996; Horak, 1996; Milanovic and Saville, 1996; Winetroube, 1997; Yue, 1997). As noted by Saville, the Lancaster team conceived the Study in terms of four sub-projects:

- Project 1 observing the context and nature of classroom activity in IELTS classes;
- Project 2 evaluating the content and nature of IELTS-course teaching materials;
- Project 3 surveying the views and attitudes of users groups towards IELTS, and
- Project 4 profiling the IELTS test-taking population.

By mid-2000, the Study had reached the end of Phase 2 of its three-phase schedule.

Table 1: IELTS Impact Study phases

Phase 1	1995-1996	Identification of areas to be targeted and development of instrumentation to collect information which allows impact to be measured
Phase 2	1997-2000	Validation and rationalisation of instruments
Phase 3	2001	Implementation of the instruments as part of a major survey

The development and validation of the instrumentation for the four sub-projects through Phases 1 and 2 had resulted in a total of 13 draft questionnaires, schedules or summary sheets intended for use in investigating the characteristics and attitudes of key IELTS stakeholders. In Phase 2 of the IIS, all these data collection instruments were trialled according to a model borrowed from language testing in order to establish the validity, reliability and practicality of the instruments, using both quantitative and qualitative methods. Kunnan (2000), in his analysis of some of the draft IIS data collection instruments, summarises such an approach as shown in Figure 1 opposite.
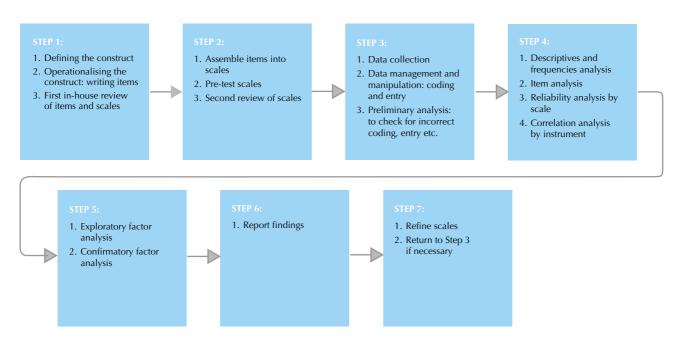
Following the early small-scale trials of the draft classroom observation instrumentation noted by Nick Saville, recent validation work on IIS Project 1 has aimed at producing a final version of a classroom observation and feedback instrument for use in Phase 3 in late 2001 and early 2002. As a result of intensive focus sessions involving other researchers on test impact in the classroom (e.g. Anthony Green, 2001) the original classroom observation package, in particular its content and practicality, has been revised; the result is a simpler version, based more closely on the COLT model (Spada and Frohlich, 1995), possibly for use on video-recorded IELTS-related classes in a sample of global centres round the world.

The initial pilot version of the IIS Project 2 classroom materials evaluation instrument was first trialled with experienced ESOL teachers and then in focus group discussion; as a result, it has been submitted to two significant revisions (see Hawkey and Saville, 2001). The significantly rationalised and shortened third version of the instrument for the analysis of textbook materials is now ready for use in the implementational phase of the IIS.

As Nick Saville reported in his earlier paper, seven questionnaires were originally developed to explore the views and attitudes of a wide population of IELTS users, namely:

1. students preparing for IELTS
2. teachers preparing students for IELTS
3. teachers preparing students for academic study (post-IELTS)
4. IELTS administrators
5. admissions officers in receiving institutions
6. students who have taken IELTS
7. academic subject teachers

Figure 1: Validation model for questionnaires using scales

STEP 1:	STEP 2:	STEP 3:	STEP 4:
1. Defining the construct	1. Assemble items into scales	1. Data collection	1. Descriptives and frequencies analysis
2. Operationalising the construct: writing items	2. Pre-test scales	2. Data management and manipulation: coding and entry	2. Item analysis
3. First in-house review of items and scales	3. Second review of scales	3. Preliminary analysis: to check for incorrect coding, entry etc.	3. Reliability analysis by scale
			4. Correlation analysis by instrument

STEP 5:	STEP 6:	STEP 7:
1. Exploratory factor analysis	1. Report findings	1. Refine scales
2. Confirmatory factor analysis		2. Return to Step 3 if necessary

Working from Kunnan's 1999 analyses and revision proposals, with additional feedback from researchers on related projects (e.g. Green, 2001), the user-group questionnaires are now being revised. The first draft revised instrument is a modular student characteristic and test attitudes questionnaire (see Hawkey, 2001), combining questionnaires 1 and 6 from Project 3 with the test-taker characteristics instrument from Project 4.
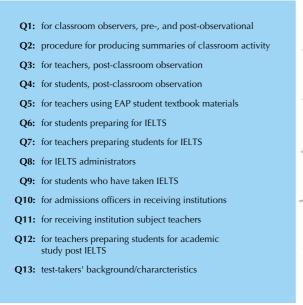
The revised test-taker characteristics instrument developed for the Impact Study, eliciting data on attitude, motivation and cognitive/meta-cognitive features and trialled in Phase 2 of the IIS with IELTS candidates, had been worked on by UCLES validation staff in collaboration with Jim Purpura; Purpura (1999) documents the use of structural equation modelling (SEM) as a method of validating the factors targeted by the instrument. Once again, qualitative validation advice has been sought in recent months from Purpura himself (see Purpura and Saville, 2001) and from Anthony Green, who is developing related instrumentation in his research into IELTS-takers and EAP pre-sessional course participants. As a result, the test-taker characteristics and attitudes instruments have been further refined, and incorporated into the modular student characteristic and test attitudes questionnaire (see Hawkey, as cited above).

While the IELTS Impact Study was originally conceived as four sub-projects, each with a focus on an area of crucial importance in considering the impact of the test and each with its own data collection instrumentation, the current process of validation and rationalisation is leading to a more fluid and integrated view of the sub-projects. It is likely that the 13 original questionnaires will actually be administered in the form of four modular instruments, as conceptualised in Figure 2 overleaf.

In Phase 3 of the Impact Study, four revised modular data collection instruments – see QA to QD in Figure 2 – will be used with appropriate samples of respondents to survey IELTS stakeholders world-wide; the results will be compiled in 2002.
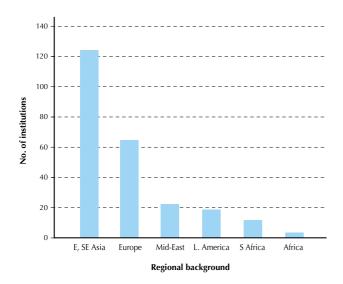
In May 2001, IIS pre-survey questionnaires were sent to a sample of over 300 institutions using IELTS (universities, centres, British Council and IDP), seeking up-to-date information on student preparation for the test and the textbooks used. Responses to this pre-survey (as high as 65% and from 41 countries) have already been analysed (see Hawkey, 2001b) and will inform decisions on where and when to implement the various data collection instruments and activities for the study. Figure 3, taken from the

Figure 2: Rationalisation of original IIS data collection instruments

Q1: for classroom observers, pre-, and post-observational

Q2: procedure for producing summaries of classroom activity

Q3: for teachers, post-classroom observation

Q4: for students, post-classroom observation

Q5: for teachers using EAP student textbook materials

Q6: for students preparing for IELTS

Q7: for teachers preparing students for IELTS

Q8: for IELTS administrators

Q9: for students who have taken IELTS

Q10: for admissions officers in receiving institutions

Q11: for receiving institution subject teachers

Q12: for teachers preparing students for academic study post IELTS

Q13: test-takers' chararcteristics

QA: classroom observation

QB: student characteristics and test attitudes

QC: teacher test attitude and preparation materials evaluation

QD: receiving institution administrators, admissions officers, subject teachers test views

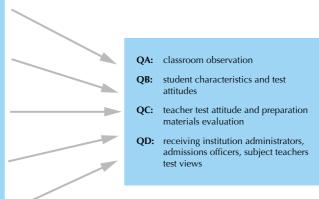(from Saville and Hawkey, forthcoming)

pre-survey report, illustrates IELTS preparation course participation by regional backgrounds and numbers of institutions across regions.

Figure 3: Institutions providing IELTS preparation courses by regional background.

Contact and collaboration with related IELTS impact research are being maintained, including those (e.g. Green, 2001, and Read and Hayes, 2000) under the British Council/IELTS Australia joint-funded research program.

As well as being the focus of UCLES presentations planned for international conferences in 2002, the IIS is described in a chapter in the forthcoming volume *Context and Method in Washback Research: The Influence of Language Testing on Teaching and Learning,* Liying Cheng, Yoshinori J. Watanabe (Eds.). The full final report of the IELTS Impact Study study will be published as a volume in the UCLES/CUP *Studies in Language Testing* series.

References and further reading

Alderson, J C, and Banerjee, J (1996): How might impact study instruments be validated? A paper commissioned by the University of Cambridge Local Examinations Syndicate as part of the IELTS Impact Study

Banerjee, J (1996): The design of the classroom observation instruments, UCLES Internal Report, Cambridge

Bonkowski, F (1996): Instrument for the assessment of teaching materials, unpublished MA assignment, University of Lancaster

Green, A (2000): Test impact and EAP: a comparative study in backwash between IELTS preparation and university pre-sessional courses. Research in progress for the PhD degree of the University of Surrey at Roehampton

Hamp-Lyons, L (1997): Washback, impact and validity: ethical concerns, *Language Testing* 14/3, 295-303

Hawkey, R (2001): IIS Student Questionnaire, UCLES EFL Validation Group Internal Paper

Hawkey, R (2001): IELTS Impact Study Pre-Survey: Draft Analyses, UCLES Internal Report

The paired speaking test format: recent studies

Herrington, R (1996): Test-taking strategies and second language proficiency: is there a relationship? Unpublished MA dissertation, Lancaster University

Horak, T (1996): IELTS Impact Study Project. Unpublished MA assignment, Lancaster University

Kunnan, A (1998): An introduction to structural equation modelling for language assessment research, *Language Testing* 15/3, 295

Kunnan, A (2000): IELTS impact study project report prepared for the University of Cambridge Local Examinations Syndicate

Messick, S (1989): Validity. In R L Linn (Ed) *Educational Measurement.* Third edition. New York: ACE and Macmillan: 13-103.

Milanovic, M, and Saville, N (1996): Considering the impact of Cambridge EFL examinations, UCLES Internal Report

Purpura, J (1999): *Learner Strategy Use and Performance on Language Tests*, Cambridge University Press

Purpura, J and Saville, N (2001): Developing a computerised system for investigating non-linguistic factors in L2 learning and test performance. Paper delivered at the ALTE Year of Languages Conference, Barcelona

Read, J and Hayes, B (2000): The Impact of the IELTS Test on Preparation for Academic Study in New Zealand. Report for the IELTS Research Committee.

Saville, N and Hawkey, R (forthcoming): A study of the impact of the International English Language Testing System with special reference to its washback on classroom materials. In Liying Cheng, Yoshinori J. Watanabe (Eds):*Context and Method in Washback Research: The Influence of Language Testing on Teaching and Learning*

Shohamy, E (1999): Language Testing: Impact. In Spolsky, B (1999)

Spada, N, and Frohlich, M (1995): *COLT Communicative Orientation of Language Teaching. Coding conventions and applications*, Sydney: NCELTR

Spolsky, B (1997): The ethics of gatekeeping tests: what have we learned in a hundred years? *Language Testing*, 14/3, 242-247

Spolsky, B (ed) (1999): *Concise Encyclopedia of Educational Linguistics*, Oxford: Pergamon

University of Cambridge Local Examinations Syndicate, the British Council, IDP Education Australia. (1998). *The IELTS Annual Review 1997/8*

Winetroube, S (1997): The design of the teachers' attitude questionnaires, UCLES Internal Report

Yue, W (1997): An investigation of textbook materials designed to prepare students for the IELTS test: a study of washback. Unpublished MA dissertation, Lancaster University

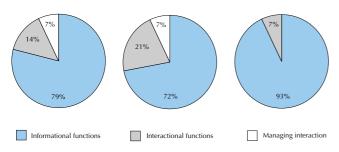**Lynda Taylor**, Senior Research and Validation Co-ordinator, UCLES EFL

In January 1999 the English Language Teaching Journal published an article by Michael Foot (*"Relaxing in pairs"* – ELTJ, Vol 53/1) in which he criticised the paired format adopted by the Cambridge EFL speaking tests. A key criticism concerned the extent to which the paired speaking test format was supported by research evidence, especially in relation to the quality of candidate performance and the candidate's affective reaction to the test format.

The decision from the early 1990s to adopt the paired format as standard for most of the Cambridge EFL speaking tests was based partly on pedagogical considerations (i.e. a more communicative approach to language teaching with the use of pair/groupwork in the classroom), but also on the findings of various studies of spoken language discourse. A previous article in *EFL Research Notes* (Issue 2, August 2000) highlighted some key studies which had informed this decision; several of these confirmed what Hughes (1989) had described as 'at least one potentially serious drawback' of the traditional one-to-one interview format: the power relationship which exists between tester and candidate. Ross and Berwick (1992) and Young and Milanovic (1992) showed how examiner-candidate discourse could be highly asymmetrical in terms of features of dominance, contingency and goal-orientation. The fixed role relationship (examiner-candidate) in a one-to-one test makes it difficult for the candidate to escape this asymmetry. The paired format, on the other hand, provides the potential for various interaction patterns: candidate-examiner; candidate-candidate; and among the three participants.

In 1999 two internal studies undertaken as part of the CPE Revision Project set out to compare the paired and one-to-one speaking test formats for CPE. An initial quantitative study suggested that the paired format was capable of generating a much richer and more varied sample of spoken language from each candidate than is usually produced in the one-to-one format (Taylor, 1999). In a second, more qualitative study (ffrench, 1999) an observation checklist was used to analyse the distribution of certain speech functions (or operations) across the individual and paired format speaking tests. Drawing on the established literature on L2 speaking ability, the observation checklist identified a total of thirty communicative language functions which characterize spoken discourse; these can be broadly categorized as *informational* (e.g. expressing opinions),

interactional (e.g. persuading), and to do with *managing interaction* (e.g. terminating a discussion). (The development of the checklist instrument was reported in *EFL Research Notes*, Issues 2 and 3).
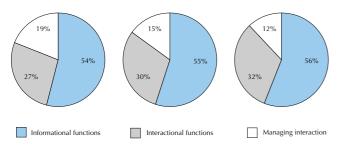
Figure 1: Distribution of speaking functions in individual test format

(taken from ffrench, 1999)

Figure 1 shows the results of an analysis of three separate <u>one-to-one</u> CPE interviews; the pie charts show the distribution of *informational*, *interactional* and *managing interaction* functions across each candidate's speech in the three tests. The dominance of *informational* functions (accounting on average for over 80% of the candidate's spoken discourse) can be seen clearly, and this is likely to be a direct result of the examiner-candidate asymmetry which previous analysts had highlighted. There is some evidence of *interactional* functions being used, but very little evidence of *managing interaction* functions (and in one test no evidence at all).

Figure 2: Distribution of speaking functions in paired test format

(taken from ffrench, 1999)

Figure 2 shows the results of the same analysis applied to three <u>paired</u> format CPE interviews. In this set of tests *informational* functions account for on average 55% of the candidates' spoken discourse in the test. The

remainder of the candidates' speech is made up of *interactional* and *managing interaction* functions, with plenty of evidence that both types of talk are occurring.

Overall, the range of functions elicited by the paired speaking test format proved to be much greater than for the one-to-one format (26 functions out of a possible 30 for the paired format, compared with 14 out of 30 for the one-to-one); in the paired format, then, the impact of examiner-candidate asymmetry seems to be reduced, the level of 'interactiveness' appears to be potentially much higher and the language sample elicited for assessment purposes is as a result considerably richer.

Recently two independent studies have provided additional empirical support for the paired speaking test format. In January 2001 the English Language Teaching Journal published a further article on the subject of testing speaking: *"Oral testing in pairs – a secondary perspective."* (ELTJ, 55/1). The writers, Gyorgi Egyud and Philip Glover, argue in favour of paired oral testing on the basis of their experience and study of oral testing of English in the Hungarian secondary school system. They comment as follows:

- *Students like pairings* (according to student responses to a questionnaire indicating preference for paired/single format)

- *Pairings give students a better opportunity to produce their best* (according to paired and singleton transcript evidence for the same speaking test task which suggests the paired candidates have the chance to produce better output)

- *Pairings help produce better English than one-to-one* (given the interrogative nature of the one-to-one interview due to the unequal examiner/test-taker relationship)

- *Pairings support good teaching* (good washback from teaching to testing and vice versa is important)

The writers conclude: *"We are convinced that the paired format offers students and teachers opportunities for development, and an escape route from the prison of dire one-to-one situations."*

In a separate study, completed as part of a Master's in Education at the University of Manchester, Ann Humphry-Baker focused on the relatively under-researched area of candidate perceptions. Ann is a Team Leader in Switzerland for the Cambridge EFL Speaking Tests and her dissertation, entitled *Speaking Tests: Students' Perception and Performance*, set out to investigate 'how test-takers feel about the speaking test'. Using

European language testing in a global context

questionnaires with 130 candidates who had taken a variety of Cambridge EFL tests (PET, FCE, CAE, CPE or CEIBT), Ann explored their affective reactions to the test. Candidates agreed or strongly agreed with the majority of the twelve statements in her questionnaire, including *I like paired tests* and *The test gave me a good opportunity to speak*; a higher level of disagreement came in response to statements such as *I had enough time to speak*; *I performed well in the test*; *I enjoyed the test*. Ann's study is relevant for students, teachers and oral examiners, and provides the basis for valuable awareness-raising activities in the classroom or in examiner training/coordination sessions.

References and further reading

Egyud, G and Glover, P (2001): Oral testing in pairs – a secondary perspective, *English Language Teaching Journal*, 55/1, 70-76

Foot, M (1999): Relaxing in pairs, *English Language Teaching Journal*, 53/1, 36-41

ffrench, A (1999): Study of Qualitative Differences between CPE Individual and Paired Test Formats, Internal UCLES EFL Report

Hughes, A (1989): *Testing for Language Teachers*, Cambridge University Press

Humphry-Baker, A (2000): *Speaking Tests: Students' Perception and Performance*. Unpublished MEd Dissertation, University of Manchester

Ross, S and Berwick, R (1992): The discourse of accommodation in oral proficiency interviews, *Studies in Second Language Acquisition*, 14/2, 159-176

Taylor, L (1999): Study of Quantitative Differences between CPE Individual and Paired Test Formats, Internal UCLES EFL Report

Young, S and Milanovic, M (1992): Discourse variation in oral proficiency interviews, *Studies in Second Language Acquisition*, 14/4, 403-424

Marianne Hirtzel, European Projects Officer, UCLES EFL

Bringing language testers together

The Association of Language Testers in Europe – ALTE – is an association of institutions within Europe, each of which produces examinations and certification for language learners. Each member provides examinations of the language which is spoken as a mother tongue in their own country or region. The concept of ALTE was initially formed by the Universities of Cambridge and Salamanca late in 1989, and at the first meeting of the association in 1990 there were eight founder members. Since then membership has grown so that there are now 27 members, representing 24 European languages (for more details see the ALTE website – www.alte.org).

2001 was designated as the European Year of Languages and ALTE members decided to mark the year by organising a conference on language testing. As the title of the conference reflects, the aim was not only to look at language testing in Europe, but also at what is happening in other parts of the world. Participants had the opportunity to exchange ideas, present the results of research projects and develop their knowledge. Over 350 delegates attended, coming from over 35 countries and the conference was unique in being the largest on language testing ever held in Europe.

The conference took place from 5 to 7 July 2001 in a newly restored building of the Universitat Pompeu Fabra in Barcelona and was organised locally by the Generalitat de Catalunya (also a member of ALTE). The venue and facilities were excellent, with plenty of space and well-equipped presentation rooms. The conference was officially opened by Jordi Vilajoana, Minister for Culture in the Generalitat de Catalunya, who welcomed delegates and highlighted the importance of linguistic education, and by Michael Milanovic, Manager of ALTE, who spoke about the work of ALTE in language testing.

Main themes of the conference

The main themes of the conference covered the relationship between new technology and testing, as well as innovations in testing using Information Technology, the use of examinations within educational systems, and European projects such as the Common European Framework and the

European Language Portfolio. There was also a strand dealing with the testing of speaking. Other presentations were on the testing of younger learners and cultural considerations in language assessment.

Plenary presentations

There were 6 plenary presentations.

- **Joseph Sheils** (Modern Languages Division, Council of Europe, Strasbourg) gave the opening plenary on the language policy of the Council of Europe and the promotion of plurilingualism, which is a very important element of the work of the Modern Languages division. He mentioned various initiatives in connection with the European Year of Languages, especially the introduction of the European Language Portfolio and the Common European Framework. These are designed to help contribute to the development of language opportunities available to all and the European Year as a whole puts language policy centrally on the agenda.

- **Raffaele Sanzo** (Ministero della pubblica Istruzione, Rome) presented the Progetto Lingue 2000 in Italy, which is an example of how foreign language learning and testing is being dealt with within a national context. The presentation included sections on the way new learning groups have been set up, language learning objectives and the way the teaching practice is organised.

- **Anne Lazaraton** (University of Minnesota, USA) spoke about qualitative research methods in language test development and validation. She gave a detailed analysis of the nature of qualitative research and how this approach has been used, with concrete examples of work recently carried out.

- **Wolfgang Mackiewicz** (Freie Universitat Berlin, Germany) spoke about higher education and language policy in the European Union, outlining the importance given by the EU to the principle of linguistic and cultural diversity and examining how this can be supported in the face of increased globalisation and integration. He also spoke about the work of the European Language Council, which is working on many projects for the improvement of language studies in higher education.

- **John Trim** (Project Director for Modern Languages, Council of Europe, 1971-1997) gave a presentation on the Common European Framework (*Common European Framework of Reference for Languages: Learning, teaching, assessment*) and its implications for language testing. He highlighted the particular sections of the document which focus on language testing and which it is important for those involved in language testing to take into consideration in the testing process. These include a description of the many factors involved in language use, the calibration of language proficiency at a number of levels and methodological options for language testing and assessment.

- **Charles Alderson** (University of Lancaster, UK) gave the closing plenary, on the current state of language testing and future developments which may be predicted for the coming decade. He included an analysis of trends in language testing which have been developing in recent years and pointed out areas where he felt change should be considered. Many issues were raised which could contribute to future discussions.

Other presentations

The themes of the conference attracted a large number of proposals for presentations and the final programme offered a wide range of topics. As well as the plenary presentations, there were approximately seventy other presentations in parallel strands, each chaired by some ALTE members. ALTE member organisations were also well represented as presenters, with presentations on the testing of most European languages, including Dutch, Swedish, Polish, Irish and Spanish.

Presentations relating to particular European projects included one on the ALTE Can Do project by Neil Jones of UCLES, one on the EAQUALS-ALTE European Language Portfolio, a presentation on relating test scores to the Council of Europe Common European Framework by John de Jong of Language Testing Services and various presentations on the DIALANG project. Other contributors were Antony Kunnan from California State University, USA (*Articulating a fairness model*), John Read from Victoria University of Wellington, New Zealand, (*Investigating the impact of a high-stakes international proficiency test*) and Elana Shohamy from Tel Aviv University, Israel (*The role of language testing policies in promoting or rejecting diversity in multilingual/ multicultural societies*).

Publication of papers

All speakers at the conference have been invited to submit papers to be considered for selection for a publication, which is expected to include about eighteen papers. The remainder will be available on the ALTE web site at www.alte.org

Conference Exhibition

The conference included a large exhibition area, with materials produced by ALTE member organisations as well as other European organisations.

There was also a demonstration area where participants could try out various ALTE products:

The UCLES EFL seminar programme

- BULATS (Business Language Testing Service) CD produced by 4 ALTE members and available in English, French, German and Spanish

- ALTE Can Do CD with Can Do statements in 12 languages

- ALTE Multilingual Glossary of Testing Terms: 450 testing related terms in 12 languages

- Language Learner Questionnaire CD developed by University of Cambridge Local Examinations Syndicate (UCLES)

- Quick Placement Test developed by UCLES and published by Oxford University Press

Many participants welcomed the opportunity to share ideas with other language testers from a new perspective and within a new forum. In particular participants from outside Europe found it very valuable to see at first hand what is happening in Europe.

Other support for the European Year of Languages

UCLES Plurilingual certification award

In collaboration with ALTE, UCLES has decided to support the European Year of Languages by launching a competition to award certificates to language learners who achieve qualifications in a range of European languages. Prizes of 500 euros each will be awarded to the winners in four different categories (under 19 and over 19 in Europe; under 19 and over 19 outside Europe) who can show they have passed examinations from the ALTE framework in at least three different languages. Examinations must be passed between 1 December 2000 and 31 July 2002.

For full details of how to enter and an application form, look at the ALTE web site (www.alte.org) or the UCLES EFL web site (www.cambridge-efl.org).

The EFL Research and Validation Group contributes to ongoing staff development in a variety of ways. One of these involves the planning and management of a monthly seminar programme open to all UCLES EFL staff; the programme is designed to maintain and develop the staff's knowledge base in applied linguistics, testing and measurement issues, the development of systems/technology, business/marketing considerations, etc. as these apply to our examinations.

In addition to internally organised sessions, the programme regularly includes contributions from visiting speakers who are acknowledged specialists in their respective fields. Contact of this sort between our internal staff and external specialists is invaluable in ensuring that the issues facing us (and language testers in general) are more widely known and understood within the academic community; in addition, we benefit greatly from the input of experts in critical fields as we seek to formulate policy and implement good practice in relation to our language tests.

In May 2001 Dr Susan Hunston of University of Birmingham led a workshop session for EFL staff on 'Aspects of corpus use in applied linguistics'. Her workshop touched upon a number of themes: the main kinds of information to be obtained from a corpus, e.g. frequency, phraseology and collocation; work based on critical linguistics and some studies of variation between language varieties and registers; the importance of phraseology to current theories of language, particularly with regard to creativity, international uses of English, and the English of advanced learners. (See also Fiona Ball's article on page 6 in this issue describing the way corpus-based activities undertaken at UCLES help to inform the development and validation of our tests.)

In early July Professor John Read of University of Victoria, Wellington (NZ), gave two seminars to EFL staff. The first focused on testing vocabulary, following recent publication of his book *Assessing Vocabulary* (CUP, 2000). Professor Read acknowledged the increasing interest among researchers and language teachers in the acquisition of second language vocabulary, thereby creating a need for new assessment instruments; but he also highlighted the importance of a good conceptual framework for measuring vocabulary acquisition and use: *"Conventional vocabulary tests are widely used but they do not fit easily into contemporary models of language test development, where the performance task is seen as the basic unit of test design. A broad approach to vocabulary assessment is needed, one which*

incorporates the lexical component of integrated tasks as well as discrete vocabulary tests. It is also important to make decisions about the selection and design of vocabulary measures by considering in a systematic way the purpose of the assessment, whether it be for instruction, research, or programme evaluation."

Professor Read's second seminar reported on a recent study within the funded IELTS Research Program to investigate the impact of the IELTS test on preparation for academic study in New Zealand, in particular on the content and format of IELTS-focused preparation courses. (See also Roger Hawkey's article on page 12 in this issue describing recent progress in the IELTS Impact Study.)

Later in July Dr Lynne Cameron and Jayne Moon from the Department of Education at the University of Leeds led a seminar on the subject of Young Learners English. The last two decades have seen a sharp increase in the number of countries opting for an early start for English; national education ministry policies and parental aspirations are just two factors involved. This has had (and will continue to have) an impact upon the growth of the Cambridge Young Learner Tests of English (YLE).

Lynne Cameron argued that it is important to conceptualise the teaching of English to young learners (TEYL) in its own terms rather than automatically adopt constructs from traditional ELT which were developed for older learners, and she explored some of the implications of this for teaching and assessment. She highlighted two characteristics of children's learning as being central to such a conceptualisation – the search for meaning, and the primacy of spoken language. She suggested that what we know about these characteristics can inform ways of thinking about (1) the 'language' that is learnt by children, and (2) the process of literacy skills development in English.

Jayne Moon provided a brief overview of issues raised by some recent research into teaching English to young learners. This research consistently points to the importance of various conditions being in place if early TEYL programmes are to have a chance of being successful. She reported that one of the findings of such research is that learning activities/tasks play an important part in motivating children to engage in language learning. Activities/tasks mediated by teachers or more mature learners/other adults in the social setting of a classroom provide experiences and opportunities

for young learners which enable them to develop various language skills and strategies. She went on to discuss findings from current research into children's learning/language learning about the role of activities and the conditions which affect performance/learning on them in order to draw out implications for teaching and assessment.

We hope to include a special focus on the Cambridge assessment of Young Learner English in the next issue of Research Notes.

Further Information

UCLES provides extensive information on the examinations and assessment services referred to in this newsletter. For further information, visit the UCLES EFL website

www.cambridge-efl.org

or contact

EFL Information
University of Cambridge Local Examinations Syndicate
1 Hills Road
Cambridge CB1 2EU
United Kingdom

Tel: +44 1223 552734
Fax: +44 1223 553068
e-mail: eflinfo@ucles.org.uk

For information on the ALTE five-level scale and the examinations which it covers, visit the ALTE website **www.alte.org**

or contact

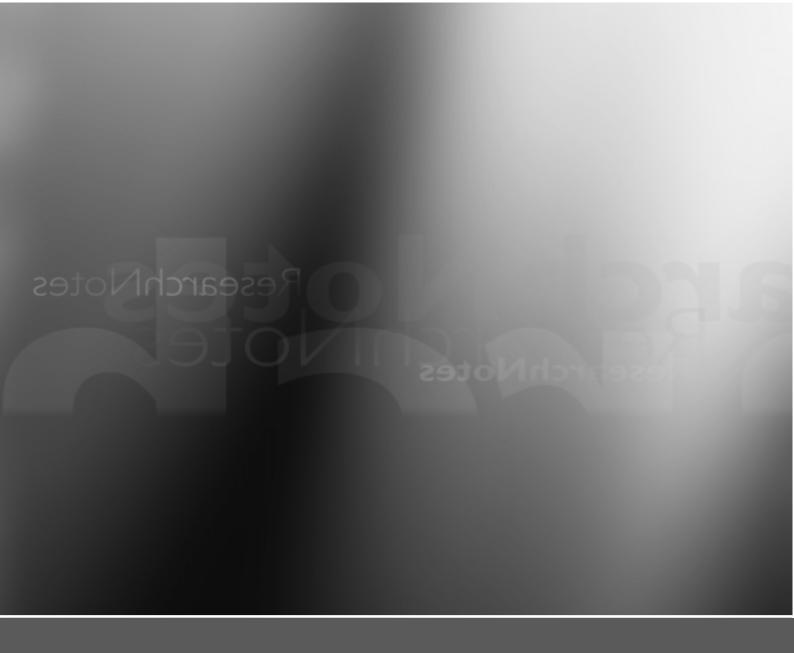The ALTE Secretariat
1 Hills Road
Cambridge CB1 2EU
United Kingdom

Tel: +44 1223 553925
Fax: +44 1223 553036
e-mail: alte@ucles.org.uk

If you would like further copies of this issue of Research Notes, or if you would like to be added to our mailing list (all registered UCLES centres receive copies automatically), please complete this form using block capitals and return it to EFL Information at UCLES. Please photocopy the form if you wish.

☐ Please send me ……. extra copies of this issue of Research Notes.

☐ Please add me to the Research Notes mailing list.

Name ………………………………………………………………………………

Job Title ……………………………………………………………………………

Institution …………………………………………………………………………

Address ……………………………………………………………………………

………………………………………………………………………………………

………………………………………………………………………………………

Country ……………………………………………………………………………

Please let us have any comments about Research Notes, and what you would like to see in further issues:

EMC/1268/2y11

ALTE

Association of Language Testers in Europe

EAQUALS

ASSOCIATE MEMBER