

ResearchNotes

Contents

| | |
|---|----|
| Editorial Notes | 1 |
| Rising to the Challenge of Asset Languages | 2 |
| Opening a new door for teachers of English: Cambridge ESOL Teaching Knowledge Test | 5 |
| Staying in Touch: tracking the career paths of CELTA graduates | 7 |
| Cambridge ESOL and the NRDC ESOL Effective Practice Project | 12 |
| Closer collaboration with other Cambridge University departments | 14 |
| ESOL Staff Seminar programme: Applied Linguistics: a personal view and Second Language Listening | 14 |
| Raising the Languages Ladder: constructing a new framework for accrediting foreign language skills | 15 |
| The Common Scale for Writing Project: implications for the comparison of IELTS band scores and Main Suite exam levels | 19 |
| Other news | 24 |

Editorial Notes

Welcome to issue 19 of *Research Notes*, our quarterly publication reporting on matters relating to research, test development and validation within Cambridge ESOL.

The theme of this issue is the development and validation of new assessment products and the opportunities for collaboration these bring. We report on products under development which take Cambridge ESOL for the first time into primary, secondary, tertiary and higher education sectors in the UK with multilingual tests within the *Asset Languages* project; into the Adult ESOL curriculum with *Skills for Life* tests, as well as reporting on the development of the *Teaching Knowledge Test*, which will assess teachers' professional knowledge about the teaching of English to speakers of other languages in a worldwide context. This issue also focuses on developing frameworks for describing and assessing all language skills, i.e. reading, listening, writing and speaking.

In the opening article Neil Jones, Karen Ashton and Ann Shih-yi Chen introduce Asset Languages, an assessment system being developed by UCLES to implement the Languages Ladder, a voluntary recognition system in the UK which seeks to give people credit for their language skills across 26 languages. Next, Mick Ashton and Hanan Khalifa outline the Teaching Knowledge Test (TKT), an award for teachers of English at any stage in their career which was developed in response to stakeholder needs.

Tony Green then reports on an impact study which is tracking the career paths of people who have taken our Certificate in English Language Teaching for Adults (CELTA).

James Simpson describes a collaborative project with researchers at Leeds and London Universities which is developing an instrument for assessing speaking of adult ESOL learners in the UK. This project has informed the development of our ESOL Skills for Life awards which will be available from April 2005. Continuing the theme of collaborative activities, Lynda Taylor outlines how Cambridge ESOL has forged closer links with other departments in the University of Cambridge which is followed by a review of former head of the University's Research Centre for English and Applied Linguistics Gillian Brown's recent staff seminar on applied linguistics and second language listening.

The final two articles deal with frameworks and comparability issues between language tests. Neil Jones describes how tests for 26 different languages are being related to the Languages Ladder framework, focusing on how objectively marked components (reading and listening) can be linked to it. Turning to writing, Roger Hawkey and Stuart Shaw draw implications from developing a common descriptive scale for assessing writing to comparing Main Suite, IELTS and BEC writing scripts and scores.

We finish with a range of news items, including updates on seminars for teachers and the forthcoming ALTE conference in Berlin together with new information available on our website such as Cambridge ESOL's latest annual review. Keep up to date with information about other new products on our website www.CambridgeESOL.org

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

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

Rising to the Challenge of Asset Languages

NEIL JONES, KAREN ASHTON AND ANN SHIH-YI CHEN, RESEARCH AND VALIDATION GROUP

Neil Jones, Dominique Slade, Karen Ashton and Ann Shih-yi Chen from the *Asset Languages* team in Cambridge ESOL presented aspects of this new product at a recent ESOL staff seminar. Guest speaker Pat McLagan then spoke about the Key Stage 2 Framework for Modern Foreign Languages (MFL hereafter). Pat is a former LEA Advisory Teacher for MFL, CILT Language Teaching Adviser, QCA MFL Subject Officer and OFSTED inspector for MFL and is currently working as a consultant for CILT.

This article is a summary of the seminar. After an introduction to *Asset Languages* it discusses the challenges of comparability both within the *Asset Languages* framework and in a wider UK and European context, the task development process, the nature of Teacher Assessment and the current state of MFL teaching in UK primary schools.

Introduction to *Asset Languages*

Asset Languages is the brand name of the assessment system being developed by UCLES to implement the Languages Ladder. It is a joint venture by two UCLES business streams: OCR and Cambridge ESOL. Cambridge ESOL has responsibility for developing the assessments themselves and conducting validation of the system.

UCLES was awarded the tender for *Asset Languages* in October 2003 by the Department for Education and Skills (DfES hereafter). The project originated from one of the three overarching objectives of the National Languages Strategy 'Languages for all: Languages for Life. A Strategy for England' (2002) which were:

- improving the teaching and learning of languages, including making the most of e-learning
- introducing a voluntary recognition system, i.e. *the Languages Ladder*, to complement existing qualification frameworks, and give people credit for their language skills
- increasing the number of people studying languages in further and higher education and in work-based training.

Key features

Key features of *Asset Languages* assessments include:

- **Modular** – *reading, listening, writing* and *speaking* skills are assessed separately.
- **Can-do approach** – assessments are based on a functional can-do approach.
- **External Assessment (EA)** and **Teacher Assessment (TA)** – two forms of assessment are available. Successful completion of EAs leads to formal qualifications whereas TAs lead to certification by an *Asset Languages* accredited teacher.

- **26+ languages** – UCLES is contracted to provide assessments in at least 26 languages. Test production is currently focusing on three languages – Spanish, French and German. Development work is also taking place for the next round of languages – Panjabi, Urdu, Japanese, Mandarin Chinese and Italian.
- **6 levels** – the framework will be constructed around six levels – *Breakthrough, Preliminary, Intermediate, Advanced, Proficiency* and *Mastery*. Development in 2004 has been on the first three levels of the ladder.
- **Paper based tests (PB)** and **Computer based tests (CB)** – teams are working on developing computer based tests as well as paper based versions.
- **Range of educational contexts** – assessments will be available for those in primary, secondary, adult, further education (FE) and higher education (HE) sectors in the UK.

Comparability issues

With the vast number of variables detailed above, issues of comparability pose a significant challenge for the *Asset Languages* team. Some aspects of comparability that will need to be taken into account include:

- **Linguistic differences:** Clearly languages differ in terms of linguistic features (e.g. the way they are written, or the way they express notions such as time).
- **Developmental differences:** How far can young children and adult learners be compared? Users of *Asset Languages* will include both foreign language learners and members of communities where the language is routinely spoken. The developmental route may be quite different in these two cases.
- **Social and cultural differences:** The social and cultural setting in which language learning takes place also impacts on methodological approaches and the attitudes of teachers.

Framework issues

In addition to the complexities of developing a framework which works internally for *Asset Languages*, developing a framework which allows for meaningful comparisons to be made across existing frameworks is vital given that 'the scheme is designed to complement existing national qualification frameworks and the Common European Framework of Reference (CEFR)' (http://www.dfes.gov.uk/languages/DSP_languagesladder.cfm).

Table 1 opposite illustrates how the DfES expects the frameworks to align.

Table 1: Alignment of Frameworks

| National Qualification Framework | National Curriculum levels | General qualifications | Asset stages | CEFR (approx) |
|----------------------------------|----------------------------|------------------------|------------------|---------------|
| Entry level | 1–3 | Entry 1–3 | Breakthrough 1–3 | A1 (A2) |
| Level 1 | 4–6 | Foundation GCSE | Preliminary 4–6 | A2 (B1) |
| Level 2 | 7–EP | Higher GCSE | Intermediate 7–9 | B1 |
| Level 3 | | AS/A/AEA | Advanced 10–12 | B2 |
| Levels 4–6 | | | Proficiency | C1 |
| Levels 7–8 | | | Mastery | C2 |

Asset Languages is a simple system to the extent that it sets out to recognise language skill, not learning effort. Different learners (of a foreign language, or a community language) will achieve an Asset level with differing degrees of effort. Certain skills – e.g. writing – are harder to acquire in some languages than others. *Asset* differs from current UK qualifications such as GCSEs and AS/A levels in that practically the latter have a dual purpose: they accredit academic achievement as well as the useful skill acquired. This means that although GCSEs and AS/A level exams aim to be ‘at the same level’ across languages, this probably does not imply direct comparability of levels of functional language ability.

Task types

The *Asset Languages* team within Cambridge ESOL has been given the challenge of developing materials that are appropriate for the MFL context in the UK. Existing ESOL task types have therefore been specially adapted by the development team. Table 2 below gives some examples of how existing Cambridge ESOL task types have been modified.

ESOL task types were modified for *Asset Languages* for various reasons. Some were modified due to the need for the same task type to apply across exams of the same level but for different contexts or to be of comparable difficulty to task types used in

Table 2: Development of Asset Languages task types

| Existing Cambridge ESOL tasks | | Asset Languages tasks | |
|-------------------------------|---|-----------------------|----------------------------------|
| Exam | Task type | Exam | Task type |
| YLE | Listen and colour and draw | Breakthrough | Listen and colour |
| KET | Matching (notices): 2 distractors | Preliminary | Matching (notices): 1 distractor |
| PET | Listening/Reading: 4 option multiple choice | Intermediate | 3 option multiple choice |

Key: YLE: Cambridge Young Learners English Tests PET: Preliminary English Test KET: Key English Test

exams of the same level across different contexts. Some changes to task types were prompted by recommendations from the DfES or the QCA (the UK awarding body for national qualifications). In the initial stages of development the focus had been on developing task types for EA. Due to the different demands of TA (see the section below), the development team are in the process of developing new task types which are suitable for this mode of assessment.

Another key feature of *Asset Languages* tasks is that the rubrics and many of the questions are in English for the first three levels that have been developed. This is primarily to provide candidates with access to the tasks. As assessments for *Asset Languages* are modular it is important that candidates who are not learning the writing system for a non-Latin script language are not denied access to tasks in the listening paper, for example.

Developing sufficient quantities of materials for both EA and TA and keeping to the tight development timelines for the upcoming languages and levels is an ongoing challenge for the busy development team.

Teacher Assessment

As mentioned earlier, TA is carried out by an *Asset Languages* accredited teacher (a classroom language teacher who has passed the *Asset Languages* accredited teacher training) with pupils in the classroom. Three key characteristics have been identified for TA. These are *flexibility*, *ongoing process* and *formative feedback*.

Flexibility

As many primary schools currently allocate a maximum of one hour per week for MFL teaching, *Asset Languages* TA therefore needs to be flexible enough to complement current teaching practices. Due to limited class time, few schools in the UK use the European Language Portfolio (ELP) to record their students’ language achievements despite it being an effective method of record-keeping that does not require much teaching time. The problem seems to be, however, that the ELP itself is not assessment material. To help teachers certify their learners’ language competence and capture learners’ achievement, TA will provide assessment material that is integrated with classroom activities. This aspect of TA links to the second feature of TA, *ongoing process*.

Ongoing process

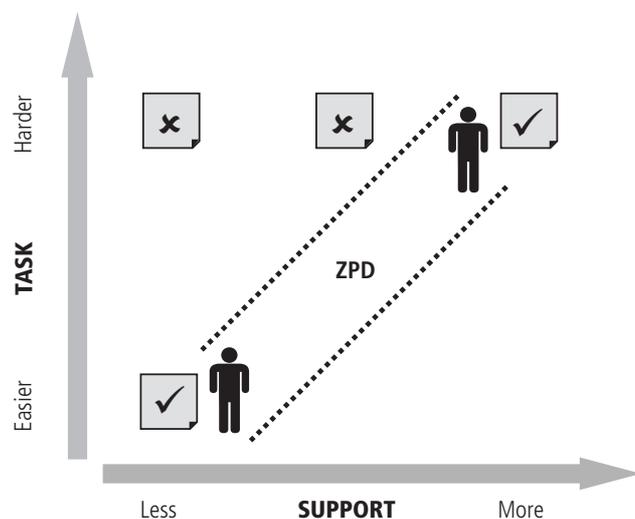
A challenge for *Asset Languages* is to develop TA materials in a way that supports teachers and the work they are currently doing in the classroom rather than imposing additional requirements on them. In order to achieve this objective Cambridge ESOL is developing a TA model which as well as providing teachers with ready-made materials will also allow teachers to develop their own assessment material using guidelines and templates developed by Cambridge ESOL. These guidelines and templates will demonstrate to teachers how they can make use of their current classroom activities to generate assessment opportunities. In other words, TA is to enhance the ongoing process of learning and teaching.

Formative feedback

Formative feedback is the third key characteristic of TA. Formative feedback in an *Asset Languages* context refers to the provision of feedback to students and feed-forward – that is giving them direction so they know what they should do in order to progress.

The model of formative feedback for TA is theory-driven and based on Vygotsky's notion of the Zone of Proximal Development (ZPD). This refers to the distance between what learners can do on their own and what they can achieve with assistance from more competent members of a group (Lantolf 2000). The application of Vygotsky's ZPD to TA is illustrated in Figure 1.

Figure 1: TA Zone of Proximal Development



Based on this model, a learner at a given point of development will be able to succeed with harder tasks with more support, or easier tasks with less or no support. In other words, the learner's ability is not viewed as a fixed point; instead, the learner's growth is developed through interaction between the learner and the

teacher. This understanding of progression remains theoretical. The impact of formative feedback on language development and the role it plays in the classroom are areas which will need monitoring and evaluating.

Key Stage 2 Framework for MFL

To complement the National Languages Strategy, the DfES introduced its Key Stage 2 framework for primary schools (7–11 year old students) to enable schools to implement language learning successfully. Its goals were stated as follows:

'Every child should have the opportunity throughout Key Stage 2 to study a foreign language and develop their interest in the culture of other nations. They should have access to high quality teaching and learning opportunities, making use of native speakers and e-learning. By age 11 they should have the opportunity to reach a recognized level of competence on the Common European Framework and for that achievement to be recognised through a national scheme. The Key Stage 2 language learning programme must ... be delivered at least in part in class time.'

(http://www.dfes.gov.uk/languages/DSP_whatson_primary.cfm)

This passage contains three specific goals: entitlement for quality teaching and learning, recognition of language proficiency and classroom assessment. The success for providing quality teaching and learning will be supported by EA and TA in *Asset Languages*. To ensure the quality of EA and TA, it is essential to maintain ongoing communication with teachers and to establish mechanisms for teachers to provide us with feedback. This mutual understanding is clearly vital to the success of developing *Asset Languages* in a way that both complements current classroom practices and is comparable to existing assessment modes currently used in UK schools.

The work being undertaken by the Cambridge ESOL and OCR *Asset Languages* team is full of challenges but is set to have great ramifications for language learning and teaching in the UK and beyond, through what it is teaching us about developing tests for new contexts.

References

- DFES (2004a) *Languages for all: From strategy to delivery*, document retrieved from <http://www.dfes.gov.uk/languages/uploads/Languages%20Booklet.pdf>
- (2004b) *The Language Ladder – steps for success*, web page, retrieved from http://www.dfes.gov.uk/languages/DSP_languagesladder.cfm
- Lantolf, J (2000) Introducing sociocultural theory, in Lantolf, J (Ed.), *Sociocultural theory and second language learning*, Oxford: Oxford University Press.

Opening a new door for teachers of English: Cambridge ESOL Teaching Knowledge Test

MICK ASHTON, EXAMINATIONS AND ASSESSMENT GROUP
HANAN KHALIFA, RESEARCH AND VALIDATION GROUP

Background

Cambridge ESOL has long been a major provider of high quality, internationally recognised awards for English language teachers. Many teachers have gained entrance to the ELT profession following successful completion of Cambridge CELTA courses, and thousands more have progressed to senior positions after passing the Cambridge DELTA.

In recent years there has been large-scale educational reform in many countries across the world. English is now being taught much earlier in the curriculum. Consequently, many more English language teachers are needed, and many teachers of other subjects now find themselves needing to teach English. The existing Cambridge ESOL teaching awards require lengthy and/or intensive preparation courses, which, for practical reasons, may not be so attractive to many of this new generation of English language teachers. The high level of English language proficiency required by the existing awards might also be a barrier for some. In order to fulfil our mission of 'providing language teachers in a wide variety of situations with access to high quality teaching awards, which will help them to achieve their life goals and have a positive impact on their learning and professional development experience', it became clear that Cambridge ESOL needed to develop an alternative framework of teaching awards, and that this should include products that can cater more closely to the needs of teachers of English in a wide range of contexts around the world.

The new In-service Certificate in English Language Teaching (ICELT) has replaced the Certificate for Overseas Teachers of English (COTE). This is now a more flexible, course-based award which teachers can follow in their place of work. Cambridge ESOL has also developed a completely new test called the *Teaching Knowledge Test* (TKT), which will be available in the spring of 2005, and is accessible for teachers who have a level of English of at least Level B1 of the Council of Europe's Common European Framework of Reference for Languages. TKT candidates are also expected to be familiar with language relating to the practice of ELT. A non-exhaustive list of teaching terminology is provided in the TKT Glossary, which can be found on our website www.CambridgeESOL.org/TKT

TKT can be taken at any stage in a teacher's career. It is suitable for pre-service or practising teachers and forms part of a framework of teaching awards offered by Cambridge ESOL. This includes CELTA (Certificate in English Language Teaching to Adults); CELTYL (Certificate in English Language Teaching to Young Learners); ICELT (International Certificate in English Language Teaching); DELTA (Diploma in English Language Teaching to Adults) and IDLTM (International Diploma in Language Teaching Management).

The Development of TKT

In late 2002, Cambridge ESOL sent out questionnaires to various teacher training institutions worldwide in order to elicit reactions to our proposal to develop a new test for teachers which would be quite different in format and concept from the existing Cambridge ESOL teaching awards. Considerable interest was expressed, which in turn led to a series of visits by Cambridge ESOL Development Managers to countries throughout Latin America, East Asia, the Middle East and Europe. Potential partner organisations have been identified and regular meetings have taken place both in Cambridge and overseas. This process of consultation continued throughout 2003 and the first half of 2004, and has enabled Cambridge ESOL to develop TKT in such a way that it will have relevance to teachers working in different educational sectors in a wide range of countries.

A Working Group, consisting of Cambridge ESOL staff and external consultants with considerable experience in teacher education and test development, was also established in 2003, and the group has met regularly to develop the TKT syllabus and produce materials. This has been an iterative process, with each version of the syllabus being sent out for review by teacher development professionals who have experience of working in the countries where interest in TKT has been expressed. Revisions to the syllabus were then made, and materials writers were commissioned to produce test items to cover the revised syllabus. At various points in the development cycle, test materials were trialled in key countries where test development staff from Cambridge were on hand to monitor and run focus groups with the participating teachers. Further revisions, consultation and trialling followed until the development team was convinced that the product met the needs of all interested parties, and that it would be possible to achieve dependable coverage of the syllabus areas in the construction of live test versions. Cambridge ESOL's well-established set of procedures for test development are fully described in Weir & Milanovic (2003).

Supporting documentation was then finalised, including the TKT Glossary of ELT terms, the TKT Handbook and sample materials for each module. Guidelines were also prepared for TKT item writers and item writer training has taken place on several occasions.

Description of TKT

TKT is a test of professional knowledge about the teaching of English to speakers of other languages. This knowledge includes concepts related to language and language use, and the background to and practice of language learning and teaching.

The tests are simple to administer, and each contains a range of tasks with a total of 80 objective-format questions.

The testing syllabus for TKT has theoretical, practical and management strands, and covers universal aspects of what a successful teacher of English needs to know. The syllabus areas also apply to Cambridge ESOL's other teaching awards such as ICALT, CELTA and DELTA, but at TKT level teachers do not need to demonstrate such a wide/deep understanding of these. It should be noted that, unlike the other Cambridge ESOL teaching awards, TKT does not include an assessment of a candidate's teaching ability in the classroom.

The broad syllabus areas have been grouped into three equally balanced modules, as shown in Table 1:

Table 1: Module contents in TKT

| Module | Title | Description |
|--------|--|---|
| 1 | Language and background to language learning and teaching | This module focuses on general knowledge of the subject matter to be taught and background to ways of teaching and learning. These aspects have been grouped together as they underpin the rest of the test. Their inclusion also makes explicit the methodology of the test. |
| 2 | Lesson planning and use of resources for language teaching | This module focuses on what occurs before specific classroom events – the planning and selection of lessons, resources and materials. |
| 3 | Managing the teaching and learning process | This module focuses on the classroom event itself – teacher and learner language, and the management of the classroom and the learning process. |

For each module, candidates are required to read and then answer questions by selecting the correct letter. Listening, speaking or extended writing are not required when taking TKT. Modules can be taken together in one examination session or separately, in any order, over three sessions. TKT modules are free-standing, and there is no aggregate score for candidates taking more than one module. There is no pass/fail in TKT, and candidates receive a certificate for each module that is taken.

TKT is designed to offer maximum flexibility and accessibility for candidates and therefore does not include a compulsory course component or compulsory teaching practice. However, it is likely that centres will wish to offer courses for TKT preparation and these may include some teaching practice, if desired. Evidence from various countries suggests that the TKT syllabus can easily be embedded into existing teacher training courses as well as providing a useful starting point for course design.

TKT candidates will also have their own portfolio. This is an electronic resource in which candidates keep a record of their teaching experience, beliefs and aspirations for the future. Through the portfolio candidates are encouraged to become reflective practitioners by analysing their teaching and how this impacts on their students' learning. The portfolio does not form part of the assessment for TKT, however.

Validation Studies

A number of validation studies have taken place since 2002 for TKT, described below.

Trialling

During the development phase of TKT, test materials were trialled over an eighteen-month period. Local Education Authorities, Ministry Departments, State and Private Universities, and British Council Institutes in several countries in Latin America, Asia and Europe participated in the trials.

The main trialling stage took place between May and July 2004, and attracted over 1,000 participants. The sample was representative of the target candidature for TKT, consisting of both in-service and pre-service teachers, working with different age groups and with a range of teaching experience. Several instruments were used during the trials. In addition to full versions of all three TKT modules, a language test was used to enable us to gauge the extent to which candidate performance on TKT might be affected by language proficiency. Questionnaires were also administered to key stakeholders and all participating teachers in order to gather feedback on the examination.

The major findings from this trial were as follows:

- The trialled versions achieved high reliability figures of 0.9 with a mean P (average facility) ranging from 0.72 to 0.81.
- Language proficiency does not appear to be an impeding factor. Candidates at CEFR A2 level scored 54%, 43% and 52% of the available total marks on Modules 1, 2 and 3 respectively.
- Age does not seem to be a factor affecting performance.

Stakeholder Perceptions

Feedback from the trialling has played an important role in the subsequent development of TKT. Positive feedback was received in terms of TKT content coverage, appropriacy, interest and relevance to local contexts. Potential candidates perceived sitting for an exam such as TKT to be a learning experience in itself. They welcomed the chance to reflect on their teaching practice and teaching knowledge.

Grading and Results

An exploratory standard-setting activity was conducted to inform the reporting of results and the grading stage of TKT. A rich assembly of ten judges with expertise in teacher training, rater-training, setting performance criteria, and language testing participated in the activity. Judges were asked to go through each module, answer each item and provide a rating on a four-point difficulty scale with 1 being the easiest and 4 being the most difficult. Convergence and divergence between the ratings were discussed and a rationale for divergence was provided. The activity proved to be very beneficial in further refining the candidate profile at each of the four bands which will be used for reporting performance on TKT. Before deciding on a score range for each band, a comparison was made between the judges' ratings and IRT item statistics available from the aforementioned trials.

TKT performance will be reported in four bands. Band 4 will reflect strong performance, whereas Band 1 candidates will demonstrate limited knowledge of the relevant syllabus areas. The trialling research indicates that for a candidate to achieve TKT Band 3, a score of at least 45–50 marks (out of 80) is required. It should be noted that the reporting of results for TKT is subject to ongoing research.

Ongoing Research

Following the extensive design, development and trialling phases, Cambridge ESOL continues to engage in a programme of research

and validation activities in relation to TKT. Future plans include further standard-setting activities and introspection studies. Such validation activities are required to ensure that satisfactory standards are met in line with the established principles of good testing practice, covering validity, reliability, impact and practicality.

Reference

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

Staying in Touch: tracking the career paths of CELTA graduates

TONY GREEN, RESEARCH AND VALIDATION GROUP

Introduction

Many readers of *Research Notes* will have memories of taking the Cambridge ESOL Certificate in English Language Teaching for Adults (CELTA) qualification or one of its forebears such as the RSA Cert TEFLA. You may have kept in touch with some of your fellow graduates as they progress along diverse career paths. As the awarding body, Cambridge ESOL routinely collects data from all CELTA course participants to learn about their backgrounds and their future intentions regarding employment. Until recently, however, we have not formally kept track of what actually happens to these participants after they leave their CELTA courses.

The impact of EFL examinations on candidates is an area that has attracted increasing attention from researchers over recent years, particularly questions of washback or the effects of an examination on the teaching and learning preceding it (Bailey 1999; Cheng & Watanabe 2003; Green 2003; Hawkey 2001). However, other aspects of test impact are of no less importance and are now beginning to attract concerted attention from researchers and testing organisations. In an era of accountability, career tracking is increasingly recognised as an important aspect of programme evaluation (Mason, Williams, Cranmer, & Guile 2003) and, in line with our wider concern with the impact of all our qualifications (Saville 2000), Cambridge ESOL is concerned to monitor the ongoing value of CELTA to those taking the course.

This article reports on a first step in addressing this area: a survey to track the careers of CELTA graduates around the world. It is hoped that reporting on data gathered to date will provide readers with a broader picture of the impact of initial teacher training in EFL on the subsequent career paths of course participants.

What is CELTA?

The CELTA is an initial qualification for people with little or no previous teaching experience. Administered by Cambridge ESOL since 1988, CELTA is the best known and most widely taken initial TESOL/TEFL qualification of its kind in the world. The course is offered at over 286 approved centres in 54 countries, providing almost 900 CELTA courses every year. The annual candidature is now over 10,000.

The course is intended to meet the needs of a variety of users as:

- an entry level qualification for English language teachers
- an opportunity for unqualified teachers already in post to gain a formal qualification
- a passport to opportunities for a career break or short-term work overseas.

Available in full and part time modes, the course includes taught units, hands-on teaching practice and observation of experienced teachers in the classroom.

The scope of the project

In 2003, Cambridge ESOL instituted a project to track the career paths of teachers who had completed the CELTA programme. The key aim was to address the following questions, adapted from an initial proposal for research made by Goldschegler, Parent, Rudolf & Freeman (1998).

1. What happens to people after CELTA? Where do they go? What do they do?
2. What impact does CELTA have on the careers of individuals?

3. What insights can this information give us about the design of the CELTA course?

To address these questions a questionnaire was developed by Cambridge ESOL which targeted details of:

- participants' backgrounds including their age, previous occupation and any previous EFL/ESL qualifications
- details of when and where they took the CELTA course
- details of jobs since completing the course
- opinions of the value and relevance of course content
- the impact of CELTA on their careers
- future career plans.

In February 2003, a letter was sent to all CELTA and DELTA (Diploma in English Language Teaching for Adults) centres requesting their involvement and asking them to distribute the printed version of the questionnaire and/or advise current or former CELTA candidates of the web address of the online questionnaire. Additionally, 400 recent CELTA graduates, who had indicated a willingness to participate in follow-up research, were contacted by email. Links to the form were also provided on the Cambridge ESOL teachers' website and in *Research Notes* issue 11 (May 2003). The findings reported below are based on 478 questionnaire returns received to March 2004.

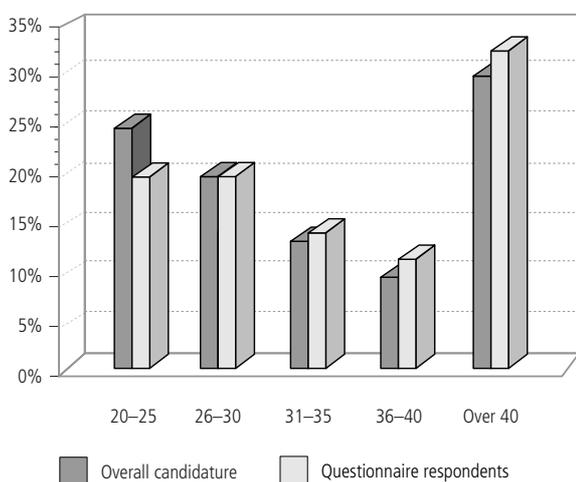
Project findings

Participants' backgrounds

The questionnaire respondents were, on average, slightly older than the CELTA candidature as a whole in 2003, with fewer respondents in the youngest and more in the oldest age categories answering the survey (see Figure 1). This is consistent with the varying lengths of time between the course and responding to the questionnaire among participants.

We would acknowledge that there may be important differences between those responding to the questionnaire and those who

Figure 1: Age of questionnaire respondents compared with overall CELTA candidature (2003)



chose not to participate. For example, respondents might be more likely to have succeeded in finding a job in English language teaching, or at least to have decided to pursue a career in ELT, and might therefore also have more positive feelings about their CELTA experience. This shortcoming must be borne in mind when considering the findings described below. It is also worth noting that most of the respondents had only recently completed their course (56% of respondents took the CELTA in 2002 or 2003) and so were unable to provide data on the longer-term impact of the course.

What were you doing in the year before you took the CELTA course?

Respondents came to CELTA from a wide range of working backgrounds ranging from financial services to tree surgery and from engineering to health and social work. However a substantial minority (38%) reported that they had been working in the field of education before taking the CELTA, and a further 17% had been studying at university or college. Three per cent reported that they had been unemployed with a further 2% 'travelling'.

Did you do any EFL teaching before you took the qualification?

39% of respondents reported that they had some ELT experience before taking the CELTA. This ranged from informal tutoring to quite extensive formal teaching. Comparing groups by age, those most likely to have previous experience were in the 36-40 age group (50%), the least likely were those aged 41-45 (30%).

A chi-square analysis showed that those lacking EFL experience were significantly ($p < .05$) more likely to feel that the course had failed to prepare them adequately for their first job, although very few respondents overall held this view (just 6% of inexperienced and 3.2% of experienced teachers).

Details of when and where they took the CELTA course

Where did you take the CELTA course?

Respondents took the course in a total of 30 different countries. This represents over half of the countries where CELTA is offered. A disproportionately large number of questionnaires came in from New Zealand (17%, compared with just over 5% in the 2003 candidature as a whole), reflecting the enthusiastic involvement of New Zealand centres in the project. A smaller proportion of respondents (35%) than in the 2003 global candidature (49%) was from the UK. 30% of respondents took the course in countries where English is not a first language.

Was the course full-time or part-time?

77% of respondents had attended the course on a full-time basis. This figure was slightly lower in L2 English speaking countries overall (72%) than in L1 English speaking countries (74% in the UK, 77% in Australia, 90% in New Zealand). Respondents who had taken the course part-time tended to be older; 41% of part-time candidates were aged over 40, compared

with 29% of full-time candidates. 47% of full-time candidates were aged 30 or under as against 30% of part-time candidates.

Details of jobs since completing the course

Please outline your teaching career since completing the CELTA, starting with your first teaching post after the course

Of those responding to the questionnaire, 83% of UK and 88% of overseas participants reported that they had found work after the course. However, it is not clear from the data how many of these had already been working in the same job before taking the CELTA. 32 respondents (7%) explicitly stated that they had not yet found a first job, some commenting that they had only just completed their courses. Many of those without work made comments such as, 'I haven't started teaching yet,' or 'not employed so far', suggesting that they were still committed to ELT. Only four respondents stated that they were no longer looking for a first ELT job, one stating that the CELTA had persuaded him that the profession was 'not for me'. A further eight respondents who had found work said that they had since left ELT for other careers (although some of these had plans to return). For those in work, the average length of time between completing the course and starting their first ELT job was 5.8 months.

From the admittedly limited data available, the oldest respondents were the least likely to have found employment. Of those aged over 45, 24% were yet to find a first post. The figure for other age groups ranged from 9% to 19% with 26 to 35 year-olds having most success. One of the older respondents (from the UK) felt that his problems in finding work reflected a wider problem in ELT, claiming, 'There is blatant age discrimination amongst EFL recruiters.'

Previous ELT experience did not seem to play a major role in securing work after the course. 35% of those who were yet to find a job following the course had some experience, compared with 40% of those who had found work. Employment rates were very similar (81% to 85%) for those taking the course part-time and full-time respectively.

Please specify town, country and dates for each institution where you taught

Among those in work, 52% of respondents taking the course in the UK also found their initial posts there. The 48% finding their first posts elsewhere obtained work in a range of locations, with 29 different countries represented. Spain, Italy, Portugal and Turkey were popular destinations, each attracting over 3.5% of UK participants. Outside Europe, China, Japan and Thailand were the most popular choices (2.9% each).

For those taking the course outside the UK, in other countries where English is a first language (L1), relatively higher proportions of candidates found their first jobs in the same country. These comprised 66% of those taking the course in Australia, 63% of those in Canada, 83% of those in South Africa and 89% of those in New Zealand. Like their counterparts in the UK, participants migrating from these countries travelled to a wide range of locations. Japan was the most popular initial overseas destination for candidates from Australia and New Zealand, attracting 5.2%

of respondents from these countries; otherwise first jobs were widely distributed. The majority (80%) of those taking the course in countries where English is a second language (L2) found their first positions in the country where they took the course. 8% moved to L1 English countries for their first job and 6% moved to countries with different first languages.

Of those finding first jobs, 53% had already moved on to a second, 22% to a third, 12% to a fourth and 3% to a fifth job by the time they responded to the questionnaire. 60% of those moving on to a second, 63% of those moving to a third, 54% of those moving to a fourth and 44% of those moving to a fifth post remained in the same country as in their previous job.

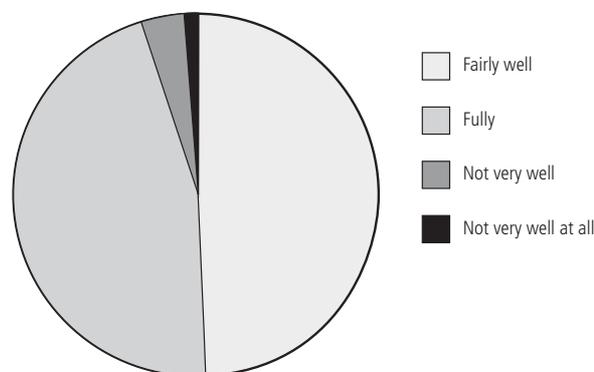
Overall, most respondents (69%) found their first jobs in private language schools (falling to 56% of third jobs). 14% found a first job in a college or university (rising to 18% for the third and 25% for the fifth job) and 7% in state schools.

Opinions of the value and relevance of course content

How well did your course prepare you for your first teaching post after the course?

Over 90% felt that the course had prepared them 'fairly well' or 'fully' for their first teaching post. Only five respondents felt that the course had prepared them 'not very well at all'. This would seem to indicate a positive level of satisfaction with the course on the part of respondents (see Figure 2).

Figure 2: How well did your course prepare you for your first teaching post after the course?



Can you give any examples of how well prepared/not well prepared you were?

Planning was mentioned 63 times, mostly as a beneficial outcome of the course although five respondents were negative about the level of detail required in planning on the course or felt that the plans they had been taught did not fit the teaching context they were working in. Fifty two mentioned the confidence that the course had given them, 30 mentioned teaching methods and 17 mentioned the ability to create and adapt materials.

Thirteen respondents mentioned classroom management. Although most of these commented that the course had improved their class management skills, three were negative, saying that the course had not prepared them adequately in this area. A further ten complained of discipline problems in their jobs that the course had not prepared them for. This probably reflects a wider concern

expressed by one respondent that, 'So much teaching nowadays is to younger students, why is the CELTA not made more general to cover all ages?' Nineteen respondents mentioned that they were now teaching children rather than, or in addition to adults. Cambridge ESOL does, in fact, offer an optional YL (young learner) extension to CELTA, or a specialist course in teaching for children, the CELTYL, which are intended to meet this need.

Thirty one respondents mentioned grammar with ten commenting on how useful the grammar input had been and ten suggesting that the grammar taught on the course had not been sufficient. One respondent with a foot in both camps commented that, 'Grammar learnt on CELTA whilst really useful, was the tip of the iceberg'. However, one complained of too much focus on grammar.

How could the course have better prepared you for your first post-CELTA teaching post?

One popular suggestion, made by 34 respondents, was an introduction to teaching children, young learners or adolescents. It appears from these responses that CELTA graduates are expected by some employers to teach young learners without further training (although this is offered by Cambridge ESOL through the YL extension). The representativeness of the student groups they had encountered on the course was another cause of concern. Fifteen respondents felt that they could have been better prepared to teach groups of different sizes (including one-to-one teaching) or groups at a variety of ability levels. Seven mentioned the need to teach groups of mixed ability. A further five would have liked training in teaching monolingual groups.

Other suggestions included planning for a greater variety of lesson types (12 responses) including planning for longer lessons than those included on the course and advice on how to plan a number of lessons quickly. Five respondents also requested more training in planning for a week of classes, or for a course. Eighteen respondents wanted more guidance on choosing and using course books with four asking for practice in lessons without a course book. Four felt that guidance on exam preparation would have been useful while six mentioned business English.

Training in aspects of the language was another theme. Thirty seven respondents mentioned grammar as an area they would have liked to learn more about and seven mentioned phonetics or phonology. Ten asked for more time on teaching practice and six for issues of classroom discipline.

Help in finding a job or more detailed careers advice were mentioned by twelve respondents. A number of these wanted more on aspects of the language teaching business and what to expect of prospective employers: 'I would allocate several hours to discuss locating and applying for jobs abroad, what to look for, how to check out the school before you go and how to negotiate a contract'.

Were there any aspects of the course which were not as useful as you had thought once you were in post?

The most common response to this question was simply, 'no' (95 responses of 318). Twenty three responded, 'no' followed by a

comment, 16 responded, 'none' and another 11, 'not really'. A further 20 commented that everything was useful.

Some of the criticisms were mutually contradictory; four felt there was too much theory on the course, two believed there was too little. Three believed there was too much attention given to using course books, one felt there was too little. However, one area that attracted more consistent criticism was lesson planning. Eighteen commented that the level of detail or rigidity of lesson planning on the course did not prove useful in their teaching posts, mostly because of time constraints. Comments included, 'The detailed lesson plans that were required were far too time consuming. Unrealistic.' And, 'This has been worthless to me. I don't have time to plan at all.'

How did CELTA training compare with any other kinds of training you have done?

'Intense' or 'intensive' was the word most often used to characterise the course. It was used by 83 respondents. 'Practical', (51 respondents), 'excellent' (28), 'challenging' (18), 'demanding' (13), 'useful' (13), 'thorough' (13) and 'rewarding' (11) were also popular descriptions. Eleven respondents commented positively on the organisation of the course, but one was negative about this aspect. Fifteen respondents were positive and one negative about the role of theory. Eight were pleased with the teaching practice, but two were not; one commenting that there was too little and another that they were inadequately prepared for it.

The impact of CELTA on their careers

Overall, how has taking the CELTA affected your career?

Although many respondents remarked that CELTA was only a first step, the majority clearly felt that the award had played an important role in opening career opportunities. Comments such as, 'I would not have my present job if I was not CELTA qualified,' and, 'Having CELTA enabled me to find a job,' were common. Over 100 made this kind of explicit mention of the positive impact CELTA had on their employment prospects or level of job security. Many spoke of 'opening doors' or 'options' (11 respondents), or of the course offering them a 'start' (23 respondents). However, the benefits of the award appeared to go beyond employment prospects. Fifty one respondents commented that the CELTA had given them confidence or made them more confident in their work.

The requirement in some areas for a degree, in addition to the CELTA, was a source of frustration for at least one respondent: 'So far I cannot get a job in many places because I do not have a degree, which seems stupid that I have the qualification and cannot get work where I would like to.' On the other hand, some respondents who did not hold degrees found that the qualification did open up opportunities.

Additional issues of recognition included difficulty in persuading employers that the CELTA was an adequate qualification in certain contexts such as the ESL sectors in the UK or Canada. There was also disappointment for L2 English-speakers that their opportunities, particularly for work outside

their own countries, seemed more restricted than those of their L1 English-speaking peers.

Future career plans

What type of post would you like to seek next?

Most who responded were intending to continue in the ELT field. In addition to other EFL teaching posts, future career moves included progressing to administrative posts such as Co-ordinator or Director of Studies, progressing to the DELTA or returning to full-time education to obtain a Masters degree.

How long do you expect to remain in EFL teaching?

Over half of those responding to these questions were unclear about how long they would remain in EFL: 13 responded 'I don't know', 6 'unsure', 4 'unknown', 3 'uncertain', 2 'undecided', 1 'undetermined', 2 'who knows?' However, many of the unspecific responses implied remaining in the field for some time with responses such as 'a few more years' (3) or 'a long time' (4). Approximately 5% of respondents said they intended to teach for up to 3 years, 6% for 3 to 6 years, 12% for as long as they continued to enjoy it or for as long as work was available and 15% for 10 years or more, or for the rest of their careers.

What careers advice would you give to someone about to embark on the CELTA?

Responses to this question took a wide variety of forms. Many complained of the long hours and low rates of pay involved. Some were very positive about the working opportunities, 'That it is guaranteed work as there is such a demand and a great way to work overseas'. Others were less sanguine, 'Be aware that ELT is a popular option for a huge number of people and you will need more than a Bachelor's Degree, CELTA and enthusiasm, to land a decent job'. Some were extremely enthusiastic about the course, 'Take the course! Teaching would have been so much more difficult without it. It's worth every penny.' Or, 'Don't hesitate, take it, enjoy it, take it seriously. You won't regret it.' Others were more cautious, 'Make sure it's really what you want to do'. And, 'I would tell that person to keep his day job, because he might not feel like teaching English all his life'.

There was advice for those taking the course, including suggestions that they should read about the topics ahead of the course, learn as much grammar as possible and gain teaching experience, 'Try out teaching for a while before (even if it is with private lessons) you do the CELTA course. I was able to absorb and digest what I learned better than those who had never taught.' Prospective trainees were warned to, 'Be prepared for a very intensive course' and to 'put everything else on hold'.

Conclusions

The tracking project is an example of the role of stakeholders in the on-going process of validation and test revision at Cambridge ESOL described by Saville (2003). It serves as an illustration of how monitoring can provide both evidence for the usefulness of our

examinations and suggestions for their further development to meet the changing needs of our candidates and test users.

This initial tracking exercise has provided some useful insights into the career paths taken by CELTA graduates. It is encouraging that CELTA has been of such benefit to so many of our graduates, although clearly attention must be given to the selection of participants before firm conclusions can be drawn from these data about overall rates of satisfaction.

From the responses, the CELTA appears to be a popular course that can have a life-changing effect on participants. Respondents were positive about most aspects of the course, even if their initial work experience sometimes limited their opportunities to put all aspects of their training into practice.

The responses showed that there is a desire for greater integration between qualifications such as those aimed at teachers of young learners or between the EFL and ESL sectors. These needs are being addressed by Cambridge ESOL in the YL extension to CELTA and through modular options for ESL training.

The global recognition of CELTA was an important factor in opening opportunities. Respondents found the qualification to be of value in the ELT marketplace in most, if not all, parts of the world.

The initial study is scheduled to end in December 2004 and, following an evaluation, Cambridge ESOL is considering how best to continue in our aim of tracking the impact of CELTA. Further information relating to the Cambridge ESOL teaching awards can be accessed at <http://www.cambridgeesol.org/teaching/index.htm>

References

- Bailey, K M (1999) *Washback in Language Testing*. Princeton, NJ: ETS.
- Cheng, L, & Watanabe, Y (Eds) (2003) *Context and Method in Washback Research: The influence of language testing on teaching and learning*, Hillsdale, NJ: Lawrence Erlbaum.
- Goldschegler, L, Parent, K, Rudolf, G, & Freeman, D (1998) *Concept Paper on Research and Documentation of the CLTS* (Unpublished internal report), Brattleboro, Vermont: School for International Training.
- Green, A (2003) *Test Impact and English for Academic Purposes: A Comparative Study in Backwash between IELTS Preparation and University Pre-sessional Courses*, unpublished PhD thesis, University of Surrey, Roehampton, London.
- Hawkey, R A (2001) The IELTS impact study: Development and implementation, *Research Notes* 6, 12–15.
- Mason, G, Williams, G, Cranmer, S, & Guile, D (2003) *How Much Does Higher Education Enhance the Employability of Graduates?* Bristol: Higher Education Funding Council for England.
- Saville, N (2000) Investigating the impact of international language examinations, *Research Notes* 2, 4–7.
- (2003). The process of test development and revision within UCLES EFL, in Weir, C & Milanovic, M (Eds) *Continuity and Innovation: Revising the Cambridge Proficiency in English Examination 1913–2002*, Studies in Language Testing Vol 15, Cambridge: UCLES/Cambridge University Press.

Cambridge ESOL and the NRDC ESOL Effective Practice Project

JAMES SIMPSON, UNIVERSITY OF LEEDS

This article describes the collaboration of Cambridge ESOL and the NRDC ESOL Effective Practice Project in the development of an instrument for assessing speaking.

Introduction: The NRDC ESOL Effective Practice Project

The ESOL Effective Practice Project (EPPP) is one of five such projects initiated by the National Research and Development Centre for Adult Literacy and Numeracy (NRDC), covering the Skills for Life areas of ESOL, numeracy, reading, writing and ICT. The project started in late 2003; data collection began in March 2004 and continues to April 2005; and the project ends in March 2006. The project teams, based at King's College London and the University of Leeds, aim to investigate the range of approaches to the teaching of ESOL in the UK. The project picks up themes from the Review of Adult ESOL Research (Barton and Pitt 2003) in its emphasis on the distinctiveness of the spoken language focus in Adult ESOL and in its recommendations for further research in the relationship between learners' outside experiences and classroom practices. It also builds on the findings of the NRDC Adult ESOL Case Study project (Roberts et al 2004) which, amongst other things, noted spoken interaction as a distinctive feature of ESOL classrooms.

Forty classes (c. 400 students) are being observed across sites which profile the diversity of adult ESOL provision in the UK. Part of the multi-method research involves establishing correlations between learner progress and pedagogic practice: the pre- and post-observation assessment instrument was developed for this purpose. In this article I describe the format and use of this instrument, and the role of Cambridge ESOL in its development. Although it is yet too early for full analysis of test scores, I draw attention to some of the questions which have arisen through its implementation.

The EPPP and Cambridge ESOL

The EPPP required a speaking test which could be administered at the beginning and the end of the observation cycle (an interval of 10–20 weeks), to ESOL learners at Pre-Entry Level, Entry Level 1 and Entry Level 2 on the National Qualifications Framework (up to ALTE Level One/Waystage). We approached Cambridge ESOL in the hope of finding an existing test which suited our needs, and for specialist training of the project team.

In consultation with staff in the Research and Validation Group, we decided to use the speaking paper of the existing Key English Test (KET) as the basis for the instrument, corresponding broadly as it does to ESOL Entry Level 2. On this basis I was given access to

retired forms of KET, the KET interlocutor frames and the KET assessment criteria to adapt and adopt. Part of this adaptation was carried out with a consultant trainer provided by Cambridge ESOL who led the training and standardisation sessions with the project team.

Training and standardisation

Training and standardisation for the administration of the test was carried out at the University of Leeds in February 2004. Re-standardisation was carried out at Cambridge ESOL in September 2004, immediately prior to the administration of the pre-observation assessment for the second round of observations. Eight project members have been trained and standardised.

Adaptations of KET for the EPPP

The speaking element of KET was not initially entirely suitable for the purposes of the EPPP. A number of issues had to be addressed, mainly to do with level, appropriacy and use.

Level

Many of the students on the EPPP, particularly those in Pre-Entry and Entry 1 classes, have language skills which are lower than those required for KET. These students fall into two groups: those with low level skills across the board, and those whose speaking skills may be good but who do not have the literacy skills to cope with Part 2 of KET (where verbal and visual prompts are used to stimulate interaction between learners). We therefore devised a lower level assessment, using the interlocutor frame from Part 1 of KET, which requires very straightforward exchanges of a predictable nature between the interlocutor and each learner in turn. We have found that this lower level assessment is easy to administer with students of a very low level of speaking ability. The assessment criteria and scales were adapted by the Cambridge ESOL trainer to reflect the standard required for the lower level test. The main adaptation is that the description for a score of 3 on the four elements of the KET scale becomes a 5 on the lower level EPPP test.

Appropriacy

The interlocutor frame of Part 1 of the KET Speaking test contains certain language which is not necessarily appropriate for the type of student we meet in ESOL classes. Many are from settled communities, but a good number are recent arrivals whose lives are in flux: refugees and asylum seekers. Thus questions such as 'How did you come to the UK?' and questions about the learners' families, had to be modified or omitted. Likewise the task types in Part 2 had to be chosen with some care. The assessment was

eventually stripped down to remove all reference to material that would offend cultural sensibilities. It is perhaps regrettable that in the interests of sensitivity only the most mundane topics were eventually chosen. These are prompts of a personal kind, but limited to the topics of *home, breakfast, going on a train or a bus, favourite room and favourite day.*

Use

KET was designed to be administered just once, after a course of language study, where all three papers are taken. The EEPP assessment differs in use quite considerably: learners are not following a course of study leading to this particular assessment, it is administered twice, and it only includes a speaking element.

The reasons for only assessing speaking for the EEPP are both practical and principled. We are keen to disrupt the ESOL classes we are working with as little as possible, and to administer tests for all language skills would raise that disruption to an unacceptable level. Moreover, the project as a whole has speaking as its prime focus.

For our pre/post-observation assessment instrument, some measures have been taken to reduce any skewing of the scores due to learner familiarity with the tasks and the interlocutor: learners do different tasks before and after; and the interlocutor is never the researcher who has observed the class (i.e. is never a 'known face'). Yet we still expect some pre/post score differences to be due to familiarity with the assessment. Learners are given only scant familiarisation in the task types before the first assessment takes place. In more recent administrations, an effort has been made to acquaint the learners with the format by circulating practice materials, thus providing the opportunity for practice before the assessment. Nonetheless, we remain concerned that we are administering the assessment to learners who may have little or no prior experience of taking a speaking test, and the effect this has on performance.

In sum then, the EEPP assessment instrument involves a paired test format, like KET, with an interlocutor and an assessor who takes no part in the interaction. However, it has been adapted according to level and appropriacy, and its use has been somewhat subverted as a pre-/post-observation instrument.

Two critical issues

Two issues recur while we administer the assessment on the EEPP: the sense that the sample of language gained in the test is not a good reflection of the learners' communicative abilities; and the matter of assessing learners with divergent speaking and literacy skills. The first of these issues involves a tension between interlocutor help and test validity on the EEPP, while the second has implications beyond the project.

Interlocutor effects vs. standardisation for validity

Standardisation of our assessment requires that little or no help is provided by interlocutors. On the other hand, interlocutors (practising ESOL tutors) find it hard not to help students who are

struggling to communicate. Thus while in most assessments the learners receive little or no assistance from the interlocutor, in some the interlocutor 'helps them along'. This interlocutor behaviour results from a wish to gain a sense of what the learner can do in normal conversation, particularly when they are very weak, nervous or reticent during the assessment. On a broader level, and cutting to the heart of some project members' concerns about language testing, we sometimes feel that our attempts to assess language make it difficult to assess communication. This raises the interesting question of how easily assessment procedures can take account of the distinction which is sometimes made between 'language' and 'communication' skills.

The 'spiky profile' of ESOL learners

High level speaking skills are coupled with low level literacy skills in large numbers of ESOL learners. Such learners are often said in ESOL circles to display a 'spiky profile'. As far as the EEPP is concerned, they fall between the two stools of our assessment: they cannot take the higher level assessment as they do not possess the basic literacy skills to tackle the task in Part 2; yet at the same time they are competent in giving the kind of personal information required for the low level assessment, and score very highly on the pre-observation instrument. This makes it difficult for any improvement to be reflected on the post-observation assessment. The implications of this for the project are not yet clear. What is clear is the difficulty of assessing the speaking skills of learners with low levels of literacy, and the need to devise assessments for speaking which are not reliant on candidates possessing even basic literacy skills.

Conclusion

This short descriptive paper comes too early to report 'findings'. Yet already during the administration of the EEPP assessments we have learned a great amount about the ESOL learners on the project. We are also very pleased that our collaboration with Cambridge ESOL is helping us to gain a sense of effective practice, of what works, in ESOL classrooms. And as the learners on the project are a cross-section of the ESOL learner population who will be taking Skills for Life ESOL exams, it is our great wish that bodies such as Cambridge ESOL will ultimately benefit from the EEPP, just as we have learned from Cambridge ESOL.

References

- Barton, D and Pitt, K (2003) *Adult ESOL Pedagogy: A Review of Research, an Annotated Bibliography and Recommendations for Future Research*, London: National Research and Development Centre for Adult Literacy and Numeracy.
- Roberts, C et al (2004) *English for Speakers of Other Languages (ESOL) – Case Studies of Provision, Learners' Needs and Resources*, London: National Research and Development Centre for Adult Literacy and Numeracy.

Closer collaboration with other Cambridge University departments

LYNDA TAYLOR, RESEARCH AND VALIDATION GROUP

During 2004 Cambridge ESOL was fortunate to be able to develop closer links with other departments of the University of Cambridge, primarily to provide expertise in language assessment but also to collaborate on projects of mutual interest and benefit.

During the first term of the 2004/5 academic year, a team of staff from the Research and Validation Group, under the direction of Dr Lynda Taylor, was responsible for teaching a course on the Assessment of Language Proficiency; this course is one of the option courses within the one-year MPhil programme offered by the Research Centre for English and Applied Linguistics at the University of Cambridge. The course covered sessions on: assessment principles/practice; designing and developing tests; assessing comprehension ability; assessing performance; assessing formal language knowledge; measurement and technological issues; alternative approaches to assessment; and assessment in context.

The Research Centre for English and Applied Linguistics was established in 1988, funded by a generous endowment from the University of Cambridge Local Examinations Syndicate (UCLES). The founding professor was Dr Gillian Brown who recently retired after 13 years in the post (see below for a seminar she recently presented at Cambridge ESOL). She was succeeded in October 2004 by Professor John Hawkins, formerly Professor of Linguistics at the University of Southern California. The primary purpose of the Centre is to undertake and promote research in English and Applied Linguistics. As well as the 12 graduate students following the nine-month MPhil course, a growing number of research students are following the PhD programme.

Although the Research Centre for English and Applied Linguistics is a free-standing department of the University, its core staff are members of the Faculty of English, which is the body responsible for overseeing the Centre's teaching programme, its graduate students and research students. In the last Research Assessment Exercise (2001) carried out by the Higher Education Funding Council, the Faculty of English, together with the Research Centre, gained the highest possible rating (5); and in the last Teaching Quality Assessment Exercise (1995), the Faculty, together with Centre, was pronounced 'Excellent' with respect to its teaching, assessment procedures and support for students. The current Centre prospectus is available online at <http://www.rceal.cam.ac.uk/index.html>

The Research and Validation Group has also forged stronger links with the University's Faculty of Education as part of the *Asset Languages* project. Karen Ashton and Ann Shih-yi Chen started work in 2004 as Research Assistants in the Research and Validation Group at Cambridge ESOL. While working on the development and validation of the *Asset Languages* assessments, they are studying for a PhD in the Faculty of Education at Cambridge University. Dr Neil Jones, Senior Research and Validation Co-ordinator, is acting as a co-supervisor.

Cambridge ESOL aims to continue developing collaborative links both within Cambridge, the UK and further afield, with academic, educational and other institutional partners. One way in which we achieve this is through inviting external speakers to give seminars at Cambridge ESOL.

ESOL Staff Seminar programme: Applied Linguistics: a personal view and Second Language Listening

CHRIS HUBBARD, PERFORMANCE TESTING UNIT

Professor Gillian Brown CBE visited Cambridge ESOL in November 2004 and presented a seminar on her views of Applied Linguistics and Second Language Listening. She began the session with a personal account of how the definition and content of the term 'Applied Linguistics' has been modified over time: from a narrowly focused view of the linguistic analysis of language features being applied directly to the realms of pedagogy and research, to a situation today where the number of areas of study which have developed as part of Masters programmes has led to a much less clear single understanding of what is meant by Applied Linguistics.

This brief history of terminology served as the foundation for the main focus of the session, which was Professor Brown's overview of the complex nature of second language listening. As the presentation began to tease out features of second language listening, so a number of pertinent observations and questions became apparent covering a variety of areas relevant to our work as language testers: features of speech which can influence the success of a listener; interpretation versus comprehension; the context of utterance; and, most provokingly, the question of how, or indeed whether, success in listening can be measured at all.

Speech contains elements which by their nature can have an effect on a listener's success to decode the message. Variability in phonological oppositions, segmenting manageable and meaningful chunks out of the acoustic blur of speech, and interpreting paralinguistic vocal features were all covered. It was argued that although it may not be necessary to be fully expert in all of these areas an ability to extract messages from a variety of verbal patterns may be one indicator of success. If so, this could be a consideration for test content.

The speaker next spent some time investigating the areas of interpretation and comprehension of meaning, and made a clear distinction between the active interpretation of speech with a resulting inference of meaning and listening comprehension. The main thrust of this division was that in a process of interpretation ideas can be understood in different ways and can be added to or modified, especially when internally filling the gaps in a message. With comprehension the focus is more on content. Professor Brown clearly demonstrated how interpretation skills were a requirement of successful listening, including correct

understanding in choice of syntax and structures, and how comprehension tests of the past had failed to address this.

The session as a whole provided a number of valid points for consideration in the areas of testing both speaking and listening skills. There is a continual challenge to create tests that reflect a candidate's ability in the real world whilst remaining valid, measurable and standardised. This seminar kept that challenge alive whilst also hinting at the directions to explore further as we seek to reflect the insights that the discipline of Applied Linguistics can provide.

References and further reading

- Brown, G (1990)(2 ed) *Listening to spoken English*, London: Longman.
—(1995) *Speakers, Listeners and Communication*, Cambridge: Cambridge University Press.
Rost, M (2002) *Teaching and Researching Listening*, Harlow: Longman.
Shockey, L (2003) *Sound patterns of spoken English*, Oxford: Blackwell.

Raising the Languages Ladder: constructing a new framework for accrediting foreign language skills

NEIL JONES, RESEARCH AND VALIDATION GROUP

Introduction

This article is about relating tests in different languages to a single interpretative framework. It aims at developing a methodology for the Languages Ladder, but relates to similar work going on elsewhere in relation to the Common European Framework of Reference (CEFR). Indeed, empirically establishing the relation to the CEFR is one of the goals of the project. In particular this article addresses procedures for linking objectively-marked tests (Reading and Listening) to the CEFR. Procedures for subjectively-marked tests (Speaking and Writing) will be dealt with in a forthcoming issue of *Research Notes*.

The Languages Ladder: a case study for framework construction

The scope of the project is wide, taking in 26 languages, across three contexts of learning (Primary, Secondary and Adult) and six major levels from beginner to very advanced. (See page 2 in this issue for an introduction.)

Implementing this complex multilingual measurement framework is clearly a huge challenge, but we can look to similar work going on elsewhere in relation to the CEFR – currently the focus of a great deal of scholarly effort. The Council of Europe has published a draft pilot Manual *Relating language examinations to the CEFR* (Council of Europe 2003), which proposes a methodology for how this can be done. Cambridge ESOL is

among the assessment bodies who have undertaken to do case studies based on the manual, and the Languages Ladder is one of those case studies.

However, while the manual relates more to single-language studies, the multilingual Languages Ladder demands a more explicit focus on the cross-language dimension. In order to equate tests to the CEFR we need to be able to divide a continuum of ability into rationally-defined intervals, and then replicate the procedure precisely across languages.

'Rationally-defined' means corresponding to the CEFR levels as defined by the illustrative scales currently provided. But while the CEFR levels are widely perceived as relevant and useful, the judgment of experts who have attempted to use them as test specifications is that they are underspecified and incomplete for this purpose (e.g. Weir 2004, Alderson et al 2004). So 'rationally-defined' must be extended to include:

- conforming to other existing well-established understanding of levels
- consistent with experience of the amount of time or learning effort required to change level
- plausible in terms of the relative size of levels in measurement units.

We will come back to this in the discussion of rationally-defined levels below.

The problem with task-centred standard-setting

Task-centred standard-setting refers to a group of procedures in which experts make judgments about the difficulty of objectively-marked test tasks and from there set cut-offs on an ability scale. Task-centred approaches have been adopted widely in CEFR-related studies – in the methodology proposed by the pilot *Manual*, the DIALANG Project, the Dutch CEF Construct Project and elsewhere. However, they can be thought inappropriate for framework construction, because they developed primarily to address pass-fail decision making, and, although they are justified by issues of policy and due process, it is now accepted that there is no ‘correct’ result.

As Zieky (2001:45) explains: ‘There is general agreement now that cut scores are constructed, not found. That is, there is no ‘true’ cut score that researchers could find if only they had unlimited funding and time and could run a theoretically perfect study.’ Thus different procedures will produce different cut scores, and any cut score represents subjective values on the part of the participants, reflecting among other things perceptions of the social cost of particular errors of classification.

Starting from scale construction

We need to locate a large number of elements (six levels, many languages) in a framework in as algorithmic a way as possible. Isolated human judgments, multiplied many times, are unlikely to achieve a satisfactory solution. We should work from a conception of the framework as a whole downwards to the placement of elements within it, rather than work upwards from individual standard-setting decisions in the hope that the resulting framework will display internal coherence. Approaches should have some chance of being open to validation.

Scale construction is logically prior to standard-setting. A coherent scale is needed, constructed using latent-trait measurement theory, to cover the whole range of ability in which we are interested. Only when this has been done can we define levels.

Of course, this is easier said than done, requiring as it does a lot of response data covering the whole range of ability. Indeed, the use of standard-setting in CEFR-related studies has to some extent been driven by the need to press ahead pending the arrival of response data (e.g. DIALANG). The Languages Ladder schedule also presents problems for a scale-based approach, as the initial three levels (Breakthrough, Preliminary, Intermediate) will be rolled out for about 10 languages before the higher levels are developed. Thus some data needed for constructing the whole scale will be unavailable. However, the proposed approach can still be of use.

Scale-based definition of levels

The scale-based approach proposed here has the following steps:

1. Construct a measurement scale defined by calibrated tasks.
2. Fix the upper and lower limits – the A1 and C2 thresholds in the case of the CEFR.

3. Interpolate the intermediate levels, according to a rationally-defined procedure.

The advantages of this approach are:

- It involves making only two decisions, and these concern the levels which are most easily judged.
- It defines the whole level system in terms of proportions: the procedure can be replicated across languages in a straightforward manner by preserving the proportions.
- Such replication makes no assumption that the scale lengths for each language will be identical (measurement precision may vary for a number of reasons).

In selecting the upper and lower cut-offs of the scale it is not necessarily an issue whether higher or lower measurable levels exist. What is critical is that agreement should be possible as to what kind of performance is characteristic of these thresholds. Learner-centred standard-setting methods should be practical to the extent that only two sample groups are needed, and such extreme high and low-ability groups are more easily identified as such (from their learning background, observable functional ability) than samples at intermediate levels.

Rationally-defined levels

Thus levels are defined as *proportional* intervals on a scale. How should those proportions look? We must be prepared to define an approach and muster arguments to defend its validity.

Observable differences and ‘natural levels’

We should first consider whether the CEFR levels have any particular meaning. Are there substantive differences – step changes, perhaps – between learners at different CEFR levels, analogous to, say, those between children at different Piagetian stages of cognitive development? Not really, one might say: surely the scale simply defines a continuum of ability, characterised through a discrete set of level descriptors. However, the author of the CEFR illustrative scales reports that the cut-offs were indeed set so as to find a best fit with an existing notion of levels. The procedure included ‘looking for patterns and clusters, and apparently natural gaps on the vertical scale of descriptors which might indicate “thresholds” between levels’ and ‘comparing such patterns to the intentions of the authors of the source scales from which descriptors had been taken or edited, and to the posited conventional or “natural levels”.’ (North 2000:272).

This, then, provides one possible substantive meaning of a level, and it relates to the particular group of learners which make it up. Elsewhere North says of the ‘natural levels’:

‘ELT professionals will find few surprises in the six levels ... since they correspond closely to the levels that have already established themselves in ELT. ... The levels have emerged in a gradual, collective recognition of what the late Peter Hargreaves of Cambridge ESOL described as “natural levels”.’ (North 2004).

‘Natural’ is a somewhat dangerous word because it suggests a rightness or inevitability about the levels which should not be

assumed. However, it does seem to capture the organic way that ELT levels developed to cater for significant groups of language learners. ‘FCE level’, for example, has been well-understood by publishers, teachers and learners since long before latent trait theory or the CEFR came along. How do such groups arise? How do they relate in terms of relative ability? This is what we consider next.

In what follows we need to understand the measurement scale against which levels are defined, that is, the meaning of a logit (the unit of a scale constructed using latent trait theory). A difference of one logit between a task and a person represents a specific probability of the person responding correctly to the task. Thus a difference of a unit entails a particular *observable difference* in performance. This might be reflected in the subjective impression of a difference made upon an observer encountering learners at different ability levels. Would we then expect natural levels to be separated by roughly the same observable difference? As we saw above, this is the basis on which the nine-level CEFR scale was defined.

CEFR and Cambridge ESOL levels compared

The nine-level thresholds of the CEFR illustrative scales are set by design to be roughly equidistant. However, the CEFR is more commonly treated as a six-level system, with the middle three levels (A2, B1 and B2) each composed of two sub-levels, and thus being about twice as wide as the other levels.

The Cambridge ESOL Common Scale links the five Main Suite exams from KET up to CPE. These Cambridge levels have been aligned to CEFR A2 to C2 using both analytic and quantitative methods; however, the match to the CEFR illustrative scale thresholds is approximate. In logit terms the lower Cambridge levels tend to be wider than the higher: KET (A2) is wider than PET (B1), which is slightly wider than FCE (B2). The Young Learner

exams constitute a three-level system from near-beginner to approximately A2 level. Given the very different nature of the candidature, the linking of YLE into the Cambridge or CEFR framework is proceeding with caution – in particular, there is as yet no ‘official’ YLE level corresponding to the A1 (Breakthrough) threshold. But for the present discussion it is notable that the logit interval defined by YLE between approximate A2 and the lowest band of the lowest level is as wide as that between the B1 and C2 thresholds.

Levels and learning hours

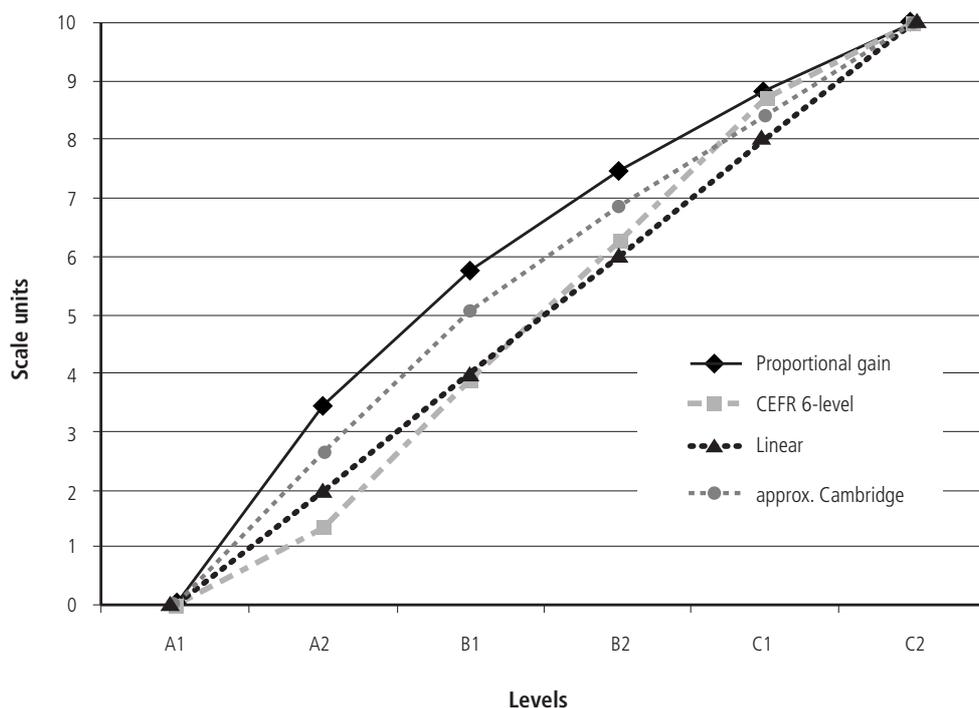
Levels are also frequently distinguished in terms of the number of ‘learning hours’ needed to achieve them. While the rules of thumb given by different sources (language schools, exam boards, the Council of Europe etc) tend to differ, there is some agreement: the higher levels take progressively more effort than lower ones. This implies that a given amount of learning makes a bigger observable difference at lower levels than at higher. This suggests that what is observable, and hence measurable, in language proficiency is *proportional gain*. The difference in the Cambridge ESOL logit bandwidths between lowest (Young Learner) and highest (CPE) levels clearly reflects this.

Learning gains are a notoriously difficult area, as what we tend to observe reflects an averaging over quicker learners, slower learners, and learners who are going nowhere. None the less, the notion of proportional gain still seems useful in explaining how observable difference slows as learning progresses.

So what should a levels system look like?

An approach to setting levels must therefore balance a number of considerations. Figure 1 illustrates the foregoing discussion by showing four scales defined according to different principles.

Figure 1:
Different approaches
to defining scales



The scales are shown anchored at the A1 and C2 thresholds (against an arbitrary 10-unit scale). The linear scale is the simplest: all six thresholds are equidistant. This implements the ‘constant observable difference’ principle.

The highest curve implements a ‘proportional gain’ model. If this model held, and the levels were indeed separated by a constant quantity of effort, then the number of learning hours between levels would be constant.

The CEFR and the Cambridge scales (the latter approximate because the A1 level is provisional) are empirically derived, although in very different ways. They differ from each other chiefly because of the double width of the three central bands in the CEFR. The Cambridge ESOL scale can be explained as follows.

In the early stages learning proceeds quickly. A relatively small amount of effort produces a very substantial change in observable behaviour – enough to warrant identifying a level and offering accreditation of it. As learning proceeds it takes progressively more time to make a substantial difference, and indeed, many learners plateau or drop out on the way. The higher levels are separated by smaller observable differences, but each level is needed because it accredits a final learning achievement or provides an interim target for those who wish to go further (e.g. Cambridge ESOL’s CAE exam was introduced at C1 explicitly to bridge a perceived gap between FCE and CPE).

If the aim is to define a progressive series of levels through objectively-marked tests, in a way which is consistent with people’s experience of how learning progresses, then the Cambridge ESOL scale seems to provide a good model.

The contrast between the Cambridge ESOL and CEFR scales is quite noticeable, although it is very difficult to know how much significance to attach to this, given the very different data and methods used to calibrate the two scales. Perhaps the measurement scales for the two methods simply operate on a different basis.

Where have all the learners gone?

The standard-setting literature discusses both task-centred and examinee-centred procedures, but focuses heavily on the former. The absence of the learner is also quite striking in the CEFR-related standard-setting movement. It appears that cultivating familiarity with the CEFR illustrative scales risks becoming a substitute for cultivating familiarity with the learners they describe.

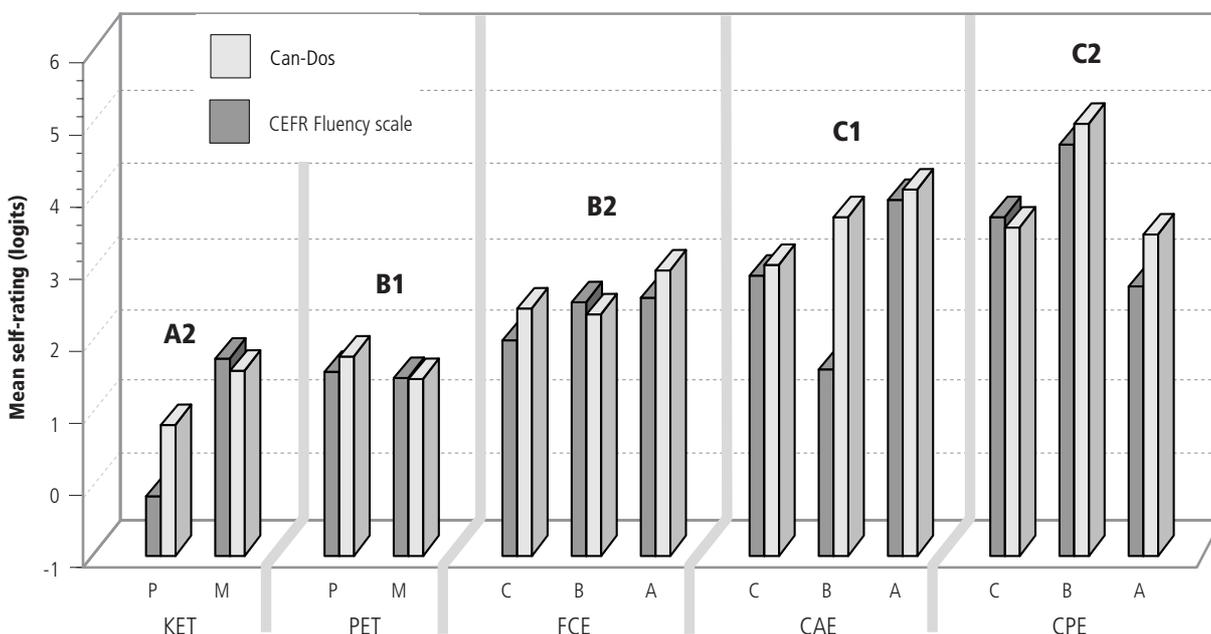
For the Languages Ladder project we shall be looking for practical ways of bringing learners back into the interpretation of test performance. One example of what can be done is the ALTE Can Do Project (Jones 2000, 2001, 2002). The Can Do project used over 7000 responses to self-report questionnaires to calibrate a number of performance scales. It succeeded in establishing an empirical link to the CEFR, achieved by the inclusion of some CEFR scales in the Can Do questionnaires, and to performance in language exams, achieved by collecting Can Do self-ratings from candidates for ALTE exams.

Studies have been undertaken for English (Cambridge ESOL), German (the Goethe-Institut) and Italian (Università per Stranieri in Perugia).

Figure 2 shows for English the mean Can Do self-rating of candidates grouped by the Cambridge ESOL exam grade which they achieved. The exams are ordered by level and within each exam the grades are ordered from lower to higher. The figure shows self-ratings on the Can Do statements and on the CEFR ‘Fluency’ statements separately estimated.

Not unexpectedly, the Can Do Project found evidence that a range of effects relating to groups of respondents – their age, first language, proficiency level, and area of language use – can affect their understanding of a scale and of the meaning of level descriptors expressed in Can Do terms. But the approach is nonetheless valuable as a practical way of giving meaning to the notion of cross-language comparability.

Figure 2: Mean self-ratings (Can Do statements, Fluency) by exam grade



Another study currently being undertaken by Cambridge ESOL is to equate the French, Spanish and German versions of the business language test CB BULATS, using plurilingual informants. This may provide a strong methodology for direct cross-language equating of tests.

We intend that Cambridge ESOL's contribution to the piloting of the *Manual* will emphasize learner-centred methodologies, providing ultimately more meaningful and practical approaches for the construction of a complex multilingual framework.

References and further reading

- Cizek, G J (2001) *Setting Performance Standards: Concepts, Methods and Perspectives*, New Jersey: Lawrence Erlbaum Publishers.
- Council of Europe (2001) *Common European Framework of Reference for Languages: Learning, teaching, assessment*, Cambridge: Cambridge University Press.
- (2003) *Relating language examinations to the CEFR. Manual; Preliminary Pilot Version*, retrieved from: [http://www.coe.int/T/E/Cultural Co-operation/education/Languages/ Language Policy/Manual/default.asp](http://www.coe.int/T/E/Cultural%20Co-operation/education/Languages/Language%20Policy/Manual/default.asp)
- DfES (2003) *The Language Ladder – steps for success*, retrieved from http://www.dfes.gov.uk/languages/DSP_languagesladder.cfm
- (2004) *Languages for all: From strategy to delivery*, retrieved from <http://www.dfes.gov.uk/languages/uploads/Languages%20Booklet.pdf>
- Jones, N (2000) Background to the validation of the ALTE Can Do Project and the revised Common European Framework, *Research Notes* 2, 11–13.
- (2001a) The ALTE Can Do Project and the role of measurement in constructing a proficiency framework, *Research Notes* 5, 5–8.
- (2001b) Relating the ALTE Framework to the Common European Framework of Reference, in Council of Europe (2001): 167–183.
- North, B (2000) *The Development of a Common Framework Scale of Language Proficiency*, New York: Peter Lang.
- (2004) Europe's framework promotes language discussion, not directives, *The Guardian*, retrieved from <http://education.guardian.co.uk/tefl/story/0,5500,1191130,00.html>
- Weir, C (2004) Limitations of the Council of Europe's Framework (CEF) in developing comparable examinations and tests, paper presented at BAAL conference September 2004.
- Zieky, M J (2001) So Much Remains the Same: Conception and Status of Validation, in Cizek (2001).

The Common Scale for Writing Project: implications for the comparison of IELTS band scores and Main Suite exam levels

ROGER HAWKEY, CONSULTANT FOR CAMBRIDGE ESOL
STUART D SHAW, RESEARCH AND VALIDATION GROUP

Background

This article reports briefly on Phases 2 and 3 of the Cambridge ESOL common scale for writing project, updating a previous article in *Research Notes*, (Hawkey July 2001), and adding findings relevant to comparisons between IELTS band scores and candidate writing performance levels in Main Suite exams – the Key English Test (KET), the Preliminary English Test (PET), the First Certificate in English (FCE), the Certificate in Advanced English (CAE) and the Certificate of Proficiency in English (CPE). The aim of such comparisons is to be able, in the long term, to link levels of writing performance as evidenced in the IELTS Writing Tests to levels of performance which have already been extensively analysed and described for our Main Suite exam writing components.

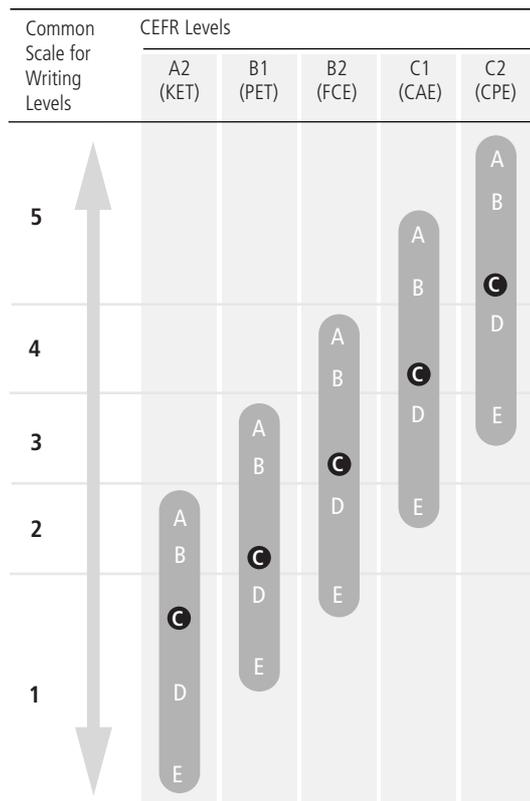
The Cambridge ESOL Common Scale for Writing (CSW) project has derived, from empirical investigation, a scale of descriptors of writing proficiency levels to appear alongside the common scale for *speaking* in the Handbooks for the Main Suite and other Cambridge ESOL exams. The scale is intended to assist test users in interpreting levels of performance across exams and locating the level of one examination in relation to another. A common scale, according to the Common European Framework of Reference for Languages (CEFR), may cover 'the whole conceptual range of proficiency' (2001:40). The location on a

common scale of proficiency of examinations for candidates at different levels should, the CEFR continues, make it 'possible, over a period of time, to establish the relationship between the grades on one examination in the series with the grades of another' (ibid, p.41).

It is clear that candidates for tests representing particular language proficiency levels (for example Common European Level B2 or ALTE Level 3) actually perform at a *range* of levels, some falling below the benchmark pass level for the exam concerned, some appearing to reach levels higher than the exam's top performance grade, (for example 'pass with merit'). Candidates may in fact be reaching a level normally associated with the exam one level higher up in the hierarchy (e.g. CAE at CEFR Level C1 rather than FCE at Level B2). Figure 1 (see also Hawkey 2001, Hawkey and Barker 2004) conceptualises the relationship between a *common* scale for writing (intended to provide descriptor bands for levels from elementary to advanced) and the levels typically covered by candidates for Cambridge ESOL examinations, each of which has its own pass level (the 'C' in Figure 1).

This article reports research aimed to contribute towards comparisons of levels of writing performance in Main Suite tests with IELTS Writing band scores.

Figure 1: Conceptual diagram of a common scale across examination levels and ranges



IELTS and the Writing Module

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 the language of communication' (IELTS Handbook September 2003:2; also see website www.ielts.org).

IELTS consists of four modules – listening, reading, writing and speaking – and the test scores provide a profile of a candidate's ability to use English at nine levels from Non User to Expert User. IELTS provides a choice between its Academic and General Training Writing Modules, which have been designed to take 60 minutes to complete and consist of two tasks (see Table 1). Task 1 requires candidates to write at least 150 words whilst Task 2, which carries more weight in marking than Task 1, requires at least 250 words. Each task is assessed independently. Detailed performance descriptors profile written performance at the nine IELTS bands. Task 1 scripts are assessed on the following criteria: *Task Fulfilment, Communicative Quality and Vocabulary and Sentence Structure*. Task 2 scripts are assessed on performance in the following areas: *Arguments, Ideas and Evidence, Communicative Quality and Vocabulary and Sentence Structure*. Scripts under the required minimum word limit are penalised¹. The format and content of the IELTS Writing Module are reproduced here as a table.

1. The assessment criteria have been revised for the Academic and General Training Modules and for Tasks 1 and 2. As from January 2005 there will be five assessment criteria: Task Achievement (Task 1) / Task Response (Task 2), Coherence and Cohesion (Task 1 and 2), Lexical Resource (Task 1 and Task 2) and Grammatical Range and Accuracy (Task 1 and Task 2). The band level descriptors for each of the assessment criteria have been revised accordingly.

Table 1: Format and Content of the IELTS Writing Module

| Academic Writing Module | General Training Writing Module |
|---|--|
| Appropriate responses in the Academic Writing Module consist of short essays or general reports addressed to tutors or to an educated non-specialist audience. There are two compulsory tasks. | The General Training Writing module requires candidates to write personal semi-formal or formal correspondence, or to write on a given topic. There are two compulsory tasks. |
| In Task 1 candidates are asked to describe some information (graph/table/chart/diagram), and to present the description in their own words. | In Task 1 candidates are asked to respond to a given problem with a letter requesting information or explaining a situation. |
| In Task 2 candidates are presented with a point of view, argument or problem and asked to present the solution to a problem; present and justify an opinion; compare and contrast evidence, opinions and implications or evaluate and challenge ideas, evidence or an argument. | In Task 2 candidates are presented with a point of view, argument or problem and asked to provide general factual information; outline a problem and present a solution; present and possibly justify an opinion, assessment or hypothesis, or present and possibly evaluate and challenge ideas, evidence and argument. |

Interpreting IELTS scores

IELTS is not a certificated pass/fail examination; rather it provides a profile of a candidate's performance. Assessment of performance in IELTS depends on how the candidate's ability in English relates to the language demands of courses of study or training, not on reaching a fixed pass mark. The appropriate level required for a given course of study or training is ultimately something which the institutions, departments or colleges concerned must decide in the light of knowledge of their own courses and their experience of overseas students taking them.

Receiving institutions are advised to consider both the Overall Band Score and the Bands recorded for each individual module, which indicate the candidate's particular strengths or weaknesses. In this way, language skills can be matched to particular courses. Receiving institutions are further advised to consider a candidate's IELTS results in the context of number of other factors, which include age and motivation, educational and cultural background, first language and language learning history. It is important for test users to remember that IELTS Band Scores reflect English language proficiency alone, and are not predictors of academic success or failure.

Demand for cross-test comparison

A frequent question asked of Cambridge ESOL is how IELTS scores align with scores from Main Suite and other Cambridge ESOL examinations. It is very difficult to make exact comparisons due to the different design, purpose, content and format of the examinations. It is also the case that candidates' aptitude and preparation for a particular type of test will vary from individual to individual and some candidates are more likely to perform better in certain kinds of tests than others. Clearly, any cross-test score alignment must be based upon a growing and continuing body of internal research of which the Common Scale for Writing is seen as an integral part. Research activity is combined with long established experience of test use within education and society,

as well as feedback from a range of test stakeholders regarding the uses of test results for particular purposes. In the final analysis, *'each test is designed for a different purpose and a different population, and may view and assess language traits in different ways as well as describing test-taker performance differently'* (Davies et al 1999:199).

The Research and Validation Group at Cambridge ESOL has spent considerable time and effort addressing the challenge of providing empirical evidence in support of the conceptual co-location of tests within a common frame of reference. That work has concentrated on defining a Common Scale for Writing through analysis of writing performance across different proficiency levels and across different domains (as realised in the writing test scripts of Main Suite and IELTS test-takers). This work has been recently extended to encompass tests which are modular (CELS) and tests from the Business English domain (BEC).

The Common Scale for Writing project: Phases 1 to 3

In Phase 1 of the CSW Project, existing writing assessment scales were used to draft a set of 'pass-level' descriptors of the writing proficiencies of candidates from CEFR A2 (Basic user, Waystage) through to C2 (Proficient user, Mastery) levels. A senior Cambridge ESOL examiner, Annette Capel, reviewed existing mark schemes and modified the descriptors for the levels represented by the Main Suite examinations. The result was a draft five-band common scale for writing characterised by criteria such as: *operational command of written language; length, complexity and organisation of texts; register and appropriacy; range of structures and vocabulary, and accuracy errors* (Saville and Capel 1995). In parallel CSW project Phase 1 research, Liz Hamp-Lyons investigated a corpus of PET, FCE, CAE and CPE exam scripts with the aim of characterising their proficiency levels through 'can do', 'can sometimes do', and 'cannot do' statements, for which she identified the following assessment criteria: *task completion; communicative effectiveness; syntactic accuracy and range; lexical appropriacy; chunking, paragraphing and organisation; register control; and personal stance and perspective* (Hamp-Lyons 1995). Hamp-Lyons noted, however, that the wide range of Cambridge ESOL exams and tasks covered by the script sample made it difficult to identify *consistent* features of writing at different levels.

Informed by the approaches and findings of Phase 1, Phase 2 of the CSW project set out to identify distinguishing features in the writing performance of ESOL learners across three Cambridge ESOL examination levels (FCE, CAE and CPE) and to incorporate these features into a scale of band descriptors common to the three levels. A corpus of 288 candidate writing performances was obtained on an identical writing task and each script was graded by more than one experienced and trained rater, using the FCE assessment scale.

As a first step in the drafting of band descriptors common to the three levels of writing performance, each of the 288 scripts was read and described qualitatively in terms of its salient features. Then, four sub-corpora of scripts, grouped according to the band scores assigned by raters, were identified for closer analysis and

the specification of their typical features. These were cross-checked for agreement through expert consultation. This qualitative analysis of the scripts in the four sub-corpora was supplemented by computer analyses of certain of the features identified as typical of each sub-corpus and of additional related features of potential relevance to the research questions. The characteristics and criteria identified were then 'rationalised into a draft scale of band descriptions for the proficiency levels specified, this scale to be proposed as a draft common scale for writing' (Hawkey and Barker 2004), and using descriptors which focused on three criteria:

- sophistication of language
- organisation and cohesion
- accuracy.

Phase 3 of the CSW project entailed the trial use of the draft Common Scale for Writing with corpora of candidate writing performances from other Cambridge ESOL exams. Qualitative analyses were performed using a similar approach to that applied to the CPE, CAE and FCE corpora in Phase 2, on IELTS, BEC and CELS candidate scripts, this time over a range of levels and tasks.

Findings on IELTS: Main Suite level comparisons

The primary purpose of the analysis, in Phase 3 of the CSW project, of a corpus of IELTS Writing scripts, was to trial and further validate the draft common scale band descriptors. This scale was modified progressively according to findings from its use with each new corpus of candidate writing.

The data for analysis were 79 IELTS writing performances including Academic Writing Task 1 scripts (description of iconic data), General Training Writing Task 1 scripts (letter writing), Academic Writing and General Training Task 2 scripts (both argumentative tasks), selected to include a wide range of IELTS band scores. Since the scripts were selected from a set of IELTS writing performances established as *certification* scripts used between 1995 and 2000, they had been multiply marked and identified as benchmark examples of particular levels. They had also been reproduced as word-processed text files to make them amenable to statistical software packages such as WordSmith Tools (Scott 2002) to reduce the potential impact of handwriting and photocopying.

The qualitative analysis of the individual IELTS writing performances included their description and rating according to the criteria and band levels of the draft common scale for writing (*sophistication of language, organisation and cohesion, accuracy*). Descriptions were achieved by providing an initial, impressionistic overview of the 79 IELTS scripts followed by a more detailed descriptive analysis of the features which emerged from each script. These descriptions and ratings were then compared with the IELTS profile and global ratings originally given to the performances in the corpus. Figure 2 illustrates the analysis of one script from the corpus.

Finally, a comparison was undertaken of the levels of IELTS writing performance with the Common Scale level bands previously identified. Figure 3, which averages the CSW band

Figure 2: Example of a script analysis and comparative IELTS: CSW bandings

| ID | Test version | Words | CSW criteria | | | Total CSW score | Overall CSW band | CEFR level | IELTS Band already assigned |
|----|--------------|-------|----------------------------|-------------------------|----------|-----------------|------------------|------------|-----------------------------|
| | | | Sophistication of language | Organisation & cohesion | Accuracy | | | | |
| 06 | 40 | 184 | 4 | 3.5 | 3.5 | 11 | 3/4 | B2/C1 | 6 |

Some fluency and range of lexis give occasional language sophistication; organisation and links clear though one cross-group analysis is missed. Accuracy errors, slight though sometimes surprisingly basic, do not impede meaning.

scores for scripts already assigned IELTS bands indicates a reasonably strong correlation between CSW and IELTS bands, although the average difference between the IELTS Bands 7 and 8 is not significant and there were too few IELTS Band 3 scripts in the sample to make any comparison.

Figure 4 takes account of the analysis of the corpus of 79 IELTS writing scripts to suggest a comparison between the IELTS Bands and the CSW levels, themselves tentatively associated with Council of Europe (CEFR) levels A2 to C2. It will be seen from Figure 4 that none of the 79 scripts in the study were assigned IELTS Bands 1 or 2 by the trained markers, only two were assigned a Band 3.

The IELTS: CSW comparisons from this study are not neat or proportional. There seem to be indications, for example, that:

- CSW level 2 (linked to CEFR level B1) could extend from the upper reaches of IELTS Band 3 to the lower reaches of IELTS 5
- IELTS 6 relates to CSW levels 3 and 4 (B2 and C1)
- CSW 4 reaches from the upper half of IELTS 6 into 7 and 8
- CSW 5 (C2) extends from high IELTS Band 8 into Band 9.

The wide band of writing performance apparently represented by CSW 4 corresponds to a CEFR level, C1, described across the skills by the Common European Framework as ‘an advanced level of competence suitable for more complex work and study tasks’ (2001:23) and in *writing* can-do terms, as follows: ‘Can produce clear, well-structured detailed texts on complex subjects, showing controlled use of organizational patterns...’ (2002:24). This level appears to extend from around the middle of IELTS Band 6 to Band 7 and even the beginnings of Band 8. (Interestingly, an overall

IELTS Band score of 6.5 is a common cut score used for university admission.)

High IELTS performance bands 8 and the top Band 9 appear to be the level of C2, CPE. Indications from the study are also that CSW level 5 (CEFR C2) stretches from high IELTS Band 8 to 9.

It is stressed that the inferences made here on IELTS band: draft CSW level relationships remain tentative, with the reminder, too, that the focus of the research referred to is on one macro-skill only, namely writing. Further similar CSW validation studies are needed, with larger samples and on different candidate test populations. It is perhaps also time to re-analyse closely and compare the IELTS Band and the CEFR level *descriptors* to see whether they explain the Band: Level comparisons emerging from corpus analyses. For example, are the CEFR’s *Threshold* constructs for the B1 level, with their implications of ‘enough language to get by’ (2001:28) but with *constraints* on range, accuracy and fluency, matched by a similar implied threshold between IELTS Bands 5 and 6? It may be; the IELTS Band 5 description (for Task 2) includes the word ‘limited’ three times, referring to ideas, lexical range and appropriacy, and sentence structures.

Findings on BEC: IELTS and Main Suite level comparisons

Given that the inferences made so far on CSW relationships with other scales remain tentative and that more validation studies are needed with different candidate test populations, a further study was made of BEC scripts. The data were drawn from the set of BEC writing co-ordination scripts used in 2002 and comprised a small corpus of 56 ‘live’ writing task scripts from BEC Higher (16 scripts),

Figure 3: Comparison of IELTS and CSW bands assigned to 79 IELTS scripts

| IELTS Bands already assigned | Common Scale for Writing Total Score (three criteria x band scores on scale of 1 to 5) and Band scores | | | | | |
|------------------------------|--|--|--------------------------------|---------------------------------|--------------------|----------|
| | Academic Module Task 1 | Academic Module Task 2 | General Training Module Task 1 | General Training Module Task 2 | Average CSW scores | CSW Band |
| 3 | - | 4, 5 | - | - | - | |
| 4 | 5.5, 6 | 5.5, 7 | 4, 4.5 | 6, 6.5 | 5.62 | 2 |
| 5 | 7, 8, 9.5 | 7, 7.5, 7 | 4, 8.5 | 7 | 7.3 | 2-3 |
| 6 | 11,11.5,12.5 | 8.5,11,11 | 11,7.5 | 12,8.5 | 10.45 | 3-4 |
| 7 | 12,11,11.5, 12 | 13,11,12 | 11, 12, 12.5 | 13.5,14 | 12.3 | 4 |
| 8 | 12, 11.5, 12.5, 12.5, 12.5, 13, 13 | 14, 11.5, 13, 12, 13, 12.5, 13.5, 12.5, 14, 12.5, 12.5, 12.5, 13 | 12, 12.5,12, | 13.5, 13, 11.5, 12.5, 12.5 11.5 | 12.6 | 4 |
| 9 | 15, 13, 13.5, | 14 | 15,14 | 15, 13.5, 15 | 14.22 | 5 |

Figure 4: Comparisons of IELTS, CSW and CEFR bands

| IELTS Bands | CSW Levels | CEFR Levels | Main Suite Exams |
|-------------|------------|-------------|------------------|
| 9 | 5 | C2 | CPE |
| 8 | 4 | C1 | CPE and CAE |
| 7 | | | CAE |
| 6 | 3-4 | B2-C1 | FCE/CAE |
| 5 | 2 | B1 | PET |
| 4 | | | |
| 3 | | | |
| 2 | | | |
| 1 | | | |

Figure 5: BEC scripts mean CSW band scores per level

| BEC scripts | n | CSW Band score means | CSW Level | CEFR Level |
|-------------|----|----------------------|-----------|------------|
| Higher | 16 | 11.16 | 4 | C1 |
| Vantage | 25 | 8.62 | 3 | B2 |
| Preliminary | 15 | 6.43 | 2 | B1 |

Vantage (25 scripts) and Preliminary (15 scripts); they consisted of a mix of Task 1 and Task 2 scripts. All the scripts were established certification scripts which had been multiply marked and identified as benchmark examples of particular writing performance levels within the BEC suite. The scripts had also been reproduced as word-processed text files.

Once again, each script in the corpus was described qualitatively and assigned CSW band scores for the three criteria (*sophistication of language; organisation and cohesion, accuracy*).

Figure 6: IELTS, BEC and draft CSW levels related

| CSW Levels | CEFR Levels | Main Suite Exams | IELTS Bands | BEC Levels inferred from corpus analysis and scoring | | |
|------------|-------------|------------------|-------------|--|---------|--------|
| | | | | Prelim | Vantage | Higher |
| 5 | C2 | CPE | 9 | | | |
| 4 | C1 | CPE, CAE | 8 | | | ● |
| | | CAE | 7 | | | ● |
| 4-3 | C1 B2 | CAE | 6 | | ● | ● |
| | | FCE | | | ● | |
| 2 | B1 | PET | 5 | ● | | |
| | | | 4 | ● | | |
| | | | 3 | ● | | |
| | | | 2 | | | |
| | | | 1 | | | |

Figure 5 shows the mean CSW band scores assigned to the 56 scripts. The averages showed a reasonable fit between the draft CSW band scale scores and the BEC levels.

The CSW ratings of all 56 scripts were then compared with the 'live' band ratings assigned to the same scripts using the BEC mark scheme. All but four of the 56 ratings correlated well. Further investigation indicated that in three of the four cases this was because the CSW scale did not penalise the scripts concerned as strictly as the BEC mark schemes for *missing or misunderstood information*, since this was a task-specific factor in the context of otherwise higher communicative performance. In the fourth case the candidate had written too few words for the three CSW criteria to be validly applied.

The analyses of the three BEC corpora appeared in general to support the hypothesised relationship between the draft CSW, CEFR levels and the IELTS levels shown in Figure 4 above. Figure 6 takes account of the CSW Project Phase 3 analysis of the IELTS and BEC corpora to suggest further tentative comparisons of levels.

Conclusion

The CSW project attempts to aid the production of a framework of descriptor bands including key criteria for the assessment of writing across exams at levels already specified by, for example, the Common European Framework. Such a scale would help comparisons of candidate performance across different exams. In this sense, inferences may be made about what a Band 3 on a FCE task might mean at CAE level, or what a Pass at BEC Vantage might mean in terms of the nine IELTS bands.

It must be pointed out that any inferences made so far across Cambridge ESOL exam bands and the draft CSW remain provisional. More similar CSW validation studies may still be needed, with larger samples and on different candidate test populations. Nevertheless, on-going research continues to refine our understanding of the relationship between Cambridge ESOL examinations and CEFR levels.

References and further reading

- Capel, A (1995) Common Scale for Writing Project, UCLES EFL internal report.
- Council of Europe (2001) *Common European Framework of Reference for Languages: Learning, teaching, assessment*, Cambridge University Press; Council of Europe.
- Davies, A, Brown, A, Elder, C, Hill, K, Lumley, T and McNamara, T (1999) *Dictionary of Language Testing*, Studies in Language Testing Vol 7, Cambridge: UCLES/Cambridge University Press.
- Hamp-Lyons, L (1995) Summary report on writing meta-scale project, UCLES EFL internal report.
- Hawkey, R (2001) Towards a common scale to describe L2 writing performance, *Research Notes*, 5, 9–1.
- Hawkey, R and Barker, F (2004) Developing a Common Scale for the Assessment of Writing, *Assessing Writing*, 9/2.
- IELTS Handbook* (2003), Cambridge: UCLES.
- Saville, N and Capel, A (1996) Common Scale for Writing, interim project report, UCLES EFL internal report.
- Scott, M (2002) *WordSmith Tools version 3*, Oxford: Oxford University Press.
- Taylor, L (2004) Issues of Test Comparability, *Research Notes* 15, 2–5.

Other news

ALTE Berlin 2005 conference

Registration for this event is now open and will close on Wednesday 11 May 2005. Please visit www.alte.org/berlin2005 for further information. The conference has received the patronage of the Council of Europe under the auspices of its Secretary General, Terry Davis.

Certificate in Further Education Teaching Stage 3 with the Certificate for ESOL Subject Specialists

This qualification is for people who have chosen to teach English in Further, Adult and Community Education in the UK – the 'Learning and skills sector'. Cambridge ESOL has developed an integrated two-module course, leading to the award of both these qualifications.

Module One is CELTA (Certificate in English Language Teaching to Adults), the best known and most widely taken TESOL/TEFL qualification of its kind in the world. Module Two focuses on the context of teaching ESOL within Further, Adult and Community Education, with emphasis on teaching and learning theory, pedagogic knowledge and teaching skills and values of teachers working with adult learners in the Learning and Skills Sector. Both modules are assessed through teaching practice and a series of assignments.

Visit the teaching awards website: www.cambridgeesol.org/teaching for further information.

Cambridge ESOL annual review

Cambridge ESOL's 2003/4 annual review is now available for download from our website: www.CambridgeESOL.org

Seminars for teachers on Certificates in ESOL Skills for Life

Following the success of the Skills for Life introductory events in September/October 2004, Cambridge ESOL is running a programme of seminars for practitioners who are teaching to the Adult ESOL Core Curriculum and preparing learners to take Skills for Life ESOL certificates.

The seminars will cover:

- format and content of Cambridge ESOL Skills for Life Certificates
- how the Certificates link with the Adult ESOL Core Curriculum
- assessment criteria and expected standards
- requirements for administering the exams.

See www.cambridgeesol.org/sfl for further details.

Cambridge ESOL multimedia CD-ROM

Aimed at potential candidates, as well as teachers and parents, this new interactive resource is an ideal introduction to Cambridge ESOL and its exams. It explores the range of exams available, the benefits of taking them and how they are run. To order copies, complete the Publications Order Form on our website: www.cambridgeesol.org/support