Research**Notes**

Contents

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**

Editorial Notes

Welcome to issue 25 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 testing Language for Specific Purposes (LSP) which is an area of growing relevance in a world where specific domains demand their own language assessments in addition to well-established General English provision. Linked to an increase in the number of domains (business, legal, medical, etc) requiring specific provision is the growth in tailor-made products for specific markets. Language testers need to be able to respond to such requirements to ensure that quality and reliability of language tests are maintained in line with existing products and frameworks.

In this issue we cover new language tests Cambridge ESOL is developing to meet the needs of language learners worldwide (e.g. Skills for Life, ICFE, ILEC), whilst maintaining the quality of our existing products and enhancing the support provision (Teacher Portfolio).

In the opening article Nick Saville discusses the role of language testing in migration and citizenship, as language testers are contributing to the ongoing debate about policy and practice in this area. ALTE, including Cambridge ESOL, is keen to ensure that language tests are used by governments or other institutions in a fair and appropriate manner.

In the following article Kate Ingham and David Thighe describe some of the issues involved in developing LSP tests, focusing on the International Certificate in Financial English (ICFE), including the relationship of test specificity to test generalisability and the importance of ensuring authenticity of test content. David Corkill and Martin Robinson then report on how Cambridge ESOL designs tasks for LSP tests, describing how the International Legal English Certificate (ILEC), an exam for the global legal community, is being developed with the assistance of legal expertise.

Clare Mitchell Crow and Clare Harrison then describe the Teacher Portfolio, a web-based professional development tool for teachers to document various aspects of their teaching career. This new provision sits alongside existing portfolios available for other groups, for example the EAQUALS/ALTE European Language Portfolio (ELP) for language learners.

Nadežda Novaković reports on the candidate profile after the first year of the Cambridge ESOL Certificates in Skills for Life (SfL). This modular exam suite follows the Adult ESOL Core Curriculum in England and is used to ascertain migrants' language proficiency. In the following article Fumiyo Nakatsuhara, a PhD candidate at Essex University and winner of the 2005 IELTS Masters Award, considers the impact of proficiency-level on conversational styles in paired speaking tests.

In the Research and Development update Stuart Shaw reports on a recent study which investigated the reliability of raters using a revised task penalty for the BEC Higher examination and Karen Ashton reports on current activities for Asset Languages. We end this issue with reports from the latest ESOL staff seminars and invited conference presentations given by Cambridge ESOL staff together with reviews of recent publications.

Language testing for migration and citizenship

NICK SAVILLE, RESEARCH AND VALIDATION GROUP

Introduction

The question of what role language testing and assessment should play in decisions made by governments and nation states regarding the admission of newcomers to their countries and the granting of citizenship is now being widely discussed. ALTE – the Association of Language Testers in Europe – has been seeking to make a significant contribution to the ongoing debate at both national and international levels with a view to ensuring that where language tests and their outcomes are used for purposes of migration or citizenship this is done in an appropriate and ethical way. In 2002 ALTE established a working group to coordinate these activities and members of this group have been working closely with the Council of Europe and with relevant government departments in their own countries. Cambridge ESOL has been taking part in the group and is contributing to discussions of the issues in the UK context.

The 20th century was characterised by the mass movement of people across national and international borders, especially for migration purposes. This trend is still continuing and Graddol (2006) observes that by 2000 the total number of international migrants had reached 175 million, or 3% of the world's population.

Factors underlying the movement of people on this scale are many and varied, but they are much the same as they have been over the past century; they include access to international education, international business operations, and more recently mass tourism, as well as economic migration and the search for asylum or safe haven.

More recent economic and socio-political developments such as globalisation, the breakdown of the former Soviet Union, and the extension of the European Union have accelerated the growth in people movement over the past 10 years. Whatever the underlying factors, the large-scale movement of people invariably has complex linguistic implications, not only for them as individuals but also for the societies they leave behind or those to which they relocate.

The role of language tests

The use of language tests in conjunction with the movement of people is nothing new. Tests like the Cambridge examinations have always provided achievable learning goals and meaningful accreditation for those who want to study or work abroad, or those who enjoy being international travellers and tourists. Some tests, like IELTS, have been used for many years by educational institutions for international study/training purposes; more recently they have been used for registration and licensing purposes by professional bodies dealing with international applicants in domains such as the health professions and government agencies.

One consequence of the large-scale and growing movement of populations around the world in recent years has been that notions of citizenship, and the role of language within citizenship, have assumed an increasingly high profile; this is particularly true within the European context. For some years now a shift has been taking place in many European countries towards more rigorous conditions for those seeking to apply for citizenship. One of the new (or renewed) conditions for obtaining citizenship is language proficiency. In some countries those applying for citizenship are asked to provide evidence that they have attained a certain level of proficiency in the official language (or one of the official languages) of the country. More and more frequently language tests are being used for this purpose. However, the socio-political contexts in which these conditions have been instituted and language tests developed, differ widely. The grounds for stipulating language as a requirement for citizenship, for example, in Slovenia or Latvia may differ from the justification in the Netherlands or Germany. At the same time, it is interesting to examine the rationale in other countries, such as Belgium, for not stipulating language and a language test as one of the conditions for attaining citizenship.

Not surprisingly perhaps, the language testing community finds itself being drawn increasingly into discussions about the role of language tests in the management of migration and decisions about citizenship; the use of language tests, sometimes in conjunction with tests of cultural knowledge, to determine whether an individual should be granted formal citizenship is a sensitive issue and one which is sometimes hotly debated.

At Cambridge ESOL we maintain an active interest in the educational and socio-political issues associated with migration and citizenship, since these touch directly upon language and education as well as culture and identity in today's society.
For example, as the British government seeks to develop and implement appropriate policies for managing migration in the UK context, exams like our ESOL Skills for Life suite introduced in 2005 are likely to play a growing role. Along with our partners in the Association of Language Testers in Europe (ALTE), we are seeking, as specialists in language testing and assessment, to participate positively in the European-wide debate over the role of language testing and assessment in this area; ALTE's status as an international non-governmental organisation (INGO) for the Council of Europe[1] makes it especially well-placed to contribute to this debate.

1. The Council of Europe is an intergovernmental organisation whose principal aims are to: protect human rights, pluralist democracy and the rule of law; promote awareness of Europe's cultural identity and diversity; seek solutions to problems facing European society; help consolidate democratic stability in Europe; and promote unity in diversity. Founded on May 1949, the Council of Europe now has 46 member states, including the 25 European Union states. Its permanent headquarters are in Strasbourg, France.

The Council of Europe designated 2005 as the European Year of Citizenship through Education. One of the aims of the year was to support democratic citizenship and participation to promote social cohesion, intercultural understanding and respect for diversity and human rights. In that context, the Council of Europe (Language Policy Division, Strasbourg) and ALTE set up a joint forum at the second ALTE Conference (in Berlin, May 2005) to focus on the political and ethical issues involved in defining and assessing the language proficiency required for citizenship and active participation of newcomers in social, occupational and democratic processes. The one-day *Language Testing and Citizenship Forum* had the following aim: to contribute to the discussion on the purpose of language assessment for citizenship throughout a wider Europe by organising a debate with key stakeholders such as politicians, policy makers and test developers. Through keynote presentations and chaired debate between presenters and audience, a number of questions were addressed:

- Why should proficiency in 'national' languages be a requirement for obtaining citizenship and a prerequisite for integration and social cohesion?
- Can a tension be observed between European policies and national policies?
- How should language proficiency be measured? At what level?
- What is the rationale behind determining a level of language proficiency for citizenship purposes?
- What is good practice and good ethical behaviour for professional language testers?
- What should a professional test developer do in order to make language tests for citizenship as ethical as possible?
- How can a European organisation of professional language testers like ALTE contribute to such a process?

One direct outcome of the *Language Testing and Citizenship Forum* at the Berlin 2005 Conference was the production of an information leaflet by the Council of Europe's Language Policy Division in cooperation with ALTE. The leaflet aims to inform interested parties of some of the key issues to be taken into account when considering the use of language tests for migration and citizenship purposes; the content of the leaflet is reproduced here in the remainder of this article.

Ethical and technical issues

Many European countries are introducing or formalising linguistic requirements for migration and citizenship purposes and national governments often require language tests or other formal assessment procedures to be used. There may be considerable differences in approaches to language tuition and to testing for residence or citizenship purposes.

Where language tests are used, and given the wide-ranging consequences of using such tests in high-stakes contexts, it is important for all test developers and policy makers to follow established Codes of Practice to ensure that all aspects of the assessment process are of high quality and that all stakeholders are treated fairly.

Professional test developers can assist policy makers in ensuring that suitable language tests are chosen and, where necessary, new tests are constructed, used and reported appropriately for these purposes; their contribution should address both the ethical and technical issues that can arise.

Council of Europe policies on Social Cohesion

Language learning and testing in Europe can be contextualised within the declaration on Education for democratic Citizenship (1999) adopted by the Committee of Ministers of the Council of Europe, which states that the purposes of education for democratic citizenship are:

- to equip men and women to play an active part in public life and to shape, in a responsible way, their own destiny and that of society;
- to instil a culture of human rights;
- to prepare people to live in a multicultural society and to deal with difference knowledgeably, sensibly, tolerantly and morally;
- to strengthen social cohesion, mutual understanding and solidarity.

The Committee of Ministers of the Council of Europe, in its Recommendation Rec (2002) 12 on education for democratic citizenship, recommends that 'all public and private, official and non-governmental, professional and voluntary actors [be involved] in designing, implementing and monitoring policies on education for democratic citizenship.'

The Council of Europe's concern for the social inclusion and active participation of migrants in society is reflected, for example, in the provisions for language tuition for migrants contained in the European Convention on the Legal Status of Migrant Workers and the European Social Charter.

International Codes of Practice and Standards

The outcomes of language tests which are used for migration or citizenship purposes have serious consequences and affect the lives of the test takers. Whereas high-quality tests and appropriate uses can facilitate the process of language learning and integration of the test takers within a multicultural society, poor-quality tests or inappropriate uses can endanger integration and may lead to social exclusion.

Professional test developers have a responsibility in raising awareness of the implication of national policies for the development of valid and reliable tests, taking into account diverse purposes and contexts of use which can exist. Their expertise and experience can ensure that the needs of both the policy makers and test takers are fully taken into account. However, the responsibility for the appropriate selection, interpretation and use of tests is a shared one: the policy makers (often the test sponsors), the users of the results (employers, government officials, etc.) and the test takers themselves all need to play their part in striving for fairness. To do so they need to be kept adequately informed and to be consulted in appropriate ways.

There is a need to be aware of the international Codes of Practice and Standards which have been developed to help ensure fairness in language-testing practices. These codes and standards typically cover:

- test development and routine test construction
- test conduct and administration systems
- test marking, grading and issue of results (including certification)
- test analysis and validation procedures, including provision of evidence to back up claims made about the tests' validity and reliability.

Key issues to be addressed

Questions of the following kind should be raised by those involved in planning and developing assessments, and well-reasoned answers supported by evidence of good practice need to be provided.

- *What is* ***the purpose of the test*** *and how does this influence the level, the content, the administration and the use of results?*

It makes a difference whether a test is meant to motivate the learners (to help them use and improve their current competence in the target language), to ascertain whether their competence is sufficient for participation in well-defined social situations (for study or work), or to make decisions which affect their legal rights, such as their right to remain in the country or acquire citizenship.

It is very important for the test takers to know the purpose, so that they can prepare accordingly. It is also necessary for other members of society to understand the intended purpose so that the results can be interpreted and used correctly.

- *Which* ***level of language proficiency*** *is required in the target language for the stated purposes, and how does this relate to international level systems, such as the Common European Framework of Reference for Languages (CEFR) developed by the Council of Europe?*

An appropriate level of proficiency in the target language should be chosen to reflect the stated purpose of the test and use of the results. In building a rationale for this, the CEFR can be used to ensure transparency and coherence of these aims. Test development and validation procedures must be employed to ensure that this level remains stable and consistent whenever the test is used.

- *Who is responsible for the* ***administration of the tests*** *to ensure that professional standards of conduct are met?*

Standardised procedures are necessary for the accreditation of suitable test centres and the administration of the tests. There must be adequate provision of information and support systems for the test takers and other test users.

- *What procedures are in place for* ***monitoring test outcomes****, including possible negative impacts, and what evidence is collected to* ***demonstrate fairness****, and that the test design and use does not lead to bias or unfair discrimination against some test takers?*

Data should be collected and analysed regularly to ensure that the test is valid and reliable and fulfils its intended purpose effectively, and to monitor for unintended consequences.

Changes in the groups taking the test, as well as the effect of general changes in the contexts in which the test is being used, can only be detected and acted upon when there is a proper system of monitoring the results of the test for the various groups of test takers.

Conclusion

Work on issues relating to migration, citizenship and language assessment continues, building on work already completed in the ALTE context specifically in relation to the debate in France and Germany. For example, early this year ALTE representatives (including staff from Cambridge ESOL) were invited to join an Ad Hoc Project Group which forms part of a mid-term Council of Europe project to look at Language, Migration and Social Cohesion. ALTE's contribution to the project is intended to provide advice and support to government departments and non-governmental organisations in their understanding and use of language assessment for migration and citizenship purposes. The ALTE group working on these issues met in Copenhagen in April and again at Sevres in May to discuss two specific contributions to the project: the development of a survey questionnaire and an extended information booklet which will elaborate on the information provided in the leaflet described above. Both will be completed and presented at Council of Europe meetings in September this year.

For a helpful and accessible discussion of current trends, and their implications for the English language in particular, see David Graddol's recently published report for the British Council (Graddol 2006).

References and further reading

ALTE website: www.alte.org

Council of Europe website: www.coe.int

Council of Europe (2002) *Recommendation Rec (2002)12 of the Committee of Ministers to member states on education for democratic citizenship*, http://www.coe.int/T/e/Cultural_Co-operation/Education/E.D.C/Documents_and_publications/By_Type/Adopted_texts/

Graddol, D (2006) *English Next*, The British Council.

Issues with developing a test in LSP: the International Certificate in Financial English

KATE INGHAM, ASSESSMENT AND OPERATIONS GROUP
DAVID THIGHE, RESEARCH AND VALIDATION GROUP

Introduction

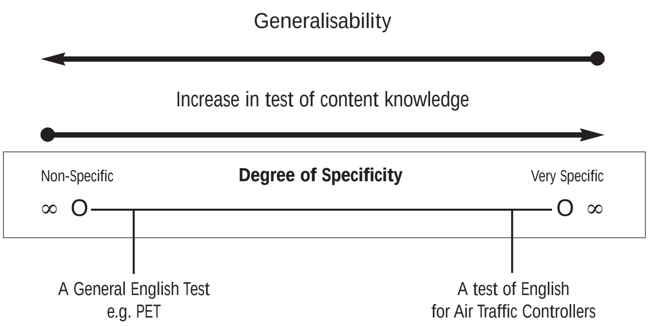
Testing Language for Specific Purposes (such as a Test of English for Engineers) refers to language assessment in which the test content arises from an analysis of specific target language use situations: these often (but not always) correspond to the language needs of a particular occupational group. Devising LSP tests presents test developers with a number of issues, including the relationship of test specificity to test generalisability; the importance of ensuring authenticity of test content; the interaction between background content knowledge and language knowledge, and for some domains, the difficulty in gaining access to relevant information on the nature of language use in that domain.

This paper examines these issues in relation to the International Certificate in Financial English (ICFE), a test of English assessing language in the context of finance and accounting which is being developed by Cambridge ESOL in partnership with the Association of Chartered Certified Accountants (ACCA).

Specificity vs Generalisability

LSP tests have often been directly contrasted with general purpose tests. This is now, however, generally acknowledged to be an over-simplification of the issue and there is growing consensus that tests do not fall into one grouping (specific purpose) or the other (general purpose), but that, in the words of Douglas (2000:1), '... there is a continuum of specificity from the very general to the very specific...': all tests are devised for some purpose and fall at some point along the specificity spectrum. The concept of a spectrum or continuum of specificity raises the question of where on the continuum a test should be placed and the related issue of how generalisable the LSP test is intended to be. As illustrated in Figure 1, generalisability is often held to decrease in proportion to the specificity of the test: the more specific a test (such as English for Air Traffic Controllers), the less possible to generalise from that to other language use situations. This is accepted as a fundamental issue in LSP, to which there are no straightforward answers.

Figure 1: The concept of the 'specificity continuum' (adapted from O'Sullivan:178)



Background content knowledge

In general purpose language testing, background knowledge of topic or cultural content is viewed as a 'confounding variable', which should be minimised as it has the potential to lead to measurement error. For LSP tests, however, subject specific content is arguably a defining feature of the test. Nonetheless, the question of 'separability', that is, how to distinguish between language knowledge and specific background knowledge in analysing candidates' results on a specific purpose language test has been a recurring concern. Bachman and Palmer (1996) argued in relation to a test for trainee doctors, that it should be possible to control for background medical knowledge in interpretation of performance on a language test, by, for example, the administration of 'knowledge tests' alongside the LSP test. The difficulty in assessing the *extent* of the test taker's background knowledge and its interaction with language proficiency has been addressed by Clapham (1996) who concluded that background knowledge was undoubtedly a significant factor in the process of testing reading, but the extent varied with the specificity of the test and the language proficiency of the candidate. There has more recently been an acceptance that until more is known about how the mind deals cognitively with ability and knowledge, specific background knowledge and language performance need to be treated as being 'inextricably linked' (Douglas 2000:39).

Access to information on language use within the domain

With an increase, in the second part of the 20th century, in the number of people needing to learn English for education, technology and commerce, the main drive behind the development of LSP was practical rather than theoretical. As a result, LSP itself may be said to have suffered from a lack of theoretical underpinning. A key analytical tool has been the use of Needs Analysis to assess the linguistic requirements of a particular target group. Some analyses resulted in long detailed lists of needs for which empirical verification was held to be lacking. Widdowson, for example, described many LSP Needs Analyses as being made up of 'observational lists with no basis in theory'

(Widdowson 1983:8). Alderson, Davies and others have raised similar concerns (Alderson 1988, Davies 1990, Skehan 1984). A further criticism of some needs analyses was that they lacked objectivity, were 'influenced by the ideological perceptions of the analysts' (Robinson 1991:7) and took insufficient account of the students themselves. Nonetheless, assessment of language needs can still inform LSP course and test design. As Clapham has said, 'We now know that such analyses can become too detailed, and also paradoxically, too limited in scope. However, this does not mean that they are unnecessary' (Clapham 1996:5).

Analysis of texts and spoken discourse from particular target language use situations is important in revealing how the target language use community communicates and disseminates information. The growth of corpus linguistics and the corresponding development of electronic databases of texts can help in enabling the identification of specific syntactic patterns and use of specific lexis among particular occupational groups or discourse communities. At present, however, there is a limited number of such corpora available and genre analysis plays an important role when considering communication between members of the occupational group or discourse community in question. According to Swales (1990), texts belonging to a particular genre share common features with regard to the organisation of information, rhetorical conventions and lexico-grammatical patterns which practitioners within that discourse community need to access and use in order to operate with any degree of effectiveness. Bhatia (1993) developed earlier work by Swales and has extensively researched language use in professional contexts, particularly discourse within business settings. Nonetheless, due to the confidential nature of the work done by some occupational groups (such as lawyers or accountants), access to texts from those domains may not be easily acquired. Swales (1996 cited in Flowerdew and Wan 2006) refers to such texts as 'occluded', genres to which access is normally denied to those outside the participating discourse community. One task for the test developer in such circumstances therefore lies in obtaining subject-specific assistance and advice. Bhatia (1993) reports on how the subject specialist or 'specialist informant' has played a role within LSP genre analysis.

An approach to LSP test development

In the light of these issues, what is an appropriate methodology to be followed when choosing to develop an LSP test aimed at the finance and accounting community? The approach of Cambridge ESOL has been to establish that need for an LSP test exists by conducting initial market research. The purpose and content of the potential test is informed by use of a Needs Analysis, and by working with subject experts (in this case a leading accountancy body) at the planning, design and development stages of draft specifications and sample papers. A Market Viability Study targeting stakeholders in the field and trialling on a representative sample of the target candidature are used to obtain feedback and establish 'proof of concept', that is, demonstrate the feasibility of the idea.

The need for a test of Financial English

Globalisation and the increase in cross-border finance have led to internationally recognised frameworks for financial reporting, such as the introduction of International Financial Reporting Standards (IFRS). One result of this is that accountancy has become a global profession, affording finance professionals the opportunity to pursue international careers. An international certificate in financial English could assist employers operating in the international finance arena in the recruitment and training of employees, and universities with the selection and placement of students whose first language is not English but who are required to study through the medium of English. A qualification in financial English could also help finance professionals and recent graduates to improve their job prospects and career mobility.

Initial telephone-based research focused on ascertaining the importance of English for members of accountancy bodies in non English-speaking countries. The survey population was made up of 41 contacts in thirty countries. 49% of the respondents felt that finance professionals in their country would need to use English every day in their work. 22% thought that English would be needed for work on most days. 73% of the respondents considered that finance professionals in their area needed to use and understand accountancy terminology in English 'to a great extent'.

A subsequent survey (Grewcock 2006) by Cambridge ESOL Business Support Group of a range of stakeholders in finance and accounting (which included large accounting firms such as Deloitte & Touche, Ernst & Young, KPMG and Grant Thornton; companies with in-house accounting departments and tuition providers of courses in accountancy) revealed an overwhelming consensus that formal assessment of English in a finance context was needed. As shown in Figures 2 and 3, 82% of respondents agreed or strongly agreed that a need exists with only 2% in disagreement. Moreover, almost 70% agreed that their institutions would use a test of Financial English if one existed (mainly for recruitment, screening or promotion purposes).

Developing a test of Financial English

An influential accountancy body, the Association of Chartered Certified Accountants (ACCA) which offers globally recognised accountancy qualifications, supported the need for an English language qualification aimed at accountants, and expressed interest in collaborating with Cambridge ESOL on the development of a test. ACCA itself reported a strong level of support from within its network of global contacts for the idea of offering specialist accounting-based English language support programmes.

Needs Analysis of the Target Language Use situation can play a useful role in test development. An external consultant was therefore commissioned to produce a Needs Analysis report, based on questionnaires and interviews. Sources included ACCA staff and accountants in the UK and abroad. The Needs Analysis focused on the language skills used by accountants in work contexts, details of those situations and examples of typical language style and content. Accounting genres were also examined together with typical frequency of use. This information fed in to the

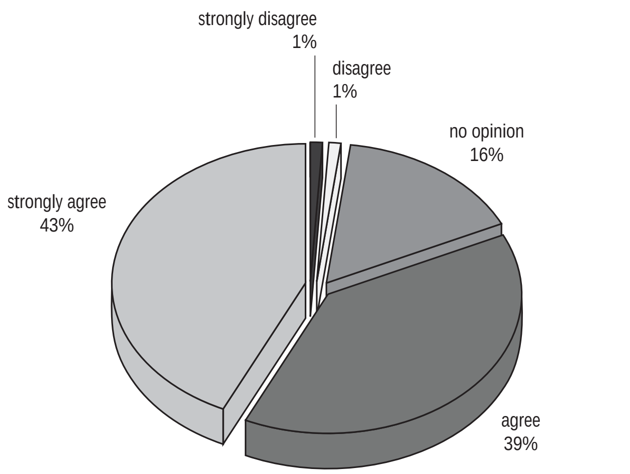


Figure 2: Response to statement – There is a need for a test in Financial English

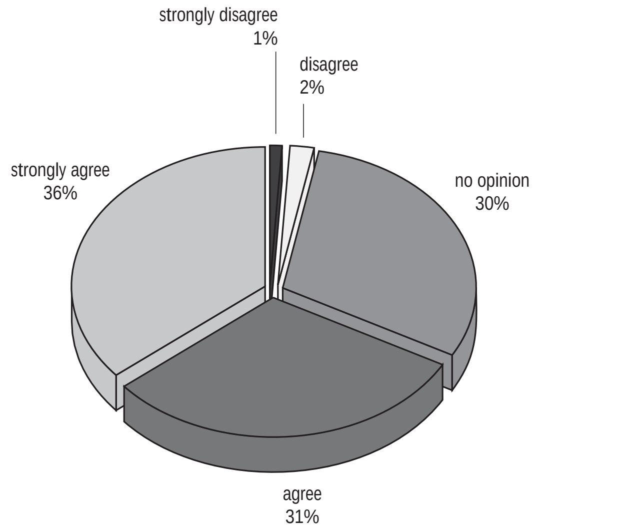


Figure 3: Response to statement – Our institution would like to use a test in Financial English

development of draft test specifications for a proposed examination – the International Certificate in Financial English (ICFE) – together with sample papers to exemplify the draft test specifications.

The Writing paper is an illustration of this: the Needs Analysis revealed that correspondence between accountancy firms and banks, tax authorities and clients is a written form of communication frequently needed by finance professionals. Furthermore, correspondence is often in the form of a response to an earlier letter and includes reference both to this text and to other documents or texts, such as tax statements, balance sheets, company accounts. This reflects the concept of *intertextuality* as identified by Kristeva (1980:69); research by others (Flowerdew and Wan 2006) has confirmed the prevalence of the interaction between texts within the tax accounting community. To reflect the findings of the Needs Analysis, one task on the ICFE Test of Writing requires candidates to draw on a previous text and compose a response to it with the use of notes. Composing the response requires the candidate to use a range of functions including clarifying, refuting, requesting information, referring the target reader to other documentation.

Cambridge ESOL worked closely with content experts at ACCA in establishing the test specifications and range of possible text sources. This relationship, together with information from the Needs Analysis led to the development of a test assessing English in the context of finance and accounting, comprising four components (Tests of Reading, Writing, Listening and Speaking). The test was set at levels B2 and C1 on the Common European Framework of Reference for Languages (CEFR) (Council of Europe 2001). Both Cambridge ESOL and ACCA were in agreement that CEF level B2 was an appropriate target language level for a finance or accounting professional.

Sample materials were developed for the four papers in order to conduct a trial on a representative target candidature sample. The papers were reviewed by subject experts at ACCA and changes made in the light of their comments. A number of specific research questions were then drawn up based on the Cambridge ESOL VRIP (Validity; Reliability; Impact; Practicality) framework.

The trial

The trial population was composed of 319 candidates studying for ACCA qualifications. Candidates came from a wide range of first language backgrounds and nationalities. In addition to trial versions of the Tests of Reading, Writing and Listening, candidates were also asked to provide information on their gender, age and study/work experience related to finance. The trial test population was also asked to complete a test of general English proficiency and asked for their opinions on a number of aspects of the test such as the authenticity of the texts featured in the test papers. Use of the General English Placement Test (GPT) showed that over half the sample was already at the target levels of CEFR B2/C1.

Trial results

What is the impact of work experience on performance?

One research question was to investigate whether there was a significant difference in scores between candidates with work experience in a finance/accounting environment and those without such work experience. As discussed earlier, there is an acceptance that specific purpose background knowledge is likely to influence and interact with specific language use. A preliminary investigation of the role of work experience on test performance was therefore undertaken in the trial.

Table 1 shows that candidates with work experience tended to score slightly lower in the General English Test than those without work experience. This difference was not significant (General English Test: $t= 1.237$, $df= 260$, $p= 0.217$, two-tailed). In the Reading, Writing and Listening components, candidates with work experience tended to score slightly higher than those without work experience. In Reading and Writing this difference was significant (Reading $t= -2.109$, $df= 286$, $p < 0.04$, two-tailed, Writing $t= -2.0$, $df= 212$, $p< 0.05$, two-tailed, Listening: $t= -0.788$, $df= 284$, $p =0.431$, two-tailed).

No definite conclusions on the effect of work experience and background knowledge on results can be drawn, as variables were not controlled for gender or first language as candidature numbers

Table 1: Work experience and scores in tests

Test	Have you worked in a finance-related area?	Mean score	Number of candidates
General English Placement Test	no	40.8	92
	yes	39.5	170
ICFE Reading	no	22.7	94
	yes	25.0	194
ICFE Listening	no	10.4	93
	yes	10.9	193
ICFE Writing	no	18.5	58
	yes	21.6	156

were too limited for this. However, it does suggest that background or subject knowledge, as expressed through work experience, may contribute to an increase in performance in ICFE in the Test of Writing and Test of Reading. When the examination becomes live, more data will be collected to enable investigation of this question in more detail.

Do the test tasks and topics have authenticity and face validity?

A candidate feedback questionnaire was given to the trial population for completion after they had taken the test papers. Course providers and stakeholders within the international finance community were also asked to complete feedback questionnaires giving their opinions of the papers. Authenticity in the questionnaires was defined as the type of texts and tasks that candidates or their colleagues may be expected to encounter in studies or work within an accounting/finance context. Respondents were asked to read statements concerning the test and to complete Likert scales from 1 to 5 indicating 1 if they strongly disagreed with the statement and 5 for strong agreement.

Opinion of test population in the trial

As may be seen in Figure 4, the majority of the test population in the trial considered the topics in the Reading component to be similar to those they need to read as part of their work or studies. Figure 5 shows that there was a similar perception of the texts and language included in the Reading paper.

Candidate responses concerning the authenticity of the Writing paper were similar. A smaller but significant majority also considered topics and language on the Listening paper to be authentic.

Opinion of stakeholders

The majority of respondents from the stakeholder group (just under 90% as illustrated in Figure 6) considered that the content of the proposed examination was appropriate and commented favourably on the texts, topics and tasks. Respondents described the trial test papers as including clear and straightforward questions within a range of task types, and highlighted the importance of the testing of all four skills.

Furthermore, as shown in Figure 7, a majority of respondents from this stakeholder group stated that their institution would give consideration to recognising ICFE as reliable evidence of English

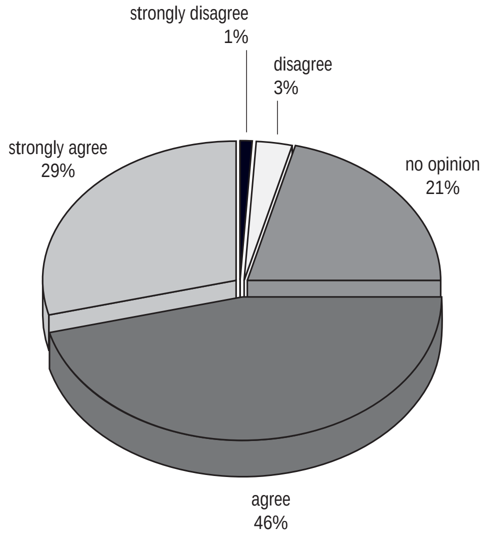


Figure 4: Response to statement – On the Reading component the topics are authentic

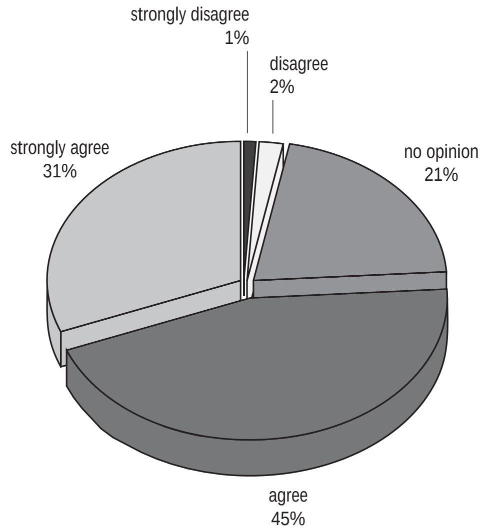


Figure 5: Response to statement – On the Reading component the texts and language are authentic

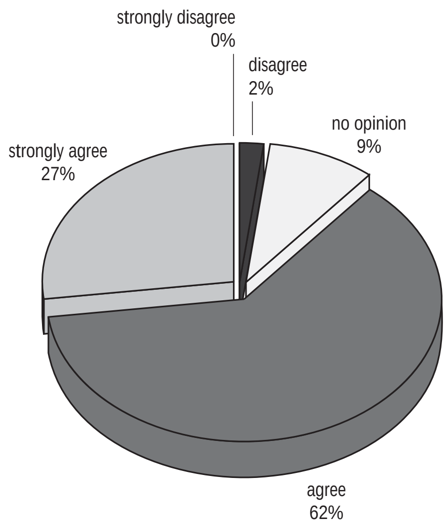


Figure 6: Response to statement – ICFE contains realistic texts, topics and tasks

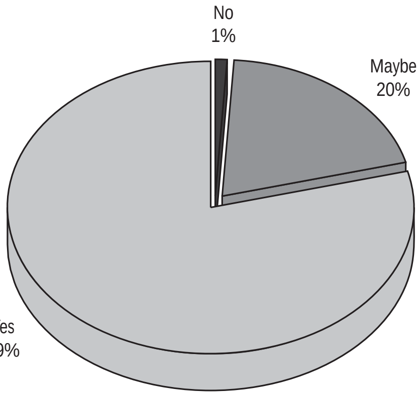


Figure 7: Response to statement – We would consider recognising ICFE as reliable evidence of English language skills

language skills. Potential uses given were varied and included the use of ICFE to promote trainee accountants and as a final target for specific courses.

Other research questions in the trial

It was also necessary to investigate aspects of ICFE not specific to tests of LSP. Candidates were asked to agree/disagree with statements concerning the time allowance for the completion of each component. In both Reading and Writing components, the majority of candidates (Reading 73%, Writing 71%) agreed or strongly agreed with the statement that there was enough time to complete all the tasks.

Cronbach's Alpha was used as a method of providing an index of the internal consistency of the Reading and Listening components in the trial. This is a measurement of how consistent the components are, for example, in awarding similar marks to candidates of the same ability. These were 0.89 and 0.79 respectively. These are similar to those found in Cambridge ESOL examinations such as BEC and FCE and indicate that the components contain enough discriminating tasks to be considered reliable.

Conclusion

This article has outlined some of the more pressing issues related to developing a test of LSP and has described the procedures adopted by Cambridge ESOL in the development stage of such a test, the International Certificate in Financial English.

The recognition of a continuum of specificity as outlined suggests that an index of the specificity of the test components, and tasks and texts within the components, would be a useful tool for investigating the construct validity of the test. O'Sullivan (2006) offers a methodology of describing the specificity of elements of an LSP test. This will be applied to components in ICFE to investigate whether any particular area of the test is more or less specific than others and then to see if this can be justified through Needs Analysis of the target language use situation and through work with subject experts such as ACCA.

The interaction of subject knowledge with language knowledge in ICFE requires monitoring. Data from live administrations will be used to examine how candidates with finance-related work experience perform in the components compared to those without this experience. A further area of future activity will be the building of a corpus of ICFE learner English using candidate responses to questions on the Writing paper. One application of this will be to explore candidates' use of finance-related vocabulary and terminology and to investigate whether this correlates significantly with performance level.

The issues described in the development of ICFE, namely, test specificity, authenticity of content, the relationship between background knowledge and language knowledge will be evaluated on a continuing basis by Cambridge ESOL in collaboration with its partner in this project, ACCA.

References and further reading

Alderson, J C (1988) New procedures for validating proficiency tests of ESP? Theory and practice, *Language Testing* 5, 220–232.

Bachman, L F and Palmer, A S (1996) *Language Testing in Practice*, Oxford: Oxford University Press.

Bhatia, V K (1993) *Analysing genre: language use in professional settings*, London: Longman.

Clapham, C (1996) *The development of IELTS: A study of the effect of background knowledge on reading comprehension*, Studies in Language Testing Vol. 4, Cambridge: UCLES/Cambridge University Press.

Council of Europe (2001) *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*, Cambridge: Cambridge University Press.

Davies, A (1990) *Principles of Language Testing*, Oxford: Blackwell.

Douglas, D (2000) *Assessing Languages for Specific Purposes*, Cambridge: Cambridge University Press.

Flowerdew, J and Wan, A (2006) Genre Analysis of tax computation letters: How and why tax accountants write the way they do, *English for Specific Purposes* 25, 133–153.

Grewcock, J (2006) *International Certificate in Financial English Viability Study*, Cambridge ESOL, internal report.

Kristeva, J (1980) *Desire in language: a semiotic approach to literature and art*, Oxford: Blackwell.

O'Sullivan, B (2006) *Issues in testing Business English: The revision of the Cambridge Business English Certificate Studies*, Studies in Language Testing Vol. 17, Cambridge: UCLES/Cambridge University Press.

Robinson, P (1991) *ESP Today: a practitioner's guide*, New York: Prentice Hall.

Skehan, P (1984) Issues in the testing of English for Specific Purposes, *Language Testing* 1, 202–220.

Swales, J M (1990) *Genre analysis: English in academic and research settings*, Cambridge: Cambridge University Press.

Widdowson, H (1983) *Learning purpose and language use*, Oxford: Oxford University Press.

Using the global legal community in the development of ILEC

DAVID CORKILL, ASSESSMENT AND OPERATIONS GROUP
MARTIN ROBINSON, ASSESSMENT AND OPERATIONS GROUP

Introduction

Cambridge ESOL is developing a number of tests of English for Specific Purposes (ESP). The first of these, the International Legal English Certificate (ILEC), is an exam for the global legal community to be used by legal professionals and law students as proof of their proficiency in English within the legal domain.

Analysis of TLU

Douglas makes the point that, for an ESP test to be valid, there must have been an analysis of the target language use situation:

> *[A specific purpose assessment is] one in which test content and methods are derived from an analysis of a specific purpose target language use situation (TLU), so that test tasks … are authentically representative of tasks in the target situation, allowing for an interaction between the test taker's language ability and specific purpose content knowledge, on one hand, and the test tasks on the other.* (Douglas 2000: 19)

Widdowson (1983) believes that such analysis has often been sketchy and based on little more than observational lists. Cambridge ESOL from the outset of the development of ILEC saw the need for significantly more that a one-off analysis. It was decided that the optimum analysis of the TLU would be one that was done in collaboration with the global legal community and one that was continuous from the first design of the test through to the ongoing production of live exam material.

Therefore, in moving away from tests of general English, it was necessary to augment established practices with an ongoing dialogue with content specialists, comprising members of the global legal community in the case of ILEC, to ensure that the test is fit for purpose and accessible to candidates with a range of experience and from different jurisdictions.

The design phase

In the development of the International Legal English Certificate, Cambridge ESOL consulted widely among the global legal community. A partnership was formed with TransLegal, Europe's leading firm of lawyer-linguists, and the area of international commercial law was identified as the one with the most transnational communication and, therefore, the area to target.

In order to conduct a viability study, there needed to be a test design and sample materials. The design of the pilot test was a result of dialogue between Cambridge ESOL, as assessment specialists, and content specialists from the legal community. Both fields of expertise were necessary here: content specialists were able to provide information on what lawyers do and Cambridge ESOL was able to analyse which tasks could be authentically

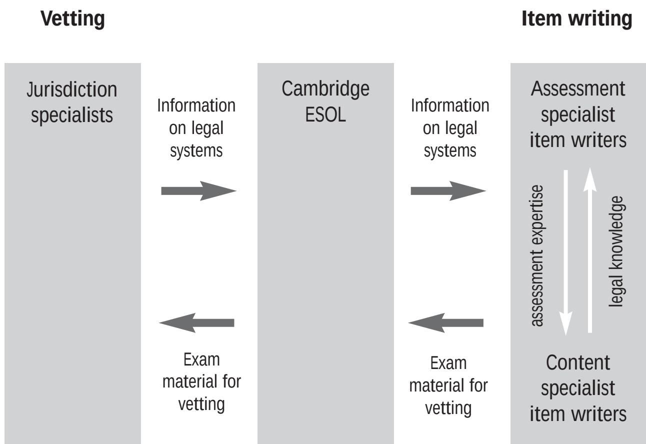


Figure 1: Model of interaction between Cambridge ESOL assessment specialists and legal content specialists

simulated in a testing situation in such a way that they could be used to assess language proficiency.

The viability study was administered to obtain feedback on the relevance, viability, structure, content and level of the proposed certificate from the major stakeholders in the legal community. These stakeholders were identified by both TransLegal and Cambridge ESOL as being suitably qualified to offer an opinion. They were also selected to represent a range of jurisdictions, both common law and civil law based.

The stakeholders contacted included:

- international law firms with overseas offices who recruit lawyers whose first language is not English
- local law firms in target markets for ILEC
- large companies with their own legal departments
- governments (e.g. Supreme Courts)
- university law faculties
- lawyer and Bar associations
- legal English training providers and language schools
- legal English tutors.

A pack was produced for ILEC including:

- background information on Cambridge ESOL and TransLegal
- information on the target candidature
- the content of the examination and what it would involve
- sample papers for all four skills tested
- information on the level of the test
- reasons to choose ILEC
- an outline of future plans for the development of the test.

In addition, two questionnaires were included, one for legal organisations and one for instructors of legal English. The feedback was positive, with 82% of respondents stating that there was a need for such an exam and 88% and 82% of law firms stating that the test material and the language skills tested, respectively, were authentic. More qualitative feedback came from the questionnaire responses:

> *I think the ILEC is a great project and will provide us with very helpful information in future recruitment procedures.* (Freshfields, Germany)

> *A certificate would be particularly convenient in the recruiting process, obviating the need for internal testing or guess work, and would assist in the determination of continuing English training needs for our staff.* (Advokatfirman Vinge KB, Sweden)

> *I do think this is valuable … we would support this kind of certificate and it would be useful.* (Linklaters, UK)

> *Overall I was impressed with the quality ... I believe there will be high interest from UK & US firms.* (European Young Bar Association, UK)

After further refinement to the test design, a trial test was administered and the feedback on the legal authenticity of the test was consistent with that generated by the pilot test. The trial feedback included questionnaire responses from participants, including both practising lawyers and law students. This trial also showed that the test could distinguish between those candidates with C1, B2 and below B2 levels of English language ability, according to the Common European Framework of Reference (CEFR).

Live test production phase

ILEC then moved onto a live paper production schedule and saw the first administration of the new exam in May 2006. The production of exam material is the responsibility of a Cambridge ESOL member of staff, the Subject Officer for the paper, and an external consultant, the Chair of the Item Writing team.

The role of the Chair is principally concerned with the technical aspects of writing the examination materials and ensuring that the item writers on the team are fully equipped to produce material to the best of their ability. The Chair ensures that tasks are appropriate in terms of topic, content and level and that they comply fully with the Specifications for the paper and Item Writer Guidelines. The Subject Officer is responsible for managing the production of the examination material through the various stages and ensuring that sufficient material is produced to the agreed schedule. Both the Chair and the Subject Officer also bring expertise to the partnership, beyond the particular paper, from their experience of teaching and assessment.

There are several specific issues involved with the production of ILEC papers:

- All item writers, Chairs and Subject Officers need to be informed on the areas of law that ILEC covers and, equally importantly, those that it does not.
- It is important that nuances of the law are not affected by any necessary editing of the texts.
- All of the tasks must be accessible to all of the candidates – both lawyers and law students, and regardless of the jurisdiction they work or study in.
- It is important that the scenarios and dialogues are authentic and that legal terminology is used appropriately.
- With specific legal collocations, all possible keys should be included on a mark scheme and no distractors should be possible keys.
- All papers need to be jurisdiction neutral, especially in areas relating to common versus civil law.
- The specific nature of the TLU domain means that there is an issue with what Swales (1996) calls occluded texts – texts not readily accessible to those outside the legal profession because of their confidential nature.

To address the first issue, TransLegal provided training for the Subject Officers and Chairs involved with ILEC. This was based on their own knowledge of the issues and also in response to questions from the trainees that had been elicited in advance. This knowledge was then cascaded down to the item writing teams through individual paper training.

To address the paper-specific issues, content specialists with experience of both the law and EFL teaching were recruited onto the item writing teams following initial training. These specialists vet material prior to pre-editing and attend editing meetings to ensure that tasks fall within the ILEC areas of law. A synergy thereby occurs with the new legal item writers learning the craft of item writing from the established item writers and the established item writers learning more about the law appropriate to ILEC from the legal item writers.

As is to be expected given that the legal item writers are UK based, their experience of the law is in a common law jurisdiction. To plug this gap, TransLegal set up a panel of jurisdiction specialists comprising Swedish, UK, US, German, French and Spanish legal experts, to vet material for its jurisdiction neutrality. This continuing relationship with those within the legal community also means that occluded texts are available to the item writing teams (see Figure 1).

Conclusion

The development of ILEC has seen a collaboration between assessment specialists and legal content specialists, with each bringing their expertise to the process. This has meant that the test has grown out of the community that it is intended for and has tasks that authentically simulate the target language situation while successfully assessing candidates' language proficiency.

This collaboration is built in to the test production process, ensuring that the dialogue between the assessment and legal communities continues to inform the design and content of ILEC.

References and further reading

Douglas, D (2000) *Assessing Languages for Specific Purposes*, Cambridge: Cambridge University Press.

Swales, J (1996) Occluded genres in the academy: the case of the submission letter, in Ventola, E and Mauranen, A (Eds) *Academic Writing: Intercultural and Textual Issues*, Amsterdam: John Benjamins, 45–58.

Widdowson, H (1983) *Learning purpose and language use*, Oxford:

Developing the Cambridge ESOL Teacher Portfolio

CLARE MITCHELL CROW, ESOL PROJECTS OFFICE
CLARE HARRISON, ASSESSMENT AND OPERATIONS GROUP

Introduction

This article reports on the development of a new service from Cambridge ESOL called the Teacher Portfolio. The Teacher Portfolio is a web-based tool that offers teachers a secure personal space to consolidate and document their teaching career. The benefits of this service are that it offers teachers undertaking or teaching our exams a means to record their qualifications, professional development, reflections on their teaching and work experience.

The introduction of the Teaching Knowledge Test (TKT) in 2005 was the catalyst for the development of a Cambridge ESOL Teacher Portfolio. TKT aims to increase teachers' confidence and enhance their job prospects by focusing on the core teaching knowledge needed by teachers of primary, secondary or adult learners, anywhere in the world (see the TKT website for further information: www.CambridgeESOL.org/TKT). To offer an electronic tool in which candidates are able to document and reflect on their teaching practice, while not a part of the assessment for TKT, was seen as an ideal adjunct. During the development phase, it became apparent that all teachers could benefit from the use of the service; therefore there are no restrictions on who is eligible to register and use the Teacher Portfolio.

Electronic portfolios

Electronic portfolios are currently used on an international scale in a wide variety of educational and vocational contexts. Electronic portfolios may be defined as a collection of authentic and diverse digitised artefacts, including demonstrations, resources, reflections and accomplishments that represent an individual's learning and work over time (Lorenzo and Ittelson 2005). The types of portfolios described in this article focus on those that have influenced the development of the Cambridge ESOL Teacher Portfolio.

Assessment portfolios

An assessment portfolio allows the owner to input or upload examples of their work for viewing by others for assessment purposes. They are typically structured to match the areas of assessment in a subject or course. The information may be organised in date order, or by assessment criteria. These portfolios always offer the opportunity to share examples of work with others, (i.e. to allow a teacher to view a portfolio for assessment purposes).

Showcase portfolios

A showcase portfolio allows the owner to choose (the best) examples of their work for viewing. As the structure of a showcase portfolio is typically open for the user to determine, owners generally develop their portfolios around the inclusion of key examples of their work, study or area of interest. These portfolios are commonly formatted in date order, allowing for the display of work over time. Visually, these portfolios tend to be unique, allowing the owner to control the style, colours and overall presentation. Showcase portfolios are typically shared with others, for example, when the owner is applying for a job, or wanting to highlight work to a colleague or show potential employers or educational institutions.

Development portfolios

A development portfolio allows the owner to monitor and plan their own development. They are generally offered when enrolled on a course or when attending a college or university. The structure of the development portfolio allows the owner to record their development in relation to specific criteria. These portfolios are typically shared with other people, to ensure development can be tracked, reviewed or commented on.

Reflective portfolios

A reflective portfolio is one that allows the owner to review their own development. The structure of a reflective portfolio typically has criteria relating to the reason for maintaining the portfolio (for example, study or work) and allows for the owner to add their own reflective commentary relating to their experiences. The common format is to order the work in chronological order. These portfolios are typically shared when the owner is applying for a job, or wanting to highlight work to others.

Hybrid portfolios

The majority of electronic portfolios available today combine two or more of the above areas. For example, a portfolio may combine aspects of a reflective portfolio with a development portfolio. The Cambridge ESOL Teacher Portfolio combines aspects of reflective, showcase and developmental portfolio styles.

Teacher Portfolios

For practising or prospective teachers, a portfolio offers a means of storing, organising and sharing materials and information related to their teaching in a way which encourages reflection and professional development. A teacher portfolio can be used to store material teachers and their learners produce. This could include:

- a statement of their teaching philosophy
- a description of their teaching responsibilities and context
- class profiles
- lesson plans and materials
- examples of learners' work
- learners' feedback
- self-reflections on their teaching
- a list of personal goals for improving teaching and enhancing skills.

Feedback the teacher has been given after lesson observations or a record of training courses attended can also form part of a teacher portfolio, as well as official documentation, such as certificates, references and employment records.

Teacher portfolios may allow a teacher to display their work publicly. For example, training course providers may make use of documents collected in a teacher portfolio to formally assess the performance of trainee teachers and/or to monitor their development during a course. Additionally, teachers applying for a new position can make use of a teacher portfolio as an expanded curriculum vitae to showcase their achievements, experience and beliefs to employers.

Naturally, the very act of selecting and organising material to be included in a portfolio encourages teachers to examine their own teaching and beliefs. Reflective teacher portfolios can have the specific purpose of fostering an analytical approach to one's own teaching by providing a framework for noting experiences and impressions. These notes can be used as a basis for developing further as a teacher, and may be shared with others as part of the development process.

In the UK, the Department for Education and Skills (DfES) has adopted the term Professional Development Record (PDR), rather than portfolio. They view a PDR as a confidential and voluntary collection of material that records and reflects a teacher's work. It is a way of using past experiences and present activities to demonstrate and reflect on skills learnt in order to identify future learning needs and priorities, and to inform and plan prospective development. It provides a mechanism for teachers to think about their practice in a planned and systematic way (DfES 2001).

Figure 1: Contents of the Cambridge ESOL Teacher Portfolio

Profile

- Summary profile
- Language profile
- Access details

Teaching

- Teaching qualifications
- Courses I have taught
- Teaching & learning samples
- Teaching action plan

Reflections

- Reflective diary
- Beliefs about teaching & learning

Achievements

- Employment record
- Training courses attended
- Certificates in language skills

Assets (the ability to upload files to demonstrate your work)

Referees

Portfolio options (the ability to print out a record)

Contact us

Regardless of the name, there is little argument that electronic portfolios encourage personal reflection and often involve the exchange of ideas and feedback (Lorenzo and Ittelson 2005).

Structure of the Cambridge ESOL Teacher Portfolio

The structure and content of the Teacher Portfolio was developed through a commissioning process. The authors of the content reviewed existing teacher portfolios from various countries and institutions. This review provided us with an insight into the range of portfolios available, and what type of content and structure would suit the Cambridge ESOL context.

As previously mentioned, the Cambridge ESOL Teacher Portfolio has a structure that allows for demonstration of one's development, reflection and also offers the opportunity to showcase examples of work (if chosen by the owner). Figure 1 outlines the overall structure of the Cambridge ESOL Teacher Portfolio.

Benefits of the Cambridge ESOL Teacher Portfolio

There are a number of benefits of the Cambridge ESOL Teacher Portfolio, namely:

- Ease of access – the ability to access the portfolio on any web browser, at any time
- Secure storage – access to a portfolio is granted after registration, and then logged into with an email address and password
- Portability – no requirement to print out sections of the portfolio, as the owner is able to access it anywhere an internet connection is available
- Reflection and planning – ease of storage and organisation of reflections and action plans, as the owner is able to keep a wide range of documents together in one place.

Conclusion

It is hoped that the Teacher Portfolio will encourage candidates and teachers to record their learning and teaching developments electronically, thus enabling them to become reflective practitioners through analysis of their teaching and to upload examples of their work or others' work that has impacted on their teaching. The Cambridge ESOL Teacher Portfolio was launched in May 2006. For further information visit www.teacherportfolio.cambridgeesol.org

References and further reading

DfES (2001) *Helping you Develop: Guidance on Producing a Professional Development Record*, Department for Education and Skills www.teachernet.gov.uk/_doc/840/Text-Help_to_Dev_T.pdf

Lorenzo, G and Ittelson, J (2005) *An Overview of E-Portfolios*, EDUCAUSE Learning Initiative Paper 1, July 2005 www.educause.edu/ir/library/pdf/ELI3001.pdf

Profile of Skills for Life candidature

NADEŽDA NOVAKOVIĆ, RESEARCH AND VALIDATION GROUP

Introduction

Cambridge ESOL's Certificates in Skills for Life (SfL) are part of the Government's strategy for improving adult literacy, numeracy and ESOL skills in England. These certificates provide assessment of English for Speakers of Other Languages, designed around the standards for adult literacy and the Adult ESOL Core Curriculum.

Cambridge ESOL Skills for Life certificates were launched in March 2005 and are available at Entry 1, Entry 2, Entry 3, Level 1 and Level 2 on the National Qualifications Framework. At each level, three separate modes are available: Reading, Writing, and Speaking and Listening. In order to reflect the fact that learners may have varying levels of ability in different skills, candidates may be assessed in different modes at different levels in the same exam session. They may also choose to enter any combination of the three modes at the same level, or enter for single modes, one by one, as their skills develop.

Candidature

The Skills for Life tests are designed for adult learners whose first language is not English, and who are living in or intending to settle or stay on a long-term basis in England. These ESOL learners may include refugees or asylum seekers, migrant workers, people from settled communities, and partners or spouses of people who are settled in this country for a number of years. The range of materials and task types reflects the diversity of SfL candidates' educational and employment backgrounds, their aspirations, needs, literacy levels and language learning skills.

This article provides information on the more than 100,000 candidates who have taken SfL tests since their launch in March 2005. The data presented here have been supplied by candidates themselves who are asked, at each SfL Writing examination session, to fill in a Candidate Information Sheet (CIS), which provides information on their age, gender, their first (and second) language, and number of years they have studied English. The answers provided on CIS sheets do not affect candidates' results in any way, and all the individual personal data remain confidential.

Age and gender

Nearly half of all SfL candidates are between 16 and 20 years of age (48.3%). This group is closely followed by candidates aged between 21 and 30 years (33.8%). The remaining candidates are either between 31 and 50 years old (14.3%) or older than 51 (3.6%) (see Figure 1). This is true of all five SfL levels, i.e. candidates in the age group 16–20 represent the highest percentage of candidates on all levels from Entry 1 to Level 2.

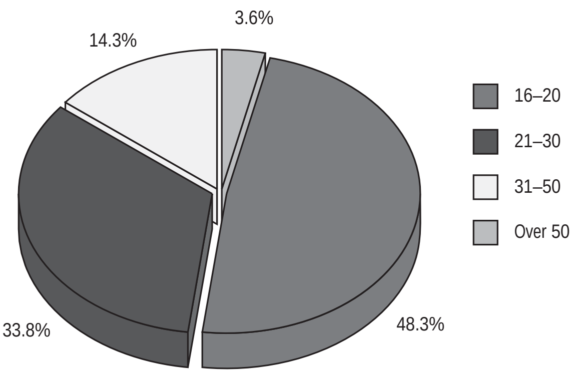


Figure 1: Age of SfL candidates

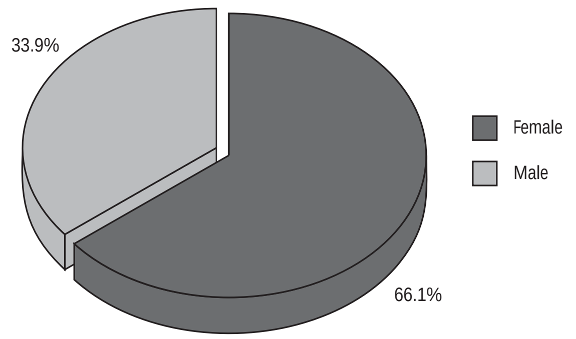


Figure 2: Gender of SfL candidates

As far as the gender of SfL candidates is concerned, the majority of test takers are female (66.1%), while 33.9% are male (Figure 2). This holds true for all five SfL levels.

First language

On CIS forms, candidates provide answers to questions relating to their first and second languages. A total of 55 languages were chosen by candidates as their first language. Furthermore, four percent of the candidates are bilingual. The majority of candidates who took SfL tests are first language Polish speakers, followed by Spanish, Farsi, French and Arabic speakers. Figure 3 shows the ten languages most frequently chosen by test takers as their mother tongue.

The Polish speaking candidates also represent the most numerous group of test takers on all five levels.

Years of learning English as a foreign language

SfL candidates are asked to provide information on how long they have been learning English prior to taking the test. At the time of sitting the exam, most candidates had spent between 6 months and

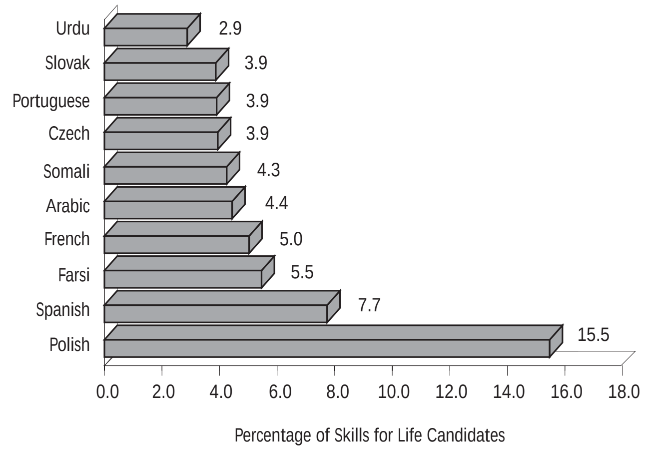


Figure 3: Most fequently chosen L1 by SfL candidates

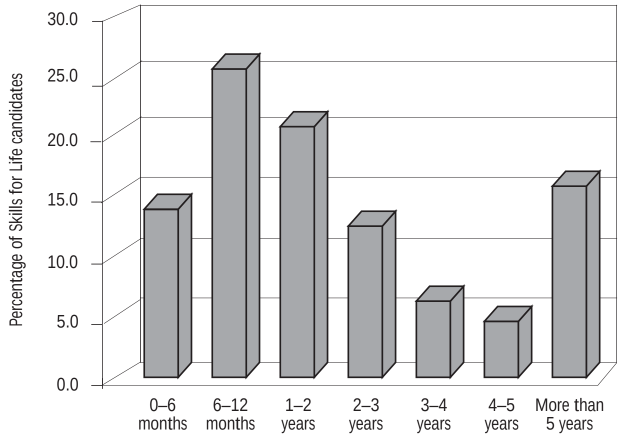


Figure 4: Time spent studying English by SfL candidates

2 years learning English (46.4%). Candidates who were learning English between 2 and 5 years represent 23.8% of the entire candidature. These are followed by candidates who learnt English for more than five years (15.5%) and those who had less than 6 months of experience as English language learners (14.4%). Figure 4 illustrates the number of months or years that SfL candidates had spent learning English before taking one of the SfL tests.

Conclusion

In brief, SfL candidates tend to be between 16 and 20 years of age, the majority are females and they have spent between six months and two years studying English prior to taking the exam. Cambridge ESOL will continue to monitor the profile of SfL candidates, noting any changes in the test user population that may lead to further test developments and revisions.

The impact of proficiency-level on conversational styles in paired speaking tests

FUMIYO NAKATSUHARA, PHD CANDIDATE, ESSEX UNIVERSITY

Introduction

As studies of oral testing have suggested desirability for paired testing, a format where non-native test takers are paired and examined together has recently become a popular tool to assess oral interactional communication ability. The pairing of test takers is also currently the standard practice in all the Cambridge Main Suite Speaking tests.

This format, however, has attracted some criticism concerning how the pairing of test takers should appropriately be conducted, as paired interlocutor characteristics could be significant variables influencing one's performance. Accordingly, studies have investigated the effect on performance of interlocutor variables such as gender, acquaintanceship, personality, cultural background and language proficiency. Among them, examination boards and teachers, interestingly, appear to take the language proficiency variable into most serious consideration (e.g. Foot 1999), despite the fact that there have been only a few attempts to explore the effect (Csepes 2002, Iwashita 1996, Norton 2005), and thus far, only little is understood. Therefore, the present study addresses the following research questions, as it seems especially essential to investigate the impact of proficiency-levels on the nature (and quality) of paired testing discourse. This is because while candidates' symmetrical contribution to the conversation is one of the greatest grounds for validating this format, if high-level candidates take a similar role to that which native speakers (NS) would take in traditional interview tests, asymmetrical (non-conversational) discourse could be generated, with higher-level students controlling the conversation.

1. Are conversational styles of dyads different between same proficiency-level pairs (SPL pairs) and different proficiency-level pairs (DPL pairs)?
2. Are dyadic interactions with different ability speakers asymmetrical? If so, how are they asymmetrical? To what extent are they asymmetrical?

Research participants

The 24 participants of this study are international undergraduate/graduate students at Essex University in the UK in 2004. These students came from a wide range of L1 backgrounds. Their length of stay in the UK varied from 6 months to 4 years and 7 months, and over two-thirds of them were female (male: female=7:17).

The subjects were divided into two groups (advanced/ intermediate) according to their proficiency level. Their proficiency-level was principally decided on the results of the commercial tests which each student had individually taken before. However, as some of the subjects' records were relatively old and the scores of speaking ability alone were not always available, all their speaking proficiency-levels were confirmed by three raters who examined their speech samples on the Cambridge Common Scale for Speaking. For the purpose of this study, the CPE and both satisfactory and strong passes on the CAE are called 'advanced', while a minimum adequate performance level of the CAE and the transitional level between the FCE and the CAE are called 'intermediate'. As a result, 10 subjects were classified as advanced, and 14 as intermediate.

Data collection

All of the subjects attended two sessions in which they were required to perform assessment tasks with two different interlocutors so that more experimental cases could be obtained from the limited numbers of subjects. Matching of pairs was randomly conducted, and this generated 12 sessions by DPL test takers (N=24) and 12 sessions by the SPL test takers (8 sessions by intermediate pairs: N=16; 4 sessions by advanced pairs: N=8).

The paired task chosen for this study was the Two-way Collaborative Task in the CAE Speaking test, which involves a problem-solving discussion. After a two-minute opening conversation for warming up, the two subjects were provided with visual and written prompts of a problem-solving task. Four alternative prompts with parallel structures were prepared:

1. Rank seven jobs in a hotel according to the salary you want to pay for them (modified from UCLES 2001b: 52)
2. Rank six things you think thirty-year-old single men might like to have (Lukey-Coutsocostas and Dalmaris 1996: 43)
3. Select two events which have had the greatest influence on people's lives (UCLES 2001a: 56–57)
4. Select three things you think are useful in crossing the Arctic on foot (modified from UCLES 2000).

In the first part of each prompt, students were asked to give opinions about several items shown on the card, and then the second part required them to discuss together to make an agreed decision.

Data analysis

All sessions were video-taped and transcribed following Conversation Analysis (CA) conventions (Atkinson and Heritage 1984). Firstly, the data were quantitatively analysed with a framework based mostly on Young and Milanovic (1992), Young (1995) and Kormos (1999). Secondly, a qualitative analysis with CA was conducted for interpreting and elaborating the statistical results.

Measures of conversational styles

Although 'conversation' could be defined in a number of ways, van Lier (1989: 495) suggests that conversation involves the following properties: 'face-to-face interaction, unplannedness (locally assembled), unpredictability of sequence and outcome, potentially equal distribution of rights and duties in talk, and manifestation of features of reactiveness and mutual contingency'. Among these properties, the analysis here concentrates only on patterns of *interactional contingency*, *goal-orientation* and *quantitative dominance*.

Firstly, *interactional contingency* refers to 'a property of adjacent turns in dialogue in which the topic of the preceding turn is coreferential with the topic of the following turn' (Young and Milanovic 1992: 405). If an utterance is contingent upon its previous utterance by the other interlocutor, the relationship of the two utterances can be called 'reactive', and such 'reactiveness' is likely to be the key to successful topical continuity across conversational turns. Therefore, the measure of the contingency here is the proportion of topics initiated by one party that are ratified and become the topics of the subsequent turns by the other party (expressed as a percentage). A topic is regarded as ratified if the topic initiation of Participant A is followed by at least one 't-unit' (one clause plus any subordinate clauses) by Participant B in which the same topic is continued.

Secondly, *goal-orientation* involves 'the speakers' attempts to realize certain internal goals or plans through the interaction' (Young and Milanovic 1992: 405). Since conversation is related to 'unplannedness (locally assembled)' but also to 'equal distribution of rights and duties in talk', both interlocutors should ideally have an equal amount of initiative in this task-based, goal-oriented activity. Thus, the measure of *goal-orientation* is the proportion of topics initiated by one party relative to the other party (expressed as a percentage).

The third element, *quantitative dominance*, is also connected to 'equal distribution of rights and duties in talk'. Although there seem multi-dimensional means for conversational dominance (Itakura 2001), this analysis of dominance here focuses only on *quantitative dominance*, which refers to the level of contribution by the amount of talk (*ibid*. 1870). As an indicator of *quantitative dominance*, the measure is the proportion of words spoken by a speaker relative to the other speaker (expressed as a percentage).

Finally, since topic is the key concept in the above measures, a framework of analysis for identifying topics is essential. Topics were identified by finding their boundaries, which Maynard (1980) calls 'topic shift'. Following several conversation analysis studies (e.g. Button and Casey 1984, Hobbs 1990, Jefferson 1984, Maynard 1980, Reichman 1990) and considering particular features of the given interaction, I divided each set of interactions into discrete topics by means of identifying topic shift with both functional and structural criteria which often co-occurred.

The criteria employed are:

- Explicit boundary markers such as "alright", "so", "okay", "then", etc
- Topic initial elicitors like "What do you think?"
- Framing moves such as "Let's say what's the most difficult thing?"
- Long unfilled pauses
- Introduction of new information, sometimes with gestures
- Rounding off, for example, by summarising ideas that have so far been discussed or by providing reasons for terminating a topic.

Results

The first section below deals with the statistical results, describing the outputs for topic ratification (the measure of interactional contingency), topic initiation (the measure of goal-orientation) and the amount of talk (the measure of quantitative dominance). The second section presents the CA results for the same three elements.

Results of quantitative analysis

ANOVA was firstly conducted to see the effect of task difference for all the cases. As no significant main effect or interactional effect with task types were found, results of Independent-Samples T-tests are only presented here.

Topic ratification

SPL pairs were more reactive, ratifying 91.62% topics initiated by the other partner than DPL pairs which ratified 73.66% topics.[1] Moreover, DPL pairs ratified topics 23.07 % more asymmetrically than SPL pairs. While SPL pairs showed 12.58% absolute mean difference, DPL pairs showed 35.65% absolute mean difference.[2] However, the advanced/intermediate speakers in DPL pairs did not generate either direction of asymmetry.[3]

Topic initiation

When advanced and intermediate speakers were paired, advanced speakers tended to initiate 37.02% more topics than intermediate speakers.[4] SPL pairs, nevertheless, also had a similar (non-significantly different) level of variability when the proportion of one party's topic initiation was compared with that of the other partner.[5]

Amount of talk

When advanced and intermediate candidates were paired, the advanced candidates tended to dominate the conversational floor about 20% more than the intermediate candidates.[6] SPL groups, nonetheless, also had a similar (non-significantly different) level of variability when the proportion of one party's amount of talk was compared with that of the other partner.[7]

Results of qualitative analysis and interpretation of the quantitative results

Interactional contingency

Firstly, as found above, SPL pairs tended to be more reactive by ratifying more topics than DPL pairs. Reasons, nevertheless, for accepting topics seemed to vary between advanced-advanced pairs and intermediate-intermediate pairs. Advanced pairs were likely to expand topics by means of accepting the other's topics. For example, when advanced learner *L* in line 3 of extract 1 ratified the topic concerning "the first man on the moon" provided by *J* in line 1, she developed the topic by mentioning "the controversy surrounding the fact of landing on the moon", which generated further reactive interaction and topic continuation.

Extract 1: SPL pair (J and L: advanced);
Topic: Influential events

1→	*J:*	I agree, what about the first man on the moon, nineteen [sixty-nine?
2	*L:*	[.hhhh
3→	*L:*	I don't kno::w, because uh uh I heard that even some people say that
4		that is not true. That- Hu hu[h
5	*J:*	[How? You mean no one
6		landed on the moon [yet?
7	*L:*	[Yeah, that is just the event someone make (.)
8		made up that was eh:: so I don't know.
9	*J:*	So you don't think [that that is real?
10	*L:*	[Huh huh huh

On the other hand, intermediate pairs tended to ratify the other's topic due to the necessity of negotiating meanings to have mutual understanding. In line 5 of extract 2, although intermediate speaker *T* started to ratify the topic about "substitutability of instruments" initiated by *S* in line 1, the information in line 5 and following lines did not expand on what S gave in line 1. Rather, *T* just confirmed her understanding about the preceding content, and in fact the topic lasted over ten more lines consisting of only such negotiation.

Extract 2: SPL pair (S and T: intermediate);
Topic: Influential events

1→	*S:*	Hhh .hhh Bu- but if- if we didn't have a phone,
2	*T:*	uh huh?
3	*S:*	we can we could write and we could communicate with other people.
4		It's not fast (.5), but we (.) could communicate
5→	*T:*	Eh old time, just like we- our ancestor did. Just [like that.
6	*S:*	[Uh, Yes, yes.

Secondly, concerning the non-directional asymmetry of ratification in DPL pairs, two different reasons could be suggested.

1. SPL pairs: N=24, Std=16.16; DPL pairs: N=24, Std=25.47; t(44)=2.88, Mean difference=17.96, p=.006
2. SPL pairs: N=12, Std=14.94; DPL pairs: N=10, Std =19.01; t(20)=-3.189, Mean difference=-23.07, p=.005
3. Adv: N=10, Mean=65.24, Std =39.42; Int: N=12, Mean=75.68, Std =17.37; t(20)=-.829, Mean difference=-10.44, p=.417
4. Adv: N=12, Mean=68.51, Std =24.97; Int: N=12, Mean=31.49, Std =24.97; t(22)=3.63, Mean difference=37.02, p=.001
5. SPL pairs: N=12, Mean=37.43, Std =21.40; DPL pairs: N=12, Mean=51.37, Std =33.32; t(22)=-1.22, Mean difference=-13.94, p=.235
6. Adv: N=12, Mean=59.76, Std =9.04; Int: N=12, Mean=40.24, Std =9.035; t(22)=5.29, Mean difference=19.51, p<.001
7. SPL pairs: N=12, Mean=21.48, Std =14.70; DPL pairs: N=12, Mean=23.17, Std =12.50; t(22)=-3.03, Mean difference=-1.68, p=.765

One reason is that some intermediate subjects wished to ratify topics which were introduced by their higher-proficiency partner, but they failed to do so due to their insufficient language ability. Alternatively, some advanced learners seemed purposefully less reactive occasionally, when they were not interested in the topic introduced by their lower-proficiency partners. Nonetheless, it is worth noticing that the asymmetry in DPL pairs might not actually be perceived as asymmetrical, compared to the degree of asymmetry in examiner-candidate interaction in interview tests; according to Kormos (1999: 172–173) who measured ratification in the same way as this research, while the candidates accepted 99% of the examiners' topics, the examiners merely ratified 52% of the total number of topics introduced by the examinees. In extract 3, after abandoning a pianist topic, the next topic was introduced by *W*.

Extract 3: DPL pair (W: advanced, X: intermediate); Topic: Hotel staff

1→ *X:* I think (.) uhm::: it's a bit difficult to find such a pianist
2 *W:* Uh huh
3 *X:* maybe e- it, you know, we need to looking for a professional pianist
4 for this moment, so
5 *W:* Yeah
6 *X:* Yeah. We need to pay more money than other people.
7→ *W:* Yes, So basically if you need a <u>cleaner</u>=

Goal-orientation

CA data confirmed that advanced speakers tended to be more goal-oriented by initiating more topics than intermediate speakers. Furthermore, it was found that the advanced member of the DPL pairs also helped their lower-proficiency partners' topic initiation. For example, intermediate speaker *N* in extract 4 tried to introduce a new topic about "contrasting necessary things and extra things". However, because he could not produce the relevant vocabulary ("extra") to his idea, *M* provided him with the word so that *N* could complete his topic initiation.

Extract 4: DPL pair (M: advanced, N: intermediate); Topic: What 30-year-old men want

1→ *N:* but .hh think about holiday, expensive car, it's kind of (.), how's it?
2 (.5)
3 *N:* [Not not
4→ *M:* [Extra, extra [( )
5 *N:* [Yeh, extra things, yeh:: so: (.) should be lower ranked,
6 eh:: compared to family or house. What do you think?

Another type of assistance was where advanced subjects sometimes gave their intermediate partners a slot where the intermediate partners could actually initiate a new topic. In lines 3 and 4 in extract 5, advanced speaker *F* withheld her topic initiation, and she uttered a filler "Eh:" and implied a need for a new topic by a discourse marker "OK" followed by another filler "and:n:". Nevertheless, in this example, since *E* did not take the opportunity to introduce a topic, the failure was displayed as unnatural one-second silence in line 4, and as a result *F* initiated a new topic in line 5.

Extract 5: DPL pair (F: advanced, E: intermediate); Topic: Crossing the Arctic

1 *F:* eh >I totally agree with [you<. huh
2 *E:* [Uh
3→ *F:* Eh:, OK, and:n:
4→ (1.0)
5→ *F:* So thi[s one, this one and this
6 *E:* [yeh this one, this

Both of these supportive behaviours correspond to accommodation strategies which occur in interview tests (Lazaraton 1996, Ross and Berwick 1992). Such supportive practices by advanced speakers seemed to encourage less proficient partners to control more topics, and the interaction produced became more collaborative than it would have been without any interlocutor support. Thus, some advanced learners were more goal-oriented not only for completion of the paired task, but also for its collaborative completion. This may be greatly important in terms of scores as well. This is because, according to Galaczi (2004), who researched the paired-task part of the FCE tests, when paired candidates were collaborative, higher marks tended to be given on the Interactive Communication scale.

This speech accommodation also seems to explain why the above quantitative analysis did not find any significant difference in the level of asymmetry of topic initiation between DPL and SPL pairs. Since some advanced candidates attempted to give their lower proficiency partners more initiative, it minimised the gap in unbalanced topic-initiation to a similar variability level of the SPL pairs.

Quantitative dominance

In DPL pairs, the lower proficiency member could sometimes play rather a passive role, especially when advanced learners spoke much faster than them. Moreover, it seems that advanced learners were better at personalising the given topic, and dominate some conversational space with their own stories. In line 3 of extract 6, when "the difficulty of being a porter" was being talked about, *W* integrated her own experience on the previous day.

Extract 6: DPL pair (W: advanced, X: intermediate); Topic: Hotel staff

1 *X:* So, (.) uh, porter? Eh: sometimes, porter need to carry
2 a lot of bagg[age.
3→ *W:* [Eh, The porter here would never carries anything, in
the university
4 *W:* You have to be your own port[er.
5 *X:* [Ah yeah. Huh hah hah hah
6 *W:* I felted that yesterday, when moving out. Oh my god.
7 *X:* .hh I see. Eh:
8 *W:* Although I had three friends to help me out, but <u>still</u> [it was very
difficult.
9 *X:* [uh:
10 *X:* Yeah yes, I see

Furthermore, there seemed to be another accommodative strategy where advanced speakers adjusted the amount of talk according to their interlocutors. Some advanced subjects who were talkative with talkative partners, spoke a little less to the intermediate, reticent speakers, and encouraged the partners to speak more. Therefore, their adjustment contributed to relatively equal quantitative dominance from both parties. In this sense, a similar degree of variability in the amount of talk between SPL and DPL pairs is accounted for by the identical reasons discussed for *goal-orientation*.

Discussion

The results of this study highlight several interesting features of NNS-NNS interactions related to the relationship between proficiency and conversational styles in paired tests. The observed discourse mainly showed evidence for the following three properties. The first two paragraphs respectively answer the two research questions of this study. The third paragraph is greatly related to the nature of paired testing interactions and thus to both the research questions.

Firstly, while SPL pairs may be recognised as having slightly more 'symmetrical contingency' because of the balanced reactiveness, when the measure of *goal-orientation* and *quantitative dominance* are focused upon, the data suggest that both of the pairs do not significantly differ. All in all, there are many more similarities than differences in conversational styles between SPL and DPL pairs.

Secondly, it is worth noticing that there are some asymmetrical interactional characteristics which could be drawn from proficiency differences in DPL pairs. Advanced speakers tend to be more goal-oriented and more quantitatively dominant. Nevertheless, a similar degree of asymmetry was also observed in SPL pairs, which indicates that although proficiency difference does have some influence on the conversational features of candidates in DPL pairs, the impact may not be so problematic.

Thirdly, the analysis confirms several types of conversational accommodation in paired testing situations. Such accommodation has contributed not only to assist the lower-proficiency level interlocutor but also to make their interaction more conversational and more collaborative. That is, advanced learners were more goal-oriented towards collaborative completion of the task.

Conclusion

This study has revealed that whether students are paired with SPL or DPL partners, they are likely to obtain identical conversational styles, and the pairing of students with different language levels may not be as problematic as anticipated. This finding should encourage us to continue to employ the paired format as a useful tool to assess communicative language ability. Lastly, a further interesting extension of the present study would be to involve rating scales. The existing rating scale could be refined and validated based on the major interactional characteristics of conversational styles and accommodation strategies. For example, based on the findings of this study, we could produce descriptors about whether candidates are able to:

- be reactive to, or contribute to, the topic development/ expansion
- be goal-oriented by initiating topics to achieve the goal of the task
- assist the partner in order to complete tasks more collaboratively
- keep the conversational floor while personalising the given topic.

Findings of future research along these lines should be beneficial to provide a better understanding of the whole procedure of paired tests, and it may soon be greatly needed since paired formats are likely to be more and more commonly employed as a desirable tool in the assessment of oral proficiency.

References and further reading

Atkinson, J M and Heritage, J (Eds) (1984) *Structures of Social Action*, Cambridge, New York: Cambridge University Press.

Button, G and Casey, N (1984) Generating topic: the use of topic initial elicitors, in Atkinson, J M and Heritage, J (Eds) *Structures of Social Action*, Cambridge, New York: Cambridge University Press, 167–190.

Csepes, I (2002) Is testing speaking in pairs disadvantageous for students? A quantitative study of partner effects on oral test scores, *novELTy*, 9(1).

Foot, M C (1999) Relaxing in Pairs, *ELT Journal*, 53(1), 36–41.

Galaczi, E (2004) *Peer-peer Interaction in a Paired Speaking Test: The Case of the First Certificate in English*, Unpublished PhD Dissertation, Columbia University, NY.

Hobbs, J (1990) Topic Drift, in Dorval, B (Ed.) *Conversational Organization and Its Development*, Norwood, NJ: Ablex, 3–22.

Itakura, H (2001) Describing Conversational Dominance, *Journal of Pragmatics*, 33, 1859–1880.

Iwashita, N (1996) The Validity of the Paired Interview in Oral Performance Assessment, *Melbourne Papers in Language Testing*, 5(2), 51–65.

Jefferson, G (1984) On Stepwise Transition From Talk About a Trouble to Inappropriately Next-position Matters, in Atkinson, J M and Heritage, J (Eds) *Structures of Social Action*, Cambridge, New York: Cambridge University Press, 191–222.

Kormos, J (1999) Simulating Conversations in Oral-proficiency Assessment: A Conversation Analysis of Role Plays and Non-scripted Interviews in Language Exams, *Language Testing*, 16(2),163–188.

Lazaraton, A (1996) Interlocutor Support in Oral Proficiency Interviews: The Case of CASE, *Language Testing*, 13, 151–172.

Lukey-Coutsocostas, K and Dalmaris, D (1996) *Candidate-First Certificate Practice Tests* (second edition), Hertfordshire: International Book Distributors.

Maynard, D W (1980) Placement of Topic Changes in Conversation, *Semiotica*, 30, 263–290.

Norton, J (2005) The paired format in the Cambridge Speaking Tests, *ELT Journal*, 59(4), 287–297.

Reichman, R (1990) Communication and Mutual Engagement, in Dorval, B (Ed.) *Conversational Organization and Its Development*, Norwood, NJ: Ablex, 23–48.

Ross, S and Berwick, R (1992) The Discourse of Accommodation in Oral Proficiency Interviews, *Studies in Second Language Acquisition*, 14, 159–176.

UCLES (2000) *FCE Teaching Resource*, www.cambridgeesol.org/teach/fce/speaking/data/speaking_p34.pdf

UCLES (2001a) *CAE Handbook*.

UCLES (2001b) *FCE Handbook*.

van Lier, L (1989) Reeling, Writhing, Drawling, Sketching, and Fainting in Coils: Oral Proficiency Interviews as Conversation, *TESOL Quarterly*, 23(3), 489–508.

Young, R (1995) Conversational Styles in Language Proficiency Interviews, *Language Learning*, 45(1), 3–42.

Young, R and Milanovic, M (1992) Discourse Variation in Oral Proficiency Interviews, *Studies in Second Language Acquisition*, 14, 403–424.

Research and development update

Applying and monitoring a revised task penalty in BEC Higher Writing

Prior to December 2005, when a test-taker failed to complete either task of the BEC Higher Writing paper (called Part 1 and Part 2), by not covering all the essential content points of the input in their response, an upper boundary was set on the mark they could achieve of 2.2. This upper limit was referred to as 2.2 T.

A trial was undertaken in April 2005 to investigate the efficacy of applying a revised task penalty to deal with content point omission. It was proposed that an appropriate task penalty for each content point (and guidance on non-achievement of more than one content point) should be added to the task-specific mark scheme at the Team Leader (TL) co-ordination meeting.

The main aim of the April 2005 trial (Trial 1) was to compare the original live 2.2 T Part 1 and 2.2 T Part 2 December 2004 performances with re-ratings of the same performance when measured against the revised penalty mark scheme. A secondary aim was to survey the trial examiners in order to ascertain how they perceived the alternative rating scale (Shaw and Thighe 2005).

Findings from Trial 1 were encouraging. The overall effect on the weighted means in both parts of the Writing test remained approximately the same as the then current scale. In terms of the aggregate score for the paper, the effect of the revised scale was to marginally inflate the overall weighted mean.

Initial examiner feedback on the scale was mixed. On a positive note, some examiners perceived that the revised task-specific mark scheme allowed for greater flexibility in applying task penalties without an automatic downgrading to 2.2 T. In the main, examiners considered the proposed system to be fairer than its operational counterpart. The revised approach was also considered to be fairer to candidates who would otherwise mostly be judged as borderline or as having failed. Trial 1, however, revealed problems particularly with the Part 1 task-specific mark scheme and examiners expressed reservations regarding the complexity of the new scale.

As a consequence the mark scheme was revised and a sliding penalty mechanism introduced to the scale (as in FCE Writing Part 1) for both Parts 1 and 2, and a multiple re-marking trial of the revised scale employing both qualitative and quantitative methods was undertaken. The main aim of the research study was to report on the findings of a second trial in order to determine the efficacy of the changes made to the revised rating scale.

The study involved a multiple rating exercise: five raters – identified as representative of the 'universe' or worldwide population of raters for BEC Higher – rated a total of 200 writing performances. All raters in the study were highly experienced BEC examiners. The 200 writing performances were identified as 2.2 T scripts and the number marked reflected the relative contributions of the principal cohorts.

A principal aim of the marking trial was to improve both the reliability and validity of the assessment process for BEC Higher Writing by refining the rating scale. It was hoped that the combined use of quantitative methodologies (application of criteria and scales to sample language performance) and qualitative methodologies (insightful and intuitive judgements derived from 'expert' participants) would offer valuable insights into how the revised scale functions. Quantitative methodologies included correlational analyses, inter-rater reliabilities, analysis-of-variance (ANOVA) and FACETS analyses. The qualitative dimension of the trial comprised a feedback response form which was distributed to participants to supplement the quantitative findings.

Specific conclusions gleaned from the trial can be related to the various statistical and qualitative methods employed throughout the analysis of the trial data. The descriptive statistics and analysis of variance indicated that the raters were generally homogeneous in the marks awarded. Mean examiner intercorrelations were consistently high. Statistical significance tests indicated that the strength of the correlational findings was such that there was evidence of a good relationship between the examiners. Examiner inter-rater reliability was also encouragingly high – of the order of 0.75 for Part 1 (the task-specific mark scheme deemed most problematic in the April trial). FACETS analyses revealed that whilst examiners were not equally severe in their assessments any differences in severity (between the most and least severe examiner) were marginal. FACETS also indicated that all of the examiners were operating within an acceptable range of consistency of performance.

The effect on the weighted means using the new scale was estimated from the sample. It was found that the mean mark of candidates in Writing was very similar using the then current marking system and the new system.

Examiners were enthusiastic about the trial claiming it to be a very positive experience. The examiners were favourably disposed towards the new rating approach. Feedback responses indicated that in the main the examiners felt they understood the new descriptors and could use the marking scheme more accurately than its predecessor. Examiners also considered the new approach to be fairer and, encouragingly, all trial examiners were confident in their application of the revised mark scheme, believing the revisions made to the April trial rating scale brought greater overall clarity.

Following successful trialling, the latest version of the revised mark scheme was introduced during the December 2005 session.

Reference

Shaw, S D and Thighe, D (2005) *Task Penalty in BEC Higher Paper 2: application of a revised task penalty*, Cambridge ESOL, internal validation report no. 673.

Asset Languages update

With more data available for the current live test versions, and good linkage within stages between pretest and current live data for most languages, assessments administered in the May 2006 live session were anchored using National Curriculum estimates collected during pretesting, rather than being dependent on the grading of older tasks. Just under 100 reading and listening assessments were graded in this way for Panjabi, Urdu, Mandarin Chinese, Japanese, Spanish, French, German and Italian. Collection of cross-stage pretest data is continuing in order to provide further information on the vertical link.

Development work has been continuing for Advanced level (approximately B2 on the Common European Framework of Reference (CEFR)). Assessments will be available at Advanced level in French from November this year with Spanish and German assessments available from March 2007. A recent standardisation event took place for Advanced French Speaking. Experts from existing UK qualifications (e.g. A level) together with Asset Languages staff worked together to rate pilot materials for Asset Languages French Speaking against CEFR benchmarked materials. Experts were in agreement regarding the standard and felt that the standard set for Asset Languages Advanced was comparable with the B2 CEFR samples.

The last update of Asset Languages mentioned that several projects have been initiated to investigate the relationship between Asset Languages levels and the levels of current qualifications within the UK education system. In one such project, over 200 candidates have sat both Asset Languages and GCSEs in all four skills for French. Self-assessments and teacher ratings of candidates were also obtained using the Asset Languages Can Do statements. Another project has been initiated to investigate the relationship of Asset Languages and National Curriculum grades. Performances of candidates at each grade for Asset Languages are being compared with exemplar material from the National Curriculum. The results from these studies will help teachers to understand the relationship between the different frameworks more clearly and help with their reporting of results.

The Asset Languages website has been updated to include a section on research. Working with centres and language experts in the UK is vital in ensuring that we are providing a high quality product. The website provides an opportunity to update centres on specific research projects that we are undertaking and also to invite them to participate in research. For example, we are currently inviting centres who are taking both Teacher Assessment and External Assessment for Asset Languages to contact us. Teacher Assessments are administered, marked and certificated by the classroom teachers trained by Asset Languages. External Assessments, which lead to an externally recognised qualification, are subject to Cambridge ESOL's rigorous statistical analysis and grading procedures. Research and analysis into the grades teachers are awarding for Teacher Assessment will monitor grade comparability between the two modes of assessment.

For more information on Asset Languages visit www.assetlanguages.org.uk

ESOL staff seminar programme

Cambridge ESOL staff (and specially invited colleagues from within Cambridge Assessment or outside the organisation) have the opportunity to attend monthly seminars and workshops on a range of language testing and teaching topics. Since October 2005 ESOL staff have enjoyed seminars on a variety of areas, summarised below. We also report on invited presentations by Cambridge ESOL staff at international events.

Reviewing examinations: FCE and CAE

In November, two internal speakers updated staff on the current review of Main Suite exams, which began in 2004 with a web-based survey for candidates, the ESOL community and local secretaries. Various issues have been raised in this context and Anne Gutch gave an outline of these issues and summarised the responses to the surveys. She then described other activity regarding the review, which includes consultation and validation activity, and outlined some suggested possible changes to content which are being considered.

Ardeshir Geranpayeh reported on empirical work that was carried out on the CAE examination. He described analysis of a live CAE session to investigate the factors that underlie the exam. Ardeshir reported on the use of Exploratory Factor Analysis techniques to build several plausible construct models for each CAE paper and for the exam as a whole. He then reported how the viability of each model was tested by Structural Equation Modelling (SEM) techniques. The best fit indices reported came from a Correlated-Trait (CT) model. It confirmed that the

Cambridge model of language proficiency in the CAE examination is based on a componential view of communicative language ability whereby each component assesses a substantively different aspect of language proficiency.

Please see the FCE and CAE Review Bulletins for further more detailed information at http://www.cambridgeesol.org/about_us/news.htm

Emerging markets for Cambridge ESOL

In December, our regional development manager, Duncan Rayner, presented a session on emerging markets in Asia. As the Asian economies play an increasingly important role in the global economy, young people throughout the Asia region are paying more attention to studying English. They see good English skills and qualifications as key to their future careers. This trend is generating new business opportunities for Cambridge ESOL in Asia. In particular as a result of the booming IT and business process outsourcing (BPO) sectors in India it has become apparent that there are limitations to the supply of English speaking graduates and that large scale English skills training is needed for graduates who are not educated in the elite English medium education sector. Assessment of the English proficiency of these graduates before they can work in BPO industries is an area that Cambridge ESOL is involved in. India's main regional competitor, China, is one of Cambridge ESOL's largest markets. The candidature for business English qualifications such as BEC is growing significantly as graduate students want qualifications in English that will help them get jobs in international companies. Other important markets in South and South East Asia include Vietnam, Malaysia, Thailand, Indonesia and Sri Lanka. In these countries Cambridge ESOL has a network of centres which include British Council offices, language schools, international schools and non-institutional open centres. Exams tend to be taken by school and university students who are looking for high quality English language qualifications.

Stakeholder relationships

In January, four ESOL colleagues presented on various aspects of the many stakeholder relationships which Cambridge ESOL maintains with people and institutions worldwide. There exists a complex web of relationships between Cambridge ESOL and many different types of external stakeholder, including candidates and teachers, test centre administrators and receiving institutions, item writers and examiners, academic researchers and ELT publishers. Colleagues from the Business Support, Operations, and Research and Validation Groups described the nature of some of the key stakeholder relationships for their own Groups and highlighted some of the challenges they face.

Nic Underhill presented on the Professional Support Network (PSN) which consists of some 15,000 independent consultants carrying out roles such as oral and writing examiner, team leader, seminar presenter, item writer, centre inspector and so on. As well as being a vital component of our quality management system, they provide a valuable link to other professional communities in the overlapping network of relationships and we are developing systems to better support their contribution to our work. Lynda Taylor, Sarah Corcoran and Stephen McKenna also contributed to this session.

CIE and work with national education systems

In February, Ann Puntis, the Chief Executive of Cambridge International Examinations (CIE), our sister organisation, led a seminar on CIE's work with national education systems. A large proportion of CIE's work relates to the delivery of national examination systems, either in whole or in part, around the world and involves this part of Cambridge Assessment in working with Ministries of Education and other Government departments, principally in the 14–19 sector. Ann outlined what the term 'localisation' meant for CIE's work in this capacity and she reviewed the types of countries in which CIE works, the factors that underpin localisation, and CIE's strategies for expansion in its work with governments. This was a fascinating insight into stakeholder groups and activities less well-known to Cambridge ESOL staff.

The future of language

In March we welcomed Professor David Crystal to Cambridge who spoke about 'The future of language'. The end of the twentieth century was a revolutionary period for language and languages. The 1990s was the decade when people finally acknowledged that the world had a genuine global language, English, now spoken by around a third of the planet's population. One of the unexpected consequences was a remarkable increase in the number of 'new Englishes', expressing national identities. It was also the decade in which the crisis affecting endangered and minority languages was reported. At least half of the world's languages are so seriously endangered that they will become extinct in the course of the present century: that is one language dying on average every two weeks. And it was the decade when most people took an active interest in Internet technology, for the first time sending emails, joining chat groups, and exploring the World Wide Web, and thus having to come to terms with a range of unprecedented linguistic processes, varieties, and strategies.

These developments have major implications for teachers and examiners of English. The arrival of new Englishes forces a re-evaluation of the balance of linguistic power in the world: new varieties are complicating the simple model of British vs American English, and providing new usages (especially in vocabulary, grammar, and pronunciation) which raise questions of what counts as 'correct'. The disappearance of so many languages is a loss for humanity, which everyone needs to address – and not least teachers of English, who need to be among the front ranks of those who affirm the importance of linguistic diversity. The Internet offers a global presence for both of these portfolios, allowing us to encounter new Englishes at first hand and providing those who wish to revitalize endangered languages with a welcome public forum. But in addition the Internet is radically altering the stylistic balance of English, adding new and especially informal styles to what was previously available, and speeding up the process of language change. The arrival of instant messaging and blogging in the new millennium shows that the impact of the technology is continuing. For professionals working with language, things are never going to be the same again.

Further staff seminars from 2006 will be reported in future issues of *Research Notes*.

Invited presentations

Since the start of 2006 several members of the Research and Validation team have accepted invitations to be guest speakers at various events worldwide.

In February Barbara Stevens gave a plenary presentation entitled 'Reflexiones sobre el Portfolio Europeo de las Lenguas' at a one day event Jornada Porfolio europeo de les llengües secundària in València. She also gave a workshop on 'Reino Unido, la inmigración y los exámenes de idioma' in Toledo at Crisol de Culturas, Seminario Internacional Consejo de Europa in March. Also in March, Neil Jones participated in an international symposium on the potential role of the Common European Framework in foreign language education held at the University of Foreign Studies in Osaka, Japan where he spoke about the impact of the CEFR on language testing in Europe. In the same month Lynda Taylor gave a presentation on the development and role of IELTS in assessing English language skills for higher education contexts as part of a joint seminar program run by the Universities of Newcastle and Northumbria. In May she was a plenary speaker at the annual conference of the Academic Committee for Research on Language Testing (ACROLT) in Israel where she discussed the contribution of qualitative research methods to our understanding of constructs in language testing; while in Israel she also gave a seminar at the National Institute for Testing and Evaluation (NITE) in Jerusalem on some of the theoretical and practical issues associated with offering suitable testing provision for language learners with disabilities. In June Nick Saville participated in the Hellenic American Union's second Language Testing Forum in Athens where he presented a model for investigating the impact of language assessment within state educational contexts.

Recent publications of interest

Studies in Language Testing

Three more titles have recently appeared in the Studies in Language Testing series, published jointly by Cambridge ESOL and Cambridge University Press.

The Impact of High-stakes Testing on Classroom Teaching: A case study using insights from testing and innovation theory by Dianne Wall (Volume 22) contains an account of one of the first data-based studies of examination 'washback'. Through a detailed analysis of the impact of examination reform in one specific educational setting, it considers the effects of an examination which was meant to serve as a 'lever for change'; the study's findings show clearly how the intended outcome was altered by factors in the examination itself, as well as by characteristics of the context, the teachers and the learners. Key features of the volume include:

- a comprehensive review of the literature on language testing, examination impact and innovation in education
- a detailed and rigorous case study analysis which is relevant to examination reform movements in many countries
- valuable insights and advice for educators seeking to implement curriculum change through examination reform.

Volume 22 provides not only a helpful model for researching washback and impact but also practical guidelines for the planning and management of change within an educational context. As such, it will be of particular relevance to all who are involved in the process of curriculum and examination reform, as well as to academic researchers, university lecturers, postgraduate students, and practising teachers.

Volume 17 – *Issues in Testing Business English: The revision of the Cambridge Business English Certificates* by Barry O'Sullivan – explores the testing of language for specific purposes (LSP) from a theoretical and practical perspective, with a particular focus on the testing of English for business purposes. A range of tests – both past and present – is reviewed, and the development of business English testing at Cambridge ESOL is discussed. The description of the revision of the Business English Certificates (BEC) forms a major part of the book and offers a unique insight into an approach to large-scale ESP test development and revision. The volume:

- reviews the historical development of testing English for business and presents current thinking and practice in this area
- provides a sound theoretical rationale in support of the testing of English for business and other specific purposes
- presents a case study of the development and revision of an internationally recognised test of English for business
- offers a systematic methodology for describing LSP tests together with practical guidance on test development issues.

Demand is steadily growing for language tests with a specialised focus which will suit the needs of key professional domains as diverse as business, law, the aviation industry, and teacher education. O'Sullivan presents a multicomponential view of specificity and is able to clearly distinguish between different tests and tasks using his approach. This volume will be of particular relevance to language test developers and researchers interested in language testing for specific purposes and contexts of use; it will also be of interest to ESP teachers – particularly those teaching English for business, as well as to lecturers and postgraduates working in the field of LSP. *Issues in Testing Business English* is the third volume in the SiLT series (following Volumes 15 and 16) to document both a historical perspective and a study of test revision with a focus on their implications.

Impact Theory and Practice: Studies of the IELTS test and Progetto Lingue 2000 by Roger Hawkey (Volume 24) describes two recent case studies to investigate impact in specific educational contexts. One analyses the impact of the International English Language Testing System (IELTS) – a high-stakes English language proficiency test used worldwide among international students; the second focuses on the *Progetto Lingue 2000* (Year 2000 Languages Project) – a major national language teaching reform programme introduced by the Ministry of Education in Italy. Key features of the volume include:

- an up-to-date review of the relevant literature on impact, including clarification of the concept of impact and related terms such as washback
- a detailed explication of the process of impact study using actual case study examples
- practical guidance on matters such as questionnaire design, interviews, permissions and confidentiality, data collection, management and analysis
- a comprehensive discussion of washback and impact issues in relation to language teaching reform as well as language testing.

Language teaching and testing programmes have long been considered to exert a powerful influence on a wide range of stakeholders, including learners and test takers, teachers and textbook writers, test developers and institutions. However, the actual nature of this influence and the extent to which it may be positive or negative have only recently been subject to empirical investigation through research studies of impact. With its combination of theoretical overview and practical advice, this volume offers another useful manual on how to conduct impact studies and will be of particular interest to both language test researchers and students of language testing. It will also be relevant to those who are concerned with the process of curriculum and examination reform.

Other publications

As a responsible and ethical examination board Cambridge ESOL is committed to offering equal opportunities – removing barriers and setting appropriate goals while at the same time protecting exam integrity and ensuring fairness to all. We continue to try and raise awareness among test stakeholders and the wider community of the special arrangements we offer to students with special requirements, e.g. a long-term or temporary disability. Language tests can represent an important incentive for learning, particularly where they are used for university admissions purposes or as part of job recruitment procedures. There is also a growing legal obligation in many countries to protect people from discrimination on the grounds of disability. In a recent article in *English Teaching Professional* (Issue 41, Nov 2005), Ruth Shuter describes aspects of the special arrangements provision now available to candidates taking our examinations, and she encourages us to think more deeply about the integration of students with special requirements in all types of assessment and in all aspects of education.

Issue 60/1 of the *ELT Journal* (Jan 2006) contains a helpful review of the TKT Teaching Knowledge Test Course by Spratt, Pulverness and Williams which was jointly published in 2005 by Cambridge ESOL and Cambridge University Press. Cambridge ESOL's new Teaching Knowledge Test (TKT) was introduced in 2005 as part of our framework of teaching qualifications, which includes the well-established CELTA and DELTA; TKT is designed to provide support for English teachers at any stage of their career to help them consolidate their existing knowledge or move on to higher level qualifications. Further information about TKT is available from: www.cambridgeesol.org/TKT

Cambridge University Press recently published a major new reference grammar of English. *The Cambridge Grammar of English: A Comprehensive Guide* is authored by Ronald Carter and Michael McCarthy, both at the University of Nottingham and highly respected experts in the field of applied linguistics and ELT. Based on extensive corpus research, including analysis of the Cambridge Learner Corpus (CLC) compiled jointly by Cambridge ESOL and Cambridge University Press, their new grammar offers comprehensive coverage of spoken and written English based on real everyday usage. The book is designed to be user-friendly through a two-part structure: in the first section, A–Z entries give more attention to lexico-grammar and other language areas that tend to be neglected in most grammar reference books; the second section covers traditional grammatical categories such as tense, clause structure and parts of speech, including the latest insights into how grammar varies between spoken and written language. An accompanying CD-ROM makes the *Cambridge Grammar of English* even more accessible with: the whole book in handy, searchable format; audio recordings of all example sentences from the book; and links to the *Cambridge Advanced Learner's Dictionary* online for instant definitions of new vocabulary. For more information see the Cambridge University Press online catalogue: www.cambridge.org/uk/catalogue