

4

Research Notes

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

FEBRUARY 2001

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

Tel: +44 1223 553355
Fax: +44 1223 460278
e-mail: efl@ucles.org.uk

www.cambridge-efl.org



UNIVERSITY of CAMBRIDGE
Local Examinations Syndicate

EAUWLS
ASSOCIATE MEMBER



© UCLES 2001

ResearchNotes

Introduction

Welcome to the fourth issue of *Research Notes*, the UCLES EFL newsletter about current developments in research, test development and validation issues.

Stakeholders expect examinations to be fair, accurate and above all reliable. In this issue Neil Jones addresses the issue of reliability in relation to the other features of UCLES EFL exams. Nick Saville continues the strand on test development which he began in the last issue. Here, he looks at the test development and revision process. The practical application of the issues raised by Nick Saville are considered by Lynda Taylor and Neil Jones in an article about the IELTS Speaking Test revision.

External researchers working in the field of language testing regularly approach UCLES asking permission for access to our data and materials to help in their projects. Lynda Taylor looks at the ways in which UCLES can help in these requests, and how researchers should approach UCLES.

Also in this issue is the announcement of the winners of the first IELTS MA dissertation award – as well as details of how to enter for this year's award.

In the next issue of *Research Notes*, Neil Jones will be following up his articles on the ALTE Can-do project and on the comparison of computer-based and paper and pencil tests. The forthcoming volumes in the Studies in Language Testing Series dealing with various UCLES revision projects will also be discussed.

Research Notes is intended to reach a wide audience of people involved in Cambridge examinations around the world and also people who are interested in the theoretical and practical issues related to language assessment. We would be very interested to hear your views on the newsletter – whether you find it interesting and useful, how appropriate you find the level of presentation and if there are any topics you would like us to cover. You can e-mail research.notes@ucles.org.uk or write to us at the address on page 18.

Research Notes is delivered to all UCLES EFL centres and other key contacts. If you would like to receive additional copies or if you would like a personal subscription to the newsletter, please complete and return the form on page 18.

Contents

Introduction	1
Reliability as one aspect of test quality	2
Test development and revision	5
Revising the IELTS Speaking Test	9
External requests for UCLES data/ materials	12
Studies in Language Testing	13
The UCLES EFL Seminar Programme	15
Announcement of the winners of the IELTS MA Dissertation Award	16

Reliability as one aspect of test quality

Neil Jones, Research Co-ordinator, UCLES EFL

Given their international recognition for a range of purposes, it is important that Cambridge Main Suite exams and IELTS should meet high standards of quality and fairness. The reliability of the exams and their sub-test components is a key aspect of fairness and therefore needs to be reported in appropriate ways to the users of the examinations. (See for example the *IELTS Users' Guide* – in press.)

Traditionally the quality of a test is assessed in relation to two key qualities: reliability and validity. Bachman and Palmer (1996) put impact and practicality alongside them as the factors which together decide a test's usefulness for a given context and purpose. (See the article on page 5 by Nick Saville for the application of these criteria in relation to the revision of UCLES EFL exams.)

'Reliability' is a word whose everyday meaning adds powerful positive connotations to its technical meaning in testing. Reliability is a highly desirable quality in a friend, a car or a railway system. Reliability in testing also denotes dependability, in the sense that a reliable test can be depended on to produce very similar results in repeated uses. But while this is a necessary condition of a good test it is by no means sufficient. A highly reliable test could at the same time be a poor measure of the skill of interest – in fact, it might be entirely uncorrelated with it. This is where validity comes in – does the test measure what we want it to measure?

The pursuit of high reliability has been a dominant goal of much testing, even though 'inconsistencies between higher reliability and better measurement' were identified as early as 1945 (Gulliksen 1945, Tucker 1946).

If the reliability of the items were increased to unity, all correlations between the items would also become unity and a person passing one item would pass all items and another failing one item would fail all items. Thus the only possible scores are a perfect score of one or zero... Is a dichotomy of scores the best that can be expected from a test with items of equal difficulty?

(Tucker 1946)

This 'attenuation paradox' – the fact that increasing reliability beyond a certain point actually decreases the information provided by a test – was named by Loevinger (1954).

To paraphrase Bachman (1990, p 161), reliability is concerned with minimising the effects of measurement error, while validity is concerned with maximising the effects of the language abilities we want to measure. Bachman represents reliability and validity as 'complementary aspects of a common concern in measurement – identifying, estimating, and controlling the effects of factors that affect test scores' (1990, p 160). However, it is evident that there is a potential tension between them. When high reliability is achieved, for example, by narrowing the range of task types or the range of skills tested, this restricts the interpretations that can be placed on performance in the test, and hence its validity for many purposes.

The use of reliability coefficients as evidence of test quality is also problematic to the extent that, strictly, the estimated reliability is not a feature of the test, but rather of an administration of the test to a given group of examinees. Candidates of widely-ranging ability are easier to rank reliably, and so will produce higher reliability indices than groups that are more equal in level. Reliability indices found for different tests should not be compared without taking into consideration factors such as this; nonetheless, it is difficult to expect most end users of such information to do more than accept it at face value.

Reliability in the context of Cambridge exams

In this section these general considerations are discussed in relation to the specific features of the Cambridge EFL exams.

The Cambridge exams are designed to achieve a particular balance of features which together account for their overall quality and usefulness in the contexts they are used. This section reviews some of these features and their impact on reliability; some of the differences between the certificated examinations and IELTS are also highlighted.

The Cambridge EFL Main Suite exams, such as FCE and CPE, have traditionally played a dual role: they test English language proficiency in a general sense, but they also fit into a pedagogical framework. Most candidates for these exams have attended classes as part of their exam preparation, and many will continue to study and take Cambridge exams at successive levels as they progress. There are many course-books and other material that relate to the exams. The Cambridge exam system has grown and developed over a period of many years, with new exams appearing as

the need was identified for a language test at a particular level. Because these exams test at a single level, they report scores in passing and failing grades.

The Cambridge EFL Main Suite exams, and newer exams such as the Business English Certificates (BEC), are designed to have a positive impact on language study, by encouraging good learning and teaching, and discouraging 'teaching to the test'. The aim is that good exam preparation and good language learning should be the same thing.

With each revision, Cambridge exams are modified to reflect current views on language learning and teaching. Thus recent revisions have given increasing attention to the communicative use of language, the better contextualisation of test items, and the authenticity of texts and tasks.

These are some of the features of Cambridge EFL Main Suite exams which are relevant to a consideration of reliability. Let us consider them in turn.

The Level System – exams which test at a level deal with a shorter ability range

Each Cambridge Main Suite exam is benchmarked to a criterion level of proficiency, i.e. covers a limited ability range within the ALTE/UCLES Framework of criterion levels. This needs to be reflected in the information we provide about the reliability of each exam.

The truncation of the ability range is particularly true of the highest level (CPE – Level 5), where there is evidently an upper limit to the distribution of candidates' ability. As there are fewer real shades of ability in the population, tests will tend to produce lower reliability indices, relative to tests which cover a wide range of the ability continuum (e.g. using internal consistency estimates).

IELTS, which is a non-certificated testing system, does not function in the same way; it covers a wider range of ability across its nine-band scale (which is used for both the General Training and Academic Modules); the estimates obtained for IELTS Reading and Listening tests reflect this and are comparable with other tests of this kind.

The Main Suite adopts a standard frame of reference to which exams can be related i.e. the ALTE Framework. This provides a comprehensive

description of levels of language proficiency, and is a more meaningful framework within which to estimate and report the reliability of the examinations which constitute the system (cf. discussions of consistency of classification). The end user, it can be argued, is most interested in how a Main Suite exam places a candidate at a criterion-level in relation to the overall ALTE 5-level system and a statistical indication of how accurately this can be done may be more useful than one which relates to the more local context of a single level.

The Grading System – most exams report results as grades or bands rather than a score on a continuous scale

Cambridge certificated exams report results as passing and failing grades, rather than a score on a continuous scale. Reporting results as grades raises the question of classification error. The alternative approach is to report a score on a continuous, standardised scale, together with an indication of the standard error of measurement (SEM). With this information the user of an exam result can exercise judgement as to the probability of that result being satisfactory for some purpose. It seems that users of UCLES exams are still largely happy with the notion of exam grades, and accept the fact that a few candidates are inevitably placed close to a grade boundary – just passing or just failing the exam because of the proximity of their final mark to the cut score.

The grading system for the certificated exams has a series of 'checks and balances' which are put into place following the provisional grading. This process – known as the Award – allows for borderline cases to be reviewed before pass/fail results are issued; for example this may mean Writing scripts being checked and, if necessary, remarked. The aim is to achieve a fair outcome for all candidates, including those with permanent or temporary handicapping conditions.

IELTS reports results on a nine-band scale (a global band and component level bands by skill – see IELTS Handbook). However, IELTS is not a certificated exam and does not have a pass/fail cut-off mark. Institutions using IELTS results for admission purposes set their own standards in terms of which band to accept for a given course of study. Clearly, a statement expressing SEM as a proportion of a band can help receiving institutions in setting requirements. The forthcoming IELTS user guide will include such statements.

The Construct – the model of English language proficiency which the exams operationalise is not homogeneous

The construct of overall English language proficiency which the Cambridge exams operationalise is heterogeneous. It reflects the view that a candidate's aggregate score over the whole range of language skills is the best, fairest measure of proficiency for exams of this type.

It is clearly a strength of the Cambridge EFL exams that they test a range of language skills, with several exams (FCE, CAE, CPE) having five different components lasting over 5 hours. However, the component skills are not highly inter-correlated, which sets practical limits on the possible composite reliability of the exam, (although this is typically higher than for any single component alone).

Authenticity – the exams seek to measure communicative language skills and to achieve high authenticity

In the objectively-marked papers task-based exercises have been replacing discrete point multiple-choice items in order to provide greater context and authenticity (both situational and interactional). However, a consequence of this is that the number of items in some papers might be limited to fit within practical time constraints. While this may be expected to bring about a small reduction in the estimated reliability, this is justified when the other qualities of the test are taken into account.

Most UCLES EFL exams contain Writing and Speaking components. UCLES has invested heavily to maintain standards of speaking assessment worldwide, through the continuous training of a large number of oral examiners (see Saville and Hargreaves, 1999). Traditional estimates of the accuracy and consistency of ratings are calculated using correlations, both inter- and intra-rater. These may be obtained operationally (where double-ratings are used) or by experimental methods. For the Cambridge Speaking tests double ratings are routinely made and inter-rater correlations are calculated (the correlation between first and second ratings for Main Suite exams are typically between 0.8 and 0.85).

Similar effort is devoted to training and standardising teams of writing assessors, and in the latter case it is also possible to scale assessors' marks to compensate for systematic differences in severity.

The interpretation of inter-rater correlations as estimates of reliability can be problematic; we know for example that correlations of this kind are affected by the nature of the assessment being made, e.g. the nature of the rating scale used and by the range of ability of the candidates who are assessed. When the ability range is narrow, small differences between examiners can affect the correlations which are obtained; this may on first inspection make the test seem unreliable. In practice it may not truly reflect the accuracy of the classification which is made in the case of criterion-related assessment. In other words the accuracy and consistency of the classification may be acceptable even though the inter-rater correlation is not high (e.g. around 0.65-0.70).

Conclusions

UCLES EFL recognises the need to provide end users with good evidence for the quality and fairness of their exams. Statistical evidence, including indices such as reliability coefficients or standard errors of measurement, is an important part of this, but, as this paper has argued, should be presented in context.

We are currently working to develop forms of report which are informative and helpful for end users. A great deal of effort has recently been invested by UCLES EFL into making the reporting of final grades and component-level performance as clear as possible to typical users (candidates, teachers, parents). (See for example a discussion of the new style statements of results in *Research Notes 3* – November 2000.)

It is not that there is a problem with achieving good reliability. Typical composite reliability of FCE, for example, is estimated at 0.94. IELTS is currently estimated to have a reliability of 0.94, with a SEM of 0.36 of a band (i.e. less than half a band). The issue is whether these figures give the full picture, and what other statistical indices might be more useful. An index of accuracy of classification within the ALTE Framework, for example, is a statistic which would be readily interpretable by end users.

At the same time as improving the reporting of statistical evidence, the aim is to present a picture where all of the qualities – validity, reliability, impact and practicality – are appropriately balanced (cf. Bachman and Palmer, 1996).

Test Development and Revision

References

- AERA/APA/NCME (1999): *Standards for educational and psychological testing*, Washington: AERA
- Bachman, L F (1990): *Fundamental considerations in language testing*, Oxford: OUP
- Bachman, L F and Palmer, A (1996): *Language testing in practice*, Oxford: OUP
- Gulliksen, H (1945): The relation of item difficulty and inter-item correlation to test variance and reliability, *Psychometrika* 10 (2), 79-91
- Loevinger, J (1954): The attenuation paradox in test theory, *Psychological Bulletin* 51, 493-504
- Saville, N and Hargreaves, P (1999): Assessing speaking in the revised FCE, *ELT Journal* 53 (1), 42-51
- Tucker, L R (1946): Maximum validity of a test with equivalent items, *Psychometrika* 11 (1), 1-13
- Wood, R (1993): *Assessment and testing*, Cambridge: CUP

Nick Saville, Manager, EFL Test Development and Validation Group, UCLES

Introduction

The test development procedures currently employed by UCLES EFL have been set out in a number of documents produced in the 1990s. For example a succinct summary of the approach is contained in the User's Guide for Examiners (Chapter 2) which was produced by ALTE on behalf of the Council of Europe (1997). Examples of this methodology in practice have been used within UCLES over the past decade and in projects for which UCLES have provided consultancy in other contexts. For example, the approach was used for the development of KET within UCLES (1992-4) and for the consultancy in China for the Public English Tests (PETS) Project (1997-2000), see Lynda Taylor, *Research Notes* 3 (November 2000).

Test development

The approach to test development is based on a cyclical model. In this approach it is important for the test developer to consider, as part of the process, how the VRIP qualities of Validity, Reliability, Impact and Practicality can be balanced in order to achieve the overall validity or utility of the examination and to ensure that it fulfils a useful purpose in the contexts it is to be used (cf. Bachman 1990, Bachman and Palmer 1996).

It is also important to think of the process of test development, not only as cyclical, but iterative. This involves feeding back the knowledge and experience gained at different stages of the process into a continuous re-assessment of a given test and each administration of it.

Figure 1 attempts to capture this process in diagrammatic form. The diagram offers a comprehensive blueprint for the stages that are gone through, beginning from the initial perception that a new test is necessary –

- perceived need for a new test
- planning phase
- design phase
- development phase
- operational phase
- monitoring phase.

Within each phase the work will involve a number of activities which themselves may be broken down into stages. For example the Development Phase involves validation activities, such as trialling and analysis, and the planning for implementation in the Operational Phase.

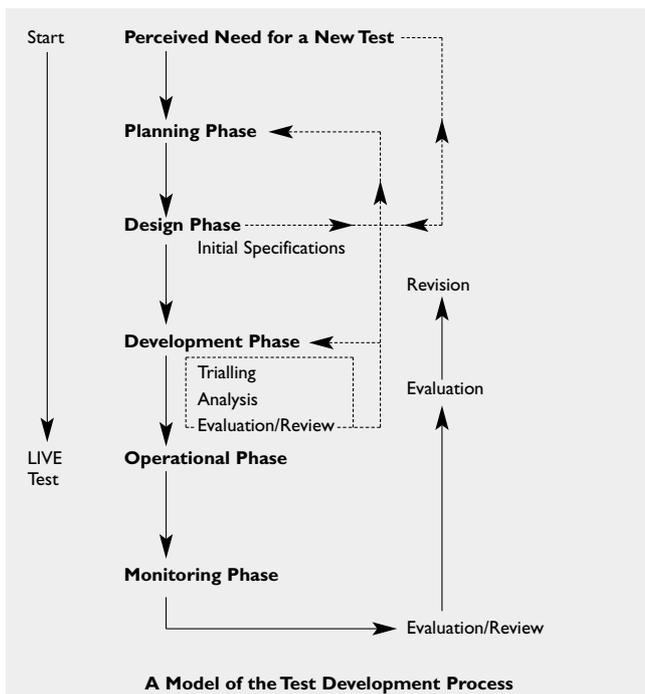


Figure 1

The cyclical nature of the development process

Figure 1 emphasises the cyclical and iterative nature of the test development process. The perceived need at the beginning of the process derives from an appraisal of the intended context of use and the relationship of the proposed new examination to any existing assessment procedures. If a new examination is produced, the monitoring and evaluation process may lead to revisions being made to it at a later stage or to another examination being developed.

A key aspect of the model is that validation is an integral part of the process. In order for validation to occur, the procedures which are implemented for the on-going production and administration of the examination should be designed so that adequate data can be collected. Failure to capture adequate data means that evidence of standards being reached and maintained (e.g. regarding validity, reliability, impact and practicality) cannot be provided.

From Monitoring and Evaluation to the Test Revision Process

Within the cyclical approach outlined above, a formal review (possibly leading into the revision process) is normally set in train by UCLES EFL several years after the introduction of the new or revised exam.

The review/revision process is initiated within a Project Review Group which meets on a regular basis within EFL. This step may be taken because the routine data collection and monitoring within the operational cycle has revealed issues which need to be addressed, or it may be the case that an appropriate moment in the cycle has been reached at which to institute a review.

In terms of internal procedures within UCLES EFL, the first stage is to set up the necessary management structures to oversee the review or revision project. As a first stage a Steering Group chaired by the Director or Deputy Director EFL is established to oversee the whole process and to manage resource allocation. A main concern is the allocation of the appropriate level of staff time for co-ordination of the project and participation in Working Groups.

The Steering Group establishes one or more Working Groups with a designated co-ordinator or co-ordinators and sets out the required terms of reference and expected membership. The membership of the Working Group includes members of the relevant Product Group as well as representatives from Validation and Business Support.

The Planning Phase

The first task of the Working Group will be to establish a Project Plan based on the Test Development model; this starts with a form of Situational Analysis as part of an initial review and will determine a project timeline with an anticipated end point (e.g. leading to a 'live' administration of the revised examination).

The Situational Analysis begins with a review of existing validation evidence relating to qualities of validity, reliability, impact and practicality (VRIP) for the existing exam. At this stage, as a result of this VRIP review, further instrumental research may need to be carried out.

In terms of impact, (especially views and attitudes) it is necessary to conduct a survey of the existing stakeholders who are involved in the examination. These include amongst others:

- Local Secretaries who administer the exam
- the language schools preparing candidates
- the receiving institutions and other users of the examination results
- the senior consultants and other professionals who are employed to work on the materials and assessment procedures (Senior Team Leaders, Team Leaders, Chairs of item writing teams, Principal Examiners etc).

While the current level of satisfaction with the examination can be determined by the regular, routine feedback which is collected from the stakeholders, it may be necessary to address specific issues and to seek views on possible directions for change before proceeding with the plan, e.g. with specially conducted surveys and other forms of data collection (cf. Nick Saville's article on Impact in *Research Notes 2*).

With regard to practicality the existing resourcing of the exam and its viability in terms of costs and revenue need to be reviewed. In order for UCLES EFL to guarantee a high quality of service, financial viability is a key factor; the resources necessary for the on-going review and development of an exam need to be taken into account within the cost structure for the development plan and the operational maintenance of the exam.

Forward planning for a review/revision is necessary within the internal budgeting cycles; suitable provision needs to be made for resources (human and financial) to carry out the necessary stages over a period (probably 3 years minimum if the process continues after the initial review). The Steering Group reviews the Project Plan and ensures that the predicted resource allocation can be provided. Procedures for adjusting the scope of the project may involve other Steering Groups getting involved before a decision to proceed can be given. When the initial Planning Phase is complete and the initial survey and findings have been reported by the Working Group to the Steering Group, the Design and Development Phases can begin as for a new test development project.

The predicted length of these phases may be determined by external constraints, such as agreements with other stakeholders, but the normal target will be to issue the public specifications for a revised examination 2 years prior to the intended release date. The end of the development phase is marked by the release of a revised handbook outlining the changes including sample materials and assessment criteria and this is normally followed by a programme of seminars and other informational events during the 12-18 months prior to the new test being administered.

The Design and Development Phases

Within the Design/Development phases (which last 1-3 years) the Working Group oversees the development of the revised specifications and sample materials. The validation activities, which include trialling/pretesting and other instrumental research, are incorporated within the work plans of the Validation Group and other relevant groups such as the Pretesting Unit.

Checklists are completed based on the work carried out in order to ensure that satisfactory standards are met in line with the established principles of good practice. These cover aspects of validity, reliability, impact and practicality; VRIP checklists were used, for example, in recent work carried out for the CPE Revision. As far as possible each aspect of the project is documented and written up with a view to future publication and public presentation in working papers, newsletters and at international conferences (cf. new SILTS volume on the CPE revision to be published early 2002).

The justification for changes and the proposed new formats or assessment procedures are normally scrutinised by both internal and external experts. In addition to the involvement of key staff within the organisation itself, the Working Group identifies suitable external academics or specialists in the field who can be involved in reviewing the changes against a background of relevant theoretical considerations. This may involve the external experts in seminars and other well-focused events or possibly in the compilation and editing of working papers or book length volumes documenting the revision process and outcomes.

During the Design/Development phases it is important to keep key stakeholders informed of the proposed developments. This is achieved by the following:

- regular meeting with publishers
- seminars within key countries for teachers (e.g. when specimen materials are fully developed but before final decisions are made)
- newsletter/bulletins – both general and specific
- reports in *Research Notes*
- consultation with Development Managers in countries and with other key external staff (STLs/TLs/OEs, Chairs, Chief Examiners).

The development of Speaking and Writing components involves specific considerations which relate to both technical and administrative issues (e.g. new rating scale development, changes to training and marking procedures etc).

For all aspects of the proposed revision the Working Group ensures that appropriate representation of the administration and business support staff is included in the project management structure. This may require a number of sub-groups to be established with specific marketing or administrative functions.

Preparing for the Operational Phase

The Operational phase overlaps with the Design/Development phases. When the internal specifications have been ratified by the Steering Group, the standard procedures for test production can begin. This may occur before the public specifications are released.

When the public specifications are released a range of activities needs to be carried out to ensure that all stakeholders are aware of the changes and any major implications for them. Forward planning takes into account attendance at conferences and events, such as seminars, or training sessions, in the key countries where the exam is used.

It may be the case that the development of some aspects of the administrative procedures will continue after the release of the public specifications. This may include refinement of the rating scales and other material for use of examiners and the procedures and materials for training them (videos etc.). Careful consideration is given to practical issues such as the re-training of examiners and the release of new materials.

In Summer 2001, the revised version of the IELTS Speaking Test will be introduced. In their article on this revision project, (included in this issue on page 9) Lynda Taylor and Neil Jones describe the process which was used in the project based on the model in Figure 1. They describe the project in terms of five phases where the Development Phase is broken down into three. This allows them to focus in particular on the validation of the assessment criteria and rating scales and on the re-training of the oral examiners.

In 2002, three revision projects will see revised examinations move into the Operational Phase with live administrations for the first time. These are:

- the BEC exams in March 2002
- the Certificates in Language Skills (CELS) in May/ June 2002 (see article in *Research Notes* 3)
- the new CPE in December 2002

References

Bachman, L F (1990): *Fundamental considerations in language testing*, Oxford: OUP

Bachman, L F and Palmer, A (1996): *Language testing in practice*, Oxford: OUP

Revising the IELTS Speaking Test

Lynda Taylor, Performance Testing Co-ordinator, UCLES

Neil Jones, Research Co-ordinator, UCLES

Issue 3 of *Research Notes* (November 2000) included an article on approaches to rating scale revision and described some of the methods UCLES EFL adopts when revising assessment criteria and rating scales. This follow-up article reports in detail on some of the developments to revise the IELTS Speaking Test.

Background to IELTS

The International English Language Testing System (IELTS) is designed to assess the language ability of candidates who need to study or work where English is used as the language of communication. Today IELTS is widely recognised as a language requirement for courses in further and higher education and is taken annually by more than 100,000 candidates at 251 approved British Council and IDP Education Australia centres in over 105 countries worldwide. (Full details of the test can be found in the IELTS Handbook or via the IELTS website at www.ielts.org.)

IELTS tests ability in all 4 skills – Reading, Writing, Listening and Speaking. The IELTS Reading, Writing and Listening Modules were most recently revised in 1995 (see Charge and Taylor, 1997, for a summary), but the Speaking Module, last revised during the Revision of the English Language Testing System (ELTS) in 1988/9, remained unchanged. The current Speaking Test lasts 10-15 minutes and consists of an oral interview between the candidate and a single examiner. Although revision of the Speaking Test was due in 1995, at that time it was still not clear what direction to take: a number of speaking-related research studies were in progress and the findings were awaited; and, as outlined in the previous article in *Research Notes* 3, the revision of any face-to-face speaking test is an especially complex matter, requiring careful management and considerable resources.

Revising the Speaking Test

The revision project began in early 1998 with identification of the issues which needed to be addressed. This was informed from a number of sources: a review of the routinely collected candidate score and test performance data for the operational IELTS speaking test; a review of theoretical and empirical studies on the test conducted between 1992 and

1998 (e.g. Ingram and Wylie, 1993; Brown and Hill, 1998; Merrylees and McDowell, 1999); a review of other research into speaking assessment, together with work on speaking test design for the other Cambridge EFL tests (see Lazaraton, in press/2001).

The revision project set out to revise (among other aspects of the test) the assessment criteria and rating scale to ensure that the descriptors matched more closely the output of candidates in relation to specified tasks and that raters could apply them in a more standardised manner. Attention focused on several areas:

- the salient/non-salient features of spoken performance for assessment purposes
- the nature of the scale/s (holistic or analytical? which criteria, and how many?)
- the behaviour of raters

Phase 1: Consultation, initial planning and design (May-Dec 1998)

Some initial investigative work was commissioned in June 1998 from specialists in the field of oral proficiency assessment: Alan Tonkyn of Reading University, UK, reported on his own study of grammatical, lexical and fluency features of IELTS candidates' oral proficiency at different bands, including the rater perspective (Tonkyn, 1998); Anne Lazaraton (then at George Mason University, USA) reported on discourse features observed during her transcription analysis of 20 IELTS speaking tests, ranging from Bands 3 to 7 (Lazaraton, 1998). This work, along with findings from earlier studies, raised the question of how well the existing holistic IELTS rating scale and its descriptors were able to articulate key features of performance at different levels or bands. It was felt that a clearer specification of performance features at different proficiency levels might enhance standardisation of assessment. For this reason, the Revision Project Working Party, which included specialists in speaking assessment and also active IELTS raters, reviewed the test specifications and rating scale descriptors to abstract the key analytical criteria and develop working definitions; they then deconstructed the existing holistic scale into several analytical subscales for more detailed investigation, deciding finally on four subscales: pronunciation, fluency and coherence, grammatical range and accuracy, and lexical resource.

Phase 2: Development (Jan-Sept 1999)

In May 1999 the Draft 1 assessment criteria and four rating subscales were applied to a subset of 4 audio-recorded test performances gathered in Australia when trialling the prototype of the revised test format. When applying the draft descriptors, careful attention was paid to features of candidate performance which distinguished the critical boundaries of Band 5/6 and Band 6/7; this exercise led to production of second draft assessment criteria and rating subscales. Draft 2 was trialled in July 1999 using a new set of 4 audio-recordings and further minor adjustments were made to some rating scale descriptors. At this point the Draft 3 assessment criteria and rating scale descriptors were considered ready for larger-scale trialling.

Phase 3: Validation (Oct 1999-Sept 2000)

This phase focused on setting up an experimental study to investigate the assessment criteria and scale functioning. The research design involved gathering a sample of video performances using the revised IELTS test format and then arranging for these to be multiply rated by experienced IELTS examiners. The video-rating option was preferred on the grounds that examiners rating audio-performances are inclined to under-rate; multiple rating of live performances was clearly not feasible.

A total of 29 video performances were filmed in the UK and Australia, using a range of materials (revised test format), proficiency levels, first languages and IELTS examiners. From these a dataset of 20 performances was selected for the multiple rating exercise. The subjects included 10 male and 10 female candidates, represented 15 different L1s, and ranged in level from Band 3 to Band 8 on the IELTS scale. (Candidates scoring below Band 3 or above Band 8 rarely appear in the live test so were not included in this study.)

The 20 performances were rated under controlled, on-site conditions by 2 teams of experienced IELTS examiners – a team of 4 in the UK and a team of 5 in Australia. They used the Draft 3 assessment criteria and rating subscales and provided independent ratings; any group discussion which took place immediately after rating was audio-recorded for future reference. Data for analysis therefore included both score data for the 20 candidates and retrospective feedback from the raters themselves.

Lynch and McNamara (1998) have put forward the view that generalizability theory and many-faceted Rasch measurement offer complementary approaches which both provide useful information for developers of performance assessments. With this in mind, candidate score data was analysed using the GENOVA and FACETS analysis programs to investigate several research questions. Some of the findings from these analyses are reported below.

Do the 4 subscales measure distinct aspects of speaking ability?

Pronunciation is clearly a distinct trait, correlating most weakly with the other scales. The three other scales appear to be more closely related, with the grammatical and lexical scales being the closest, as might be expected (see below):

	Fluency	Lexical resource	Grammatical range and accuracy	Pronunciation
Fluency	1			
Lexical resource	0.974	1		
Grammatical range and accuracy	0.951	0.981	1	
Pronunciation	0.829	0.865	0.843	1

Do the 4 subscales contribute consistently to the candidate's final score?

The FACETS graphical plot indicated that the band thresholds are fairly similarly spread in each of the 4 subscales. There is a small difference in the difficulty of the subscales, i.e. it seems slightly harder to score more highly on grammatical range and accuracy. Interestingly, this is consistent with findings from post-test analyses of other Cambridge speaking tests.

Do raters use and interpret the subscales in the same way?

Generally it appears that they do. This was confirmed by the relatively good generalizability coefficient established for the single-rater condition (see below). The FACETS analysis was useful for identifying cases of misfit which, if investigated, might throw light on problems in interpreting the markscheme as it is.

How reliable will the rating procedure be when applied by a single rater in operational conditions?

Operational requirements mean that the IELTS speaking test has to adopt a one-to-one format, i.e. one candidate and a single examiner. The reliability of the test as rated by 9 raters is predictably high but it is important to be able to model what could be expected in a single rater condition. The dependability (or Phi) coefficient is relevant to criterion-referenced interpretations and for a single rater the modelled reliability is still reasonable.

No of raters	G coefficient	Phi coefficient
9	0.956	0.952
1	0.862	0.857

Since some of the subscales (grammatical range and accuracy and lexical resource) appeared not to be strongly distinct in what they measure, one further question was considered: What if fewer subscales were used? Interestingly, although the Phi coefficient drops substantially from 0.857 with 4 subscales to 0.731 with 1 subscale, reducing the number of subscales from 4 to 3 would lead to only a small reduction in generalizability.

No of raters	No of subscales	G coefficient	Phi coefficient
1	4	0.862	0.857
1	3	0.847	0.841
1	2	0.819	0.810
1	1	0.746	0.731

Phase 4: Implementation (Oct 2000 – June 2001)

The above findings have directly influenced decisions on the nature and number of rating criteria and subscales which will be used with the revised IELTS Speaking Test. Further analyses have also been undertaken to compare the revised and existing approaches to rating.

In addition, the retrospective, qualitative feedback from raters in the early stages of trialling and during the multiple rating study has proved invaluable in helping to inform production of IELTS examiner training and standardisation materials. It has also led to further studies in the UK and Australia to investigate examiners' experience as they simultaneously deliver the test and rate candidate performance in real time.

The worldwide retraining of all IELTS raters, based on a cascade system of regionally-based, face-to-face training sessions, is scheduled to take place from January to June 2001. It is hoped that this comprehensive examiner re-training programme, which is crucial to the successful introduction of the revised IELTS Speaking Test, will be the subject of a future article in *Research Notes*.

Phase 5: Operational (from July 2001)

The revised assessment criteria and rating scales will become fully operational from July 2001 when the revised format of the IELTS Speaking Test is introduced worldwide. Following its introduction, candidate score and test performance data will continue to be systematically gathered in order to monitor the functioning of the individual subscales as well as examiner behaviour.

Conclusion

Over the past decade there have been numerous calls for a more empirically-based approach to developing assessment criteria and rating scale descriptors (Shohamy, 1990; Upshur and Turner, 1995; Milanovic, Saville, Pollitt and Cook, 1996). The project to revise the IELTS Speaking Test has provided a valuable opportunity to combine insights from expert judgement with findings from analyses of actual candidate performance and test scores; this makes it possible to redevelop criteria and rating scales which possess the essential qualities of theoretical relevance, discriminating power and assessability.

References and further reading

- Brown, A and Hill, K (1998): Interviewer style and candidate performance in the IELTS oral interview, in *IELTS Research Reports – Volume 1*, Elicos Association/IELTS Australia Pty Ltd
- Charge, N and Taylor, L B (1997): Recent developments in IELTS, *English Language Teaching Journal*, Volume 51/4, pp 374-380
- Crick, J F and Brennan, R L (1984): *GENOVA: a general purpose analysis of variance system*, Version 2.2, Iowa City, IA: American College Testing Program

External requests for access to UCLES data/materials

- Ingram, D E and Wylie, E (1993): Assessing speaking proficiency in the International English Language Testing System, in D Douglas and C Chappelle (eds) *A New Decade of Language Testing*, TESOL, Inc
- Lazaraton, A (1998): *An analysis of differences in linguistic features of candidates at different levels of the IELTS Speaking Test*, Internal report prepared for UCLES
- Lazaraton, A (in press/2001): *A qualitative approach to the validation of oral language tests*, Studies in Language Testing 14, UCLES/CUP
- Lynch, B K and McNamara, T F (1998): Using G-theory and many-facet Rasch measurement in the development of performance assessments of the ESL speaking skills of immigrants, *Language Testing*, Volume 15, pp 158-80
- Linacre, J M and Wright, B D (1993): *A user's guide to FACETS: Rasch-measurement computer program*, Version 2 (62) Chicago, IL: MESA Press
- Merrylees, B and McDowell, C (1999): An investigation of speaking test reliability with particular reference to examiner attitude to the speaking test format and candidate/examiner discourse produced, *IELTS Research Reports – Volume 2*, IELTS Australia Pty Ltd
- Milanovic, M, Saville, N, Pollitt, A and Cook, A (1996): Developing rating scales for CASE: theoretical concerns and analyses, in Cumming, A and Berwick, R (Eds) *Validation in Language Testing*, Multilingual Matters, Clevedon, pp 15-33
- Tonkyn, A (1999): *Reading University/UCLES IELTS Rating Research Project – Interim Report*
- Shohamy, E (1990): Discourse analysis in language testing, *Annual Review of Applied Linguistics*, Volume 11, pp 115-128
- Upshur, J A and Turner, C E (1995): Constructing rating scales for second language tests, *English Language Teaching Journal*, Volume 49/1, pp 3-12

Lynda Taylor, Performance Testing, Co-ordinator, UCLES

EFL Validation regularly receives enquiries from external students and researchers working on projects in the field of language testing. Sometimes these are requests for information about general testing and assessment issues and we are usually able to provide the UCLES perspective and suggest helpful references in the general language testing literature. Other enquiries are about specific UCLES tests; in such cases we normally direct enquirers in the first instance to the wide range of information now available through the UCLES EFL and ALTE websites (www.cambridge-efl.org and www.alte.org). If hard copies of documentation are needed, it is possible to download examination handbooks and other materials directly from both websites, or they can be requested from EFL Information (harding.a@ucles.org.uk) and from the ALTE Secretariat (alte@ucles.org.uk).

We frequently receive enquiries from external researchers requesting access to various types of information which are not normally within the public domain (e.g. test materials, score data, examination scripts). We are naturally keen to develop a mutually beneficial relationship with members of the external research community but we do need to give requests of this nature careful consideration and there are a number of constraints which apply. For example, granting external researchers access to candidates' test scores or examination scripts raises important issues of confidentiality and security; it would be unethical of us to release this information (which is the property of candidates and receiving institutions) other than in very special and controlled circumstances.

Following receipt of a request for sensitive or confidential data/material, we usually invite the enquirer to send us a detailed protocol of their proposed study; the protocol should outline the theoretical context for the chosen area of study, the research questions, the specific requirements for data/material from UCLES, the proposed methodology and analysis, the anticipated outcomes, etc. In the case of research students working on an MA or PhD project, we now ask for a letter from their supervisor in support of their request. After reviewing the details of the study provided by the researcher, we sometimes invite them to visit us in Cambridge to discuss their request and any issues arising in more depth; this can be helpful in explaining the UCLES perspective, in identifying common interests and in clarifying any points of difficulty. This process, which can take several weeks, leads to a final decision on whether we are able to support an external researcher in their study and, if so, what form that support should

Studies in Language Testing

take. A formal agreement is then drawn up between UCLES EFL and the external researcher clearly specifying the restrictions on access to and use of any data/material provided by UCLES; the agreement also affirms UCLES' right to see and comment on all papers/reports before publication.

Requests may be turned down for a variety of reasons: we may not actually have the data in the form required by the researcher (e.g. certain types of candidate background information); we may not have the internal resources needed to identify and assemble the requested dataset (e.g. a particular sample of writing scripts); or we may not consider the proposed study to be sufficiently well-designed to investigate the issue of interest, especially if it is a particularly complex or sensitive one (e.g. bias analyses).

Although MA and PhD students often write to ask if we can provide them with the performance and/or score data they need for their investigation and analysis, unfortunately this proves possible in relatively few cases. It should be remembered, however, that as a 'researcher in training' there is great value in having to plan and manage the data collection stages of a study yourself; it not only allows you the chance to construct a balanced sample with known characteristics according to your specific interests and needs, but it can also enable you to gather valuable additional background information on your subjects, via questionnaires, focus groups or verbal protocol analysis. Instead of providing score and performance data for analysis, we can sometimes provide specimen or retired test materials for student researchers to gather their own performance and/or score data.

Most of the cases in which we are able to provide support to external researchers are those where the interests of the external researcher overlap to some extent with the interests of EFL Validation. Last year, for example, we were able to provide two MA students at Reading University with video performances of the Cambridge Main Suite speaking tests so that they could analyse and describe the test-takers' language output. Their studies were not only of interest to the general language testing field in terms of the methodology used and the outcomes observed, but were also of special interest to UCLES EFL in the ongoing process of validating speaking test design in the Cambridge examinations.

Hopefully, this article will help to clarify the UCLES EFL position on releasing sensitive data/material in response to the requests we receive and will enable external researchers to know how best to approach the organisation with requests of this sort in the future.

Alan Davies has been centrally involved in Applied Linguistics and Language Testing for more than thirty years. Over that time he has also worked with UCLES on many occasions, most recently as a consultant, advisor and editor-in-chief of the SILT volume 6, *A dictionary of language testing*.

Volume 11 in the Studies in Language Testing Series, *Experimenting with uncertainty – Essays in honour of Alan Davies*, edited by Catherine Elder, brings together 28 invited papers surveying the state of the art in language testing from a perspective which combines technical and broader applied linguistics insights. The papers, by key figures from both within and outside the field of language testing, cover issues ranging from construct definition to the design and applications of language tests, including their importance as a means of exploring larger issues in language teaching, language learning and language policy.

The volume locates work in language testing in a context of social, political and ethical issues at a time when testing is increasingly expected to be publically accountable. It is thus particularly appropriate as a tribute to Alan Davies, whose work in this field since the 1960s has been marked by its conceptual strength and social responsiveness, seeking constantly to clarify and challenge current practice and new trends. The book represents an innovative and lasting contribution to the literature in the field. Alan's contributions to our work in Cambridge have always been of the greatest help and it is with a sense of honour that we publish this volume as a token of our respect.

The contributors are:

J Charles Alderson and Jayanti Banerjee, Lyle Bachman, Rosemary Baker, Geoff Brindley, Christopher Brumfit, Caroline Clapham, Dan Douglas, Rod Ellis, Patrick Griffin, Liz Hamp-Lyons, Batia Laufer, Brian K Lynch, John C Maher, Tim McNamara, Rosamond Mitchell, Helen Moore, Pauline Rea-Dickins, John Read, Daniel Reed and Andrew Cohen, Larry Selinker and ZhaoHong Han, Antonella Sorace and Daniel Robertson, Bernard Spolsky, Charles Stansfield and Joan Aucter, Elaine Tarone, Carolyn Turner, Cyril Weir, Henry Widdowson, Eddie Williams.

Other titles in the Studies in Language Testing Series are available from bookshops, or Cambridge University Press.

- 1 Lyle F Bachman, F Davidson, K Ryan, I-C Choi *An investigation in the comparability of two tests of English as a foreign language: The Cambridge – TOEFL comparability study*, Cambridge, 1995 (ISBN 0-521-48467-7)
- 2 Antony John Kunnan *Test taker characteristics and performance: A structural modelling approach*, Cambridge, 1995 (ISBN 0-521-48466-9)
- 3 Michael Milanovic, Nick Saville *Performance Testing, Cognition and Assessment: Selected papers from the 15th Language Testing Research Colloquium*, Cambridge and Arnhem, Cambridge, 1996 (ISBN 0-521-484465-0)
- 4 Caroline M Clapham *The development of IELTS: A study of the effect of background knowledge on reading comprehension*, Cambridge, 1996 (ISBN 0-521-56708-4)
- 5 Alison Green *Verbal protocol analysis in language testing research: A handbook*, Cambridge, 1998 (ISBN 0-521-58635-6)
- 6 *Multilingual glossary of language testing terms*, Cambridge, 1998 (ISBN 0-521-65877-2)
- 7 Alan Davies, A Brown, C Elder, K Hill, T Lumley, T McNamara *Language testing dictionary*, Cambridge, 1999 (ISBN 0-521-658764)
- 8 James E Purpura *Learner strategy use and performance on language tests*, Cambridge, 1999 (ISBN 0-521-658748)
- 9 Antony John Kunnan *Fairness and validation in language assessment*, Cambridge, 2000 (ISBN 0-521-658748)
- 10 Micheline Chalhoub-Deville *Issues in computer-adaptive testing of reading proficiency*, Cambridge, 2000, (ISBN 0-521-653800)
- 11 Catherine Elder (ed) *Experimenting with uncertainty* (ISBN 0-521-7725560) (in press)
- 12 Cyril Weir, Yang Huizhong, Jin Yan *An empirical investigation of the componentiality of L2 reading in English for academic purposes* Cambridge 2000 (ISBN 0-521-652995)

Forthcoming titles:

- 13 Kieran O'Loughlin *An investigatory study of the equivalence of direct and semi-direct speaking tests*
- 14 Anne Lazaraton *A qualitative approach to the validation of oral language tests*

The UCLES EFL Seminar Programme

The EFL 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 information 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.

In November 2000 Dr Chris Kennedy of the Centre for English Language Studies, University of Birmingham, presented a session to EFL staff on the topic of English for Specific Purposes, with particular reference to English for Business. He reviewed attempts since the 1960s to define Business English through functional and genre analysis, and touched upon the issue of how far teaching and testing can mirror the reality of language as it is used in the workplace, especially in relation to the developing concept of 'international English'. Dr Kennedy has agreed to be the editor of a *Studies in Language Testing* volume to be published in 2002; the volume will chronicle the development of the Business English Certificates (BEC) and will provide a useful statement of the UCLES position on various issues relating to the assessment of English for Business Purposes.

In December 2000 Professor David Crystal visited UCLES to give a presentation on the topic of English as a Global Language; this subject is currently attracting much attention following the publication of various books and articles (Crystal, 1997; McArthur, 1998; Graddol and Meinhof, 1999) and it has recently provoked considerable debate within the academic community (see *Applied Linguistics*, 20/2 and 21/3). In his presentation Professor Crystal briefly reviewed the way the English language has changed over recent centuries and has grown in influence to achieve the status of a world language, spoken by more than a billion people. He went on to consider the status and role of the different varieties of English which have already emerged and are continuing to emerge within the global community of English users. Finally, he touched upon the various implications this phenomenon is likely to have for English language teaching in the 21st century, and, by extension, for the testing of English. In a future issue of *Research Notes*, we plan to discuss in more detail the UCLES EFL position in relation to the notions of 'international' English and 'varieties' of English.

Contact of this sort between our internal staff and external specialists is invaluable in ensuring that the issues facing UCLES EFL (and language testers in general) are more widely known and understood within the academic community; in addition, UCLES EFL benefits greatly from the input of experts in critical fields as it seeks to formulate policy and implement good practice in relation to its language tests.

Announcement of the winners of the IELTS MA Dissertation Award 2000

As part of the tenth anniversary of IELTS in 1999, the IELTS partners – UCLES, The British Council, and IDP Education Australia – agreed to sponsor an annual award for the MA dissertation in English which makes the most significant contribution to the field of language testing.

For the first award in 2000, submissions were accepted for dissertations completed in 1999. The IELTS Research Committee, which comprises members from the three partner organisations, met in November 2000 to review the shortlisted submissions; after careful consideration, the Committee decided to announce 2 winners: Lindsay Brooks and Sally O'Hagan. The Committee was impressed at the extremely high quality of both dissertations and for this reason chose to make a double award in the inaugural year. The two award winners will each receive a cheque for £1000 to be presented at a public ceremony during the Language Testing Research Colloquium in St Louis, USA, in February 2001.

The two researchers chose to address topics of great importance to all universities catering for a linguistically and culturally diverse international student population.

Lindsay Brooks explored the attitudes of adult ESL students towards performance-based assessment (portfolio, presentation and participation) as compared to more traditional, formal tests. Sally O'Hagan investigated the attitudes of lecturers to academic writing produced by native and non-native speaker students as well as their actual rating behaviour when marking coursework essays, often without reference to pre-established assessment guidelines. Each researcher therefore chose to focus on a key language-related assessment issue which arises in the context of mainstream education and which touches upon matters of fairness and equity in relation to different stakeholders in the assessment process.

The literature reviews in both dissertations were considered to be thorough and well-argued. The two research designs were well-planned and executed, adopting sound methodologies which combined both quantitative and qualitative approaches. In both studies the findings were convincingly and lucidly presented, and any constraints on data collection and analysis were suitably acknowledged. In summary, the IELTS Research Committee considered the two studies to be excellent examples of applied linguistics research: each one systematically investigated a real-world problem observed in the course of the researcher's own experience, and

each one reported the findings in terms that are intelligible to those most likely to profit from the insights gained. The Committee agreed that these were highly original and valuable pieces of work which contribute to an increasing understanding of the attitudes of both students and lecturers in relation to performance-based assessment. The abstracts from the award-winning dissertations are presented below.

Lindsay Brooks: Adult ESL attitudes towards performance-based assessment

Ontario Institute for Studies in Education at the University of Toronto, Canada. (Academic supervisor: Professor Merrill Swain)

Performance-based assessment, such as portfolios, presentation and participation, is currently being used in many second language programs. A review of the literature prior to this research revealed that although there have been many studies and papers on performance-based assessment, including alternative assessment and authentic assessment, few have reported student reactions to this wave of assessment techniques. Therefore, the attitudes of adult English as a second language (ESL) students to performance-based assessment (portfolios, presentations, and participation) versus more traditional types of tests were surveyed by means of a questionnaire and semi-structured interviews. Quantitative and qualitative results suggest that the participants in the study perceived all four types of assessment positively. Analysis of background variables suggests that there were interaction effects for level of language proficiency and home country with regard to attitude toward assessment type. Other biographical variables showed little or no relationship to attitudes.

Sally O'Hagan Assessment of student essays: methods of marking work written by students from non-English speaking backgrounds

Department of Linguistics and Applied Linguistics, University of Melbourne, Australia. (Academic supervisor: Professor Cathie Elder)

Together with the large number of Australian students who use one or more languages other than English, overseas students from non-English speaking backgrounds (NESB) contribute to a growing linguistic diversity in many Australian universities. This thesis investigates how teachers of these students respond to this diversity by examining the ways in which staff evaluate the work of NESB students, and asks whether there are any

IELTS MA dissertation award

differences in the way staff respond to writing by native-speaking and NESB students. Two means of data collection were used to survey the essay marking methods used by staff in three academic departments. Staff sampled from each department completed a written questionnaire which asked them to describe their marking procedures and assessment criteria, and to identify any problems they experience specifically when marking essays by NESB students. One respondent from each department also provided verbal reports of their essay marking protocols, using an introspective, or 'think-aloud' methodology. The survey findings show that markers do not always respond to native-speaker and NESB essays in the same way, considering some assessment criteria to be less important for NESB essays, and commonly modifying their marking methods in some way for NESB essays. Many staff reported experiencing feelings of conflict over these differences in their responses, due to concerns over equity in the evaluation of student performance. Most staff believe that the introduction of marking guidelines could be of use in helping them to resolve this problem. In the context of a broad internationalisation agenda in higher education, which includes further international recruitment, there is a growing imperative for staff to be able to engage effectively and confidently with a diverse student body. On the basis of the survey findings, this thesis argues that further training and support is required to enable staff to do so.

To mark the tenth anniversary of IELTS in 1999, the three IELTS partners – the University of Cambridge Local Examinations Syndicate (UCLES), The British Council and IDP Education Australia: IELTS Australia – instituted an annual award of £1,000 for the MA dissertation in English which makes the most significant contribution to the field of language testing. In its inaugural year the award went to joint winners in Australia and Canada (see page 16 for full report). For this year, the entry procedures and timetable for the award are as follows:

Submission and evaluation procedures

A 1000-word synopsis of the dissertation together with reference from your supervisor should be submitted to:

Dr Lynda Taylor
EFL Division
University of Cambridge Local Examinations Syndicate
1 Hills Road
Cambridge CB1 2EU
United Kingdom

- The IELTS Research Committee, which comprises members of the three partner organisations, will review the submissions and shortlist potential award winners.
- For all shortlisted dissertations a full copy of the dissertation will be requested and a further reference may be sought.
- The Committee's decision will be final.

Time table

The following timetable will apply in 2001:

1 July	Deadline for submission of synopses and references to UCLES
1 September	Deadline for submission of full copies of shortlisted dissertations (and further reference)
October	Meeting of IELTS Research Committee
November	Announcement of award

Details of the application process for the IELTS MA Dissertation Award 2001 can also be found on the IELTS website – www.ielts.org

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 553822

Fax: +44 1223 553068

e-mail: harding.a@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:

European Year of Languages Conference

European language testing in a global context – Barcelona 5-7 July 2001

To mark the European Year of Languages, this conference will provide an unprecedented opportunity to relate the development of European language assessment to the increasingly globalised context of education and employment. Speakers from over 20 countries will consider language assessment from a wide range of theoretical and practical perspectives.

Plenary speakers

Professor J Charles Alderson, Lancaster University, UK

Current and future trends in language testing

Professor Lyle Bachman, University of California Los Angeles, USA

The use of new technologies in language assessment

Dr Wolfgang Mackiewicz, Freie Universität Berlin, Germany

European language policy in the next decade

Dott. Ispettore Raffaele Sanzo, Ministero della Pubblica Istruzione, Italy

Foreign languages within the frame of Italian educational reform

Mr Joe Sheils, Modern Languages Division, DGIV, Council of Europe,

Strasbourg, France

The role of the Council of Europe in European language policy: the portfolio for young learners

Dr John Trim, Project Director for Modern Languages, Council of Europe, 1971-1997

The Common European Framework and its implications for language testing

Registration fee – 200 euros

For further information on making a presentation or participating in the conference, please contact:

ALTE Secretariat

1 Hills Road

Cambridge

CB1 2EU

UK