

Research Notes

Contents

The classroom and the Common European Framework: towards a model for formative assessment	2
Development of Can-do Statements for KET and PET for Schools	8
Adapting testing materials for younger learners: developing KET and PET for Schools exams	12
Validating a worldwide placement test for German	15
Linking learning and assessment: Cambridge ESOL's blended learning approach	18
Impact of a Blended Learning course: observation and stakeholder views	23
Profile of Skills for Life candidature in 2007–8	28
IELTS candidate performance in 2008	30
Conference reports	33
Recent publications of interest	34
Call for IELTS Joint-funded Research Program and IELTS Masters Award proposals	35
Cambridge to host two major conferences	36

Editorial Notes

Welcome to issue 36 of *Research Notes*, our quarterly publication reporting on matters relating to research, test development and validation within Cambridge ESOL.

This issue focuses on assessment for learning or formative assessment, usefully defined in this issue as ‘any activity (not limited to testing) which engages the expertise of language assessment specialists and has the aim of supporting learning’ (see p2). In the opening article, Neil Jones considers the potential role of a language testing body with respect to language learning, and how this role might be fulfilled. He suggests that bringing assessment-for-learning techniques into the classroom depends on teacher development, which is something that the language tester could do.

The next three articles review the development of the test construct and materials for assessing the language proficiency of specific groups of candidates. Firstly, in relation to the recently-launched KET for Schools and PET for Schools tests, Szilvia Papp describes the process of developing Can-do statements and tasks for these examinations. Next, Ed Hackett describes how existing KET and PET exam materials have been adapted to meet the needs of the younger learners being attracted to these new examinations. In the third article in this section, Ardeshir Geranpayeh and Sibylle Bolton report some of the activities involved in relating the German Placement Test, a joint product of the Goethe-Institute and Cambridge ESOL, to the Common European Framework of Reference.

We then consider the concept of blended learning and its impact on our provision of support materials and courses. Reinhard Tenberg describes how we are developing online blended learning courses which complement our language testing products, taking as his example a Financial English course aimed at intermediate to advanced level learners worldwide. Roger Hawkey then presents the results of an impact study of a General English online blended learning course designed for Italian university students.

Next, we present candidate information on two exam suites, namely the profile of Skills for Life candidature 2007–8 and the performance of candidates taking the IELTS exam in 2008 (both Academic and General Training Modules). Further performance data can be found on our website – www.CambridgeESOL.org – for other years and examinations.

We finish this issue with conference reports, recent publications of interest and the calls for submissions for the 2009 IELTS Joint-funded Research Program and IELTS Masters Award. The deadline for both is the end of June, so visit the IELTS website now for further information on how to apply. We also give details of the 4th Cambridge Assessment Conference we are hosting in October and of the Language Testing Research Colloquium happening in Cambridge in April 2010.

Editorial team for Issue 36: Fiona Barker, Hanan Khalifa and Caroline Warren.

The classroom and the Common European Framework: towards a model for formative assessment

NEIL JONES RESEARCH AND VALIDATION GROUP, CAMBRIDGE ESOL

Aims of a formative assessment model

The purpose of this article is not to summarise Cambridge ESOL's current activity in the area of formative assessment; nor to trace, for example, the historical developments which saw a single exam level – the Cambridge Certificate of Proficiency in English, launched in 1913 – grow into a suite of products with the clear function of a “learning ladder”. The purpose is rather to define the potential role of a language testing body with respect to language learning, and from there to consider how this role might be fulfilled. I am using “formative assessment” in this article to mean any activity (not limited to testing) which engages the expertise of language assessment specialists and has the aim of supporting learning. This usage may in places disrespect the distinction commonly made between *formative* and *summative* assessment. The scope of such activity can be simply stated. It is:

- to promote learning
- to interpret learning in terms of a domain of use beyond the classroom.

The first of these is self-evident, and Cambridge ESOL's current exam suite, though essentially summative in conception, can already be seen to fulfil this role very successfully. Learners choose an appropriate level, and generally follow dedicated courses of study, which are well-supported by published materials. The exams themselves have over the years been revised in line with developments in language teaching, so that preparing for one should offer opportunities for effective learning. The link to learning is strengthened by the large amount of support offered to teachers, through publications, online resources, and seminars.¹

The second aim requires more explanation. The issue at heart is that what a learner knows of a language, or can-do with it in the classroom at a given point in time, stand in no simple relation to the learner's final objectives in learning the language, be they defined by the learners themselves or the educational system they are part of. This is an idea that I will develop further, but for now let us note that the objectives will most probably be conceived in terms of communicative language ability, of use in some domain or “real world”, but the skills and knowledge the learner is acquiring, and the nature of the learner's emergent competence, need not translate directly into such ability.

This being the case, one might object that the second aim is simply wrong: classroom learning should be interpreted in terms of classroom objectives, and not

extrapolated to some hypothetical future domain of use. There is some truth in this, and yet there remain serious reasons for wishing to relate classroom performance to a domain external to the classroom.

Firstly, interpretation linked to extrinsic learning objectives is good for learning. It is motivating, and provides an orientation for teachers and learners as to where they are and the ground yet to be covered – an important enabler of learner autonomy and self-direction.

Secondly, there remains a continuing role for standardised assessment alongside whatever other forms of classroom assessment may be put in place. Summative standardised assessment at the end of schooling or at important transition points, for example, has a long future. This kind of assessment must relate to final learning objectives. To engage seriously with formative assessment at classroom level an assessment body has no choice but to try and construct a single frame of reference, somehow linking to the interpretation of summative standardised assessments. We should think of a triangle linking the interpretation of formative and summative assessments to each other, and both of them to a domain of use beyond the classroom. Clearly this relates to fundamental issues of test validity.

For language testers in Europe there is a third reason why their model of formative assessment should enable interpretation of classroom performance in terms of some real-world domain of use; and it is one which lends urgency to the endeavour. This is the rise to prominence of the Common European Framework of Reference (Council of Europe 2001) as an instrument of language policy, particularly for defining learning objectives and assessing outcomes. Formative assessment as conceived in this article provides the method and the mechanism by which an argument can be built relating a specific context of learning to the CEFR. Learning contexts should be related *to* the CEFR (and not the other way round) because the CEFR is a point of *reference*. However, unless adequate thought is given to how this should be done there is a risk of language policy leading to simplistic *application* of the CEFR to the classroom, to the potential detriment of language learning. Before I go on to develop the notion of a formative approach it is worth discussing the impact of the CEFR in a little more detail.

The CEFR's impact on learning, teaching and assessment

The impact of the CEFR on language testing may be seen as part of a general trend. McNamara and Roever (2006:213)

1. See www.CambridgeESOL.org

perceive ‘a tendency for governments to use assessment as part of a general climate of greater accountability in education and, more specifically, as a means for the achievement of particular political goals. The politicisation of assessment in these ways is perhaps the most striking feature of current developments in language assessment.’ They find the CEFR to be ‘dominating language education at every level in Europe, in the most comprehensive example of policy-driven assessment yet seen.’ (ibid.:212).

The CEFR, as its subtitle “Learning, Teaching, Assessment” indicates, sets out to provide ‘a common basis for the elaboration of language syllabuses, curriculum guidelines, examinations, textbooks, etc. across Europe’ (Council of Europe 2001:1). Developed by the Council of Europe and adopted by the European Union as an instrument of policy, it has indeed within a short space of time become ubiquitous. Studies conducted by the Council of Europe in 2005 and 2006 (Council of Europe 2006, Martyniuk and Noijons 2007) concluded that in all countries the CEFR has had a major impact at the level of policy. However, at the level of the classroom the impact was found to be less, because the framework is found difficult to understand and teachers are not equipped to make use of it.

Several writers have therefore concluded that the CEFR’s major influence to date has been on assessment (Coste 2007, Fulcher 2008, Little 2007). Certainly the publication of the CEFR and the subsequent manual for relating language examinations to it (Council of Europe 2008a) has stimulated a great deal of work by language testers. It has led to increasing insistence by governments and other test users on proof of alignment, and effort on the part of testers to provide it. This influence is set to increase. A recent recommendation on the use of the CEFR by the Council of Ministers includes the call for countries to:

‘... ensure that all tests, examinations and assessment procedures leading to officially recognised language qualifications take full account of the relevant aspects of language use and language competences as set out in the CEFR, that they are conducted in accordance with internationally recognised principles of good practice and quality management, and that the procedures to relate these tests and examinations to the common reference levels (A1–C2) of the CEFR are carried out in a reliable and transparent manner.’ (Council of Europe 2008b)

The European Survey on Language Competences, scheduled for delivery in 2011, will use the CEFR as the benchmark for reporting on the language competences of secondary-school pupils across Europe, lending further visibility to the CEFR as an assessment framework.²

This somewhat one-sided adoption of the CEFR is linked to a one-sided reading of its text, as noted by Coste (2007): ‘In various settings and on various levels of discourse ... people who talk about the Framework are actually referring only to its scales of proficiency and their descriptors.’ Whatever the reason, one can identify a disjunction between the current widespread adoption of the CEFR in assessment and the framing of targets, and its generally less developed application to classroom practice and to

specific learning contexts. This situation lends urgency to the need to consider afresh the relationship of assessment to learning, looking for models that link inputs to learning to its desired outcomes. The problem is not, I believe, with the CEFR’s aim of providing comparable standards across languages: Cambridge ESOL as a member of ALTE (the Association of Language Testers in Europe) has pursued similar aims since the early 1990s. What is critical however is that the basis of comparison between one learning context and another should be carefully articulated, and that the specific aims and rationales of each context be respected.

A model of language use and learning

Our formative assessment model starts from a model of language use and learning. Having identified the CEFR as a focus of interest I will take as a starting point the elements of the model proposed by the CEFR, which describes:

‘The actions performed by persons who as individuals and as social agents develop a range of *competences*, both *general* and in particular *communicative language competences*. They draw on the competences at their disposal in various contexts under various *conditions* and *constraints* to engage in *language activities* involving *language processes* to produce and/or receive texts in relation to *themes* in specific *domains*, activating those *strategies* which seem most appropriate for carrying out the *tasks* to be accomplished ...’ (Council of Europe 2001:9, emphasis in original)

To aid presentation Figure 1 illustrates essentially what is described in the above paragraph. Some domain of language use throws up tasks which the learner/user must address by engaging in language activity, calling upon cognitive processes and strategies. Learning happens as the user monitors their own performance.

The left-hand side of the figure focuses on language ability as a cognitive trait, potential use, and generalisability. The right-hand side focuses rather on language ability as actual use in specific contexts. In assessment, the kind of language activities we set out to observe reflect our focus of interest and the nature of the inferences we may make on the basis of those observations. Bachman (2007) reviews the way language testing shifts between these two foci and the tensions between them.

The model of learning illustrated by Figure 1 appears to relate more clearly to an acquisition-rich environment where dealing with the demands of daily life provides ample opportunities for language learning to happen. This is not

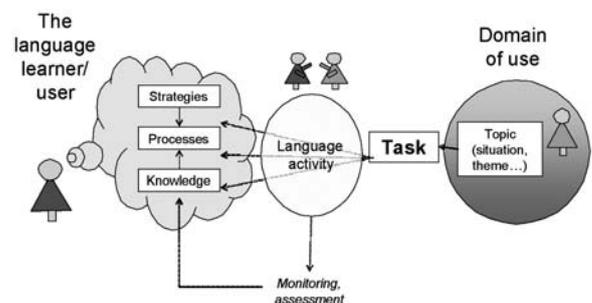


Figure 1: The CEFR’s model of language use and learning

2. See www.SurveyLang.org and <http://europa.eu/scadplus/leg/en/cha/c11083.htm>

wholly adequate to describe learning in an institutional, formal setting, which is the case that interests us here. The language classroom is not a domain of use like any other, chiefly for the reason identified at the outset: that its sole business is language learning, and it is inhabited solely by learners and teachers. A domain of use beyond it is understood: a perhaps sketchily outlined world in which the learners may one day move independently, socialising, transacting business or acquiring the culture; but the classroom is not it.

In the classroom the exposure of the learner to language is controlled. The syllabus and course material will probably structure the coverage of elements of the language system, vocabulary, and notions and functions, imposing some pedagogically-motivated progression. In this way the *content* of learning is defined. The CEFR lists exhaustively these organisational options. The choices made among these options will reflect the overall aims of the curriculum, emphasising some purposes and aspects of language more than others.

There are also important methodological choices to be made. The CEFR has been criticised for espousing a particular approach, and an outdated one at that. McNamara and Roever (2006:212) are typical when they say of the CEFR descriptor scales: 'These functional descriptors used as the basic building blocks of the scale development faithfully reflect the fundamental underlying construct of the assessment [sic], a 1970s notional/functionalist that was given its clearest expression in the work of van Ek and Trim'. The criticism is understandable, given the way readers of the CEFR are repeatedly invited to 'consider and where appropriate state' their choices with respect to content, particularly throughout chapter four – *Language use and the language learner*. Users are invited to specify among other things the domains, situations, physical locations, themes, tasks, and even the ludic and aesthetic uses of language which their learners will need. Chapter four also contains most of the descriptor scales, because, as the authors explain, observable uses of language are the most readily scalable and thus useful for describing progression. The notional/functional emphasis thus partly results from the unintended prominence of the descriptor scales in most readers' understanding of the CEFR. Chapter five – *The learner's competences*, also elicits choices, among morphological elements, grammatical elements and relations and semantic relations, but there are other prompts which invite reflection on methodological choices. The prompts in chapter six – *Language learning and teaching* – and the remaining chapters are almost entirely methodological in focus: what assumptions users make about the process of learning; which of a list of general approaches they use; what they take to be the relative roles and responsibilities of teachers and learners, and so on. These little-read invitations to methodological reflection make the CEFR more open than it is generally given credit for.

This openness, however, does not imply an absence of policy, and the Council of Europe statements of policy referred to in the text emphasise the satisfaction of learners' "communicative needs" including dealing with the business of everyday life, exchanging information and ideas, and achieving a wider and deeper intercultural

understanding; all this to be achieved by 'basing language teaching and learning on the needs, motivations, characteristics and resources of learners', and 'defining worthwhile and realistic objectives as explicitly as possible' (Council of Europe 2001:3). This conveys the CEFR's basic communicative, action-oriented approach. It is a fairly broad brief which should be coherent with the aims of most school language learning and leaves scope for a range of implementations.

Can-do statements for the classroom

Figure 2 re-draws the initial model of language use presented above to reflect the classroom domain. The tasks which engage learners in language activity are part of a controlled presentation of content. Typically, units of learning relate to some topic or theme and comprise tasks linked to presentation and practice of elements of the language system: a grammar point, an area of lexis, and so on. But much more is going on in the language classroom – or should be – than the mere presentation and practice of course content. There is also a methodological or learning focus.

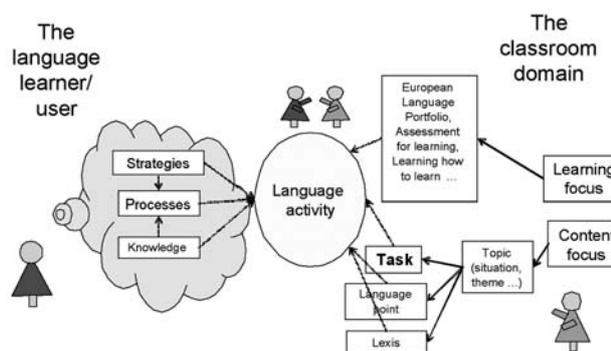


Figure 2: A model of language use for the classroom

It is the right-hand side of Figure 2 – the side focusing on actual language use in context – which differs from Figure 1. Language activities are motivated much more by explicit, structured treatment of language content and by methodology. The left-hand, cognitively-focused side is apparently the same. This can be linked to the notion of *interactional authenticity* to which standardised assessment often appeals: tasks in tests are not authentic, but in addressing the tasks the learner is claimed to engage the same mental processes as she would in the "real world" domain of use. The use of tasks in teaching makes similar assumptions. What teaching and assessment have in common, however, is the linguistic control applied in order to make tasks accessible. Linguistic control is a critically important aspect of the support or *scaffolding* provided for classroom performance, and must be taken into account in its interpretation.

An approach to formative assessment cannot be methodologically neutral, although it may well be inclusive and eclectic. We need to take a view not only on the content of learning but also on the activities that promote learning: on the enabling skills that both learners and teachers need. We might call these *can-do statements for the classroom*.

First let us look at the European Language Portfolio (ELP), a parallel development to the CEFR, closely linked to it and focusing the most coherent and principled attempts to integrate it into classroom learning. The ELP serves the pedagogical function of encouraging reflection on the language learning process and the development of learner autonomy, while at the same time documenting the learner's developing proficiency (Little 2007:649). Lenz and Schneider (2004) in an interesting introduction to a bank of ELP can-do descriptors available on the Council of Europe website, identify four different kinds of descriptor:

1. Scaled CEFR descriptors
2. Adaptations of CEFR scaled descriptors for particular learner groups, e.g. young learners
3. Context-specific achievement-oriented descriptors, e.g. related to syllabus content
4. Descriptors which do not relate directly to a language proficiency scale but rather to independent categories such as learning strategies or descriptions of cultural or intercultural experiences.

Type 3 reflect the fact that for learners it may be more meaningful to record that, for example, they can form a past tense correctly than that they can order a cup of coffee. Or as Lenz and Schneider say (2004: pp): 'being so closely related to actual classroom practice, these descriptors may be tremendously useful in improving learners' self-assessment skills and, if linked to an ELP in some way, enhance the perceived usefulness of that ELP'. Type 4 takes in a range of less directly linguistic competences which may be nonetheless of fundamental importance for learning.

The ELP philosophy shares much with the conception of *assessment for learning* put forward in the UK by the Assessment Reform Group (1999), as a reaction against standardised testing and externally-defined criteria, towards assessment that improves learning, flexibly integrated into day-to-day classroom practice. Actively involving pupils in their own learning, enabling them to assess themselves and to understand how to improve, learning how to learn (LHTL) – these are key features of assessment for learning that are also reflected in many versions of the ELP. Assessment for learning is also about developing the skills of teachers: providing effective feedback, adjusting teaching to take account of the results of assessment, and recognising the influence of assessment on motivation and self-esteem.

Lenz and Schneider (2004) consider that such learning-focused can-do statements are not scalable within a language proficiency construct; but one can imagine a formative approach where identifying a measurable progression in such skills might be possible and worthwhile. It is interesting to note that the European Commission is currently exploring the feasibility of developing educational indicators of LHTL skills, and even of creativity.

In this section I have given examples of can-do statements for the classroom – skills which enhance learning, and which a formative assessment model will need to accommodate in some way.

The nature of a learner's language competence

The current orthodoxy regarding learning is broadly constructivist: learning is not a matter of passively receiving instruction or drilling habits, but represents an active endeavour on the part of the learner to make sense of things. Within this broad approach we can identify different emphases. A more analytic view is represented by Pellegrino et al. (2001), who insist on the need to construct cognitive models of how learners represent knowledge and develop competence in particular areas – including both correct and incorrect conceptions. As Mislavy (1992:15) states: 'Contemporary conceptions of learning do not describe developing competence in terms of increasing trait values, but in terms of alternative constructs'. Pellegrino et al. (ibid.) identify three elements of an "assessment triangle": cognition, observation and interpretation. They argue that advances in understanding cognition and in the technology of assessment (using ICT to observe process as well as outcomes, sophisticated statistical models to provide interpretation), create new opportunities to devise assessments that can impact directly and positively on learning. With examples drawn mostly from maths and science, and a strong problem-solving flavour, this model may be more relevant to some aspects of language learning than others. Diagnostic assessment, or explicit, perhaps contrastive, approaches to teaching syntactic rules, are possible areas.

A different approach to analysing learning events is in terms not of the learning objectives themselves but of the situation in which learning happens. What learners do in the classroom is supported, or *scaffolded*. Teasdale and Leung (2000) depict a learner's state of knowing or understanding as inseparable from the particular situation of learning, and identify this as a problem for comparing classroom performance with performance on formally-administered tests.

Similarly, Chalhoub-Deville (2003) reviews assessment models of communicative language ability and proposes an *interactional* perspective in which 'individual ability and contextual facets interact in ways that change them both' (Chalhoub-Deville 2003:369). This 'ability – in language user – in context' model imposes a local, context-bound view of language ability, which is difficult to reconcile with the assessment goal of score generalisability. Chalhoub-Deville (ibid.) sees the way forward in identifying 'contexts which activate stable ability features' and 'attempting to account for inconsistent performance in particular contexts from a social interactional perspective'. She uses "context" to refer to different contexts of use, but her conclusions could be taken to apply to the special but fundamentally important case of interpreting classroom performance.

The notion of dynamic assessment (Poehner 2007) is based on the idea that differing degrees of support lead to differing levels of performance, and claims that a learner's response to support is an indicator of language learning aptitude. The dynamic assessment literature does not seem to address the issue of interpretation which is of central concern in this article; however, the adoption of the

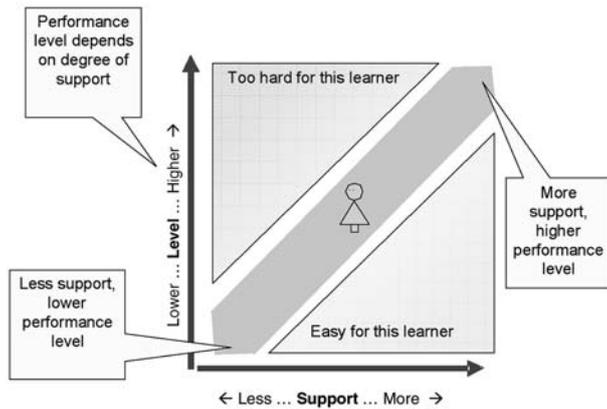


Figure 3: The zone of proximal development as the basis of a measurement model

Vygotskian notion of a zone of proximal development (ZPD) coincides with Cambridge ESOL’s thinking in this area (Jones 2006; see Figure 3) and provides a potential basis for a measurement model that factors degree of support into the interpretation of performance.

Elements of an approach to formative assessment

A worked-through approach to formative assessment will necessarily be complex and heterogeneous, eclectic and inclusive. For an assessment body it implies developing an extensive toolbox of resources, applications and techniques. Some of these will be existing tests (including those we are used to calling summative) or wraparound services. Others require significant innovation and development. Table 1 sketches the elements of a system.

Table 1: Elements of a formative assessment approach

Focus	Learning as content, knowledge, ability	Learning as process, interaction, use
Features	Discrete, analytic, testable, objective; “Knowing what students know” (models of cognition); Defined objectives	Integrated, synthetic, subjective; “Assessment for Learning”/ ELP Flexible objectives
Tools and techniques	Diagnostic tests Exercises Intelligent tutors Blended learning courseware	Teacher development ELP Familiarisation with levels (teachers, learners)
Reference level descriptions (English Profile)		
Link to domain of use/ CEFR	Explore statistical relation of formative and standardised assessments Link formative to standardised assessments through explicit statistical modelling (performance = ability + support)	For Speaking, Writing: Exemplification and comparison of more or less scaffolded production For Reading, Listening: Exemplification through calibrated bank of texts – comprehensibility of these with differing degrees of scaffolding. Treatment of scaffolding, interlocutor roles in CEFR descriptor scales
Teacher- and self-ratings, globally, and in relation to specific tasks		

Table 1 distinguishes two poles of a continuum between two rather different views of learning. They seem to be complementary, and both have a place in a formative approach.

The tools and techniques are indicative of the kind of products or activities which together would constitute a language tester’s formative offering. These might well require collaboration with, for example, publishers, teacher trainers and so on. Where the focus is more on content there is clear scope for deploying ICT-based applications: diagnostic tests, blended learning courseware etc (see, for example, Tenberg 2009). What would make these considerably more interesting is the addition of a measurement model that would, by factoring in degree of support, interpret the learner’s responses on the system in terms of extrinsic learning objectives such as CEFR level. Bringing the ELP or assessment-for-learning techniques into the classroom depends, however, on teacher development, and the language tester’s role would be to contribute to this. One of the key enabling skills which teachers need, as was found during the introduction of the Asset Languages tests in British schools (Jones and Saville 2009), is a familiarity and intuitive grasp of the framework of levels they are being asked to relate to. The language tester is well-placed to provide useful exemplification of these. The English Profile mentioned in Table 1 is a long-term collaborative research programme in which Cambridge ESOL is a partner: the major outcome of which will be a linguistic description of progression over the six levels of the CEFR, based on sophisticated analyses of corpora of learner and native speaker English. The English Profile Programme will be producing a range of valuable resources for teachers and learners. See articles on English Profile in issue 33 of *Research Notes* (Cambridge ESOL 2008).

The approaches suggested in Table 1 for linking to the CEFR are topics on the current Cambridge ESOL research agenda. The use of self-rating for this purpose fits well into a formative approach and is an area which has produced useful results in previous projects (Jones 2000, 2001a, 2001b, Papp 2009).

A note on the status of the CEFR

In this article I have taken the CEFR as representing the domain of use beyond the classroom in terms of which we would wish to interpret classroom learning. This reflects the dominance in Europe of the CEFR as a planning instrument and assessment tool. It is worth repeating that my purpose in this presentation of formative assessment is to argue in favour of treating each learning context on its own terms, and against the blanket application of the CEFR in simplistic and ill-considered ways. The authors of the CEFR have of course repeatedly stated that it is not intended to be an instrument of harmonisation; the Council of Europe now states in bold type on the website: ‘Rather than vis-à-vis the Council of Europe, it is towards one’s own learners and one’s European partners that one has a responsibility for making coherent, realistic use of the CEFR’. There is certainly a great deal of work to be done in order to guarantee its sensible use, and ensure that its

impact on the learning and teaching of languages turns out, on balance, positive. This work involves actors at every level of language education, and in every field, and there is a large role for assessment and measurement expertise. Jones and Saville (2009) present the CEFR as a fit object for the study of test impact, and propose that language testers should be proactive in foreseeing and, if possible, heading off problems resulting from its (mis)use.

Further, we should not treat the CEFR as a finished product. In particular, we should not treat the descriptor scales as definitive, but be prepared to continually enrich our conception of levels by incorporating other contexts. The scope of the CEFR is the study of a language as a foreign language in some more-or-less formal setting. As such it relates readily to many contexts and not so well to others. We should not be shy of extending it and if necessary stepping outside it, as Coste, one of the CEFR's authors, envisages (Coste 2007).

Specific issues are raised by CLIL contexts, or where language of schooling (i.e. learners acquiring education in other than their first language) is considered (see Ashton and Galaczi 2008). Explicit consideration of cognitive stage may also be needed (cf. the WIDA consortium's *English Language Proficiency Standards for English Language Learners in Kindergarten through Grade 12*, see www.wida.us/).

Lastly it is worth considering how much of the progression described by the CEFR scales feeds back to us what is actually just a widely-followed convention about how languages should be taught. Describing one's family or bedroom does not, I think, fulfil an immediate need, yet it is very much the stuff of A1 level. This may not be a problem to the extent that the scales fulfil their function of exemplifying the levels in ways recognisable to teachers, but it becomes a problem if describing your bedroom becomes a testable, criterial feature of the level – a tick-box to be checked. And if we accept that there is a conventional element to progression, then this motivates the diversification of context-specific objectives and descriptions of level.

Where would this leave the CEFR? These diverse objectives might be brought together into a single rich and complex depiction of progress (which would not fit within the pages of a single book), constituting a reference where every context could find something to relate to. Areas of language use that are currently bundled together might be separated out, enabling better characterisation of contexts in terms of profiles of skill. Thus making explicit the distinction between CALP (cognitive academic language proficiency) and BICS (basic interpersonal communicative skills) would facilitate the relation of the CLIL or L2 learning contexts to the CEFR (see Cummins 2000).

Finally it is linguistic progression, from a simple and narrow repertoire to a complex and broad one, which provides the common factor across all contexts of learning. One might assert that a linguistic characterisation can be constitutive of a level, whereas a functional one can be merely descriptive. This is why the English Profile is a particularly important project for the further elaboration of the CEFR in relation to learners of English.

References and further reading

- Ashton, M and Galaczi, E (2008) Assessment for teaching: Cambridge ESOL's CLIL Exam, *Research Notes* 34, 15–21, Cambridge, Cambridge ESOL.
- Assessment Reform Group (1999) *Assessment for learning: beyond the black box*, University of Cambridge School of Education, retrieved from: www.qca.org.uk/7444.html
- Bachman, L (2007) What is the construct? The dialectic of abilities and contexts in defining constructs in language assessment, in Fox, J, Wesche, M and Bayliss, D (Eds) *What are we measuring? Language testing reconsidered*, Ottawa: University of Ottawa Press.
- Cambridge ESOL (2008) *Research Notes* 33, retrieved from www.cambridgeesol.org/rs_notes/rs_nts33.pdf
- Chalhoub-Deville, M (2003) Second language interaction: current perspectives and future trends, *Language Testing* 20/4, 369–383.
- Coste, D (2007) *Contextualising uses of the Common European Framework of Reference for Languages*, paper presented at Council of Europe Policy Forum on use of the CEFR, Strasbourg 2007, retrieved from: www.coe.int/T/DG4/Linguistic/Source/SourceForum07/D-Coste_Contextualise_EN.doc
- Council of Europe (1954) *European Cultural Convention*, Strasbourg: Council of Europe, retrieved from: <http://conventions.coe.int/Treaty/EN/Treaties/Html/018.htm>
- (2001) *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*, Cambridge: Cambridge University Press.
- (2006) *Survey on the use of the Common European Framework of Reference for Languages (CEFR). Synthesis of results*, Strasbourg: Council of Europe, retrieved from: www.coe.int/T/DG4/Linguistic/Source/Surveyresults.pdf
- (2008a) *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR) A Manual*, Strasbourg, Council of Europe.
- (2008b) *Recommendation CM/Rec(2008)7 of the Committee of Ministers to member states on the use of the Council of Europe's Common European Framework of Reference for Languages (CEFR) and the promotion of plurilingualism*, Strasbourg: Council of Europe.
- Cummins, J (2000) BICS and CALP, in Byram, M (Ed.) *Encyclopedia of language teaching and learning*, London: Routledge, 76–79.
- Fulcher, G (2008) Testing times ahead? *Liaison Magazine* 1, July 2008.
- Jones, N (2000) Background to the validation of the ALTE Can-do Project and the revised Common European Framework, *Research Notes* 2, 11–13, Cambridge: Cambridge ESOL.
- (2001a) The ALTE Can-do Project and the role of measurement in constructing a proficiency framework, *Research Notes* 5, 5–8, Cambridge: Cambridge ESOL.
- (2001b) Relating the ALTE Framework to the Common European Framework of Reference, in Council of Europe, 167–183.
- (2006) Assessment for Learning: the challenge for an examination board, in Oldroyd, R (Ed.) *Excellence in Assessment. Assessment for Learning A supplement to the Cambridge Assessment Network Assessment for Learning Seminar held on 15 September 2006 in Cambridge, UK*, Cambridge: Cambridge Assessment.
- Jones, N and Saville, N (2009) European language policy: assessment, learning and the CEFR, *Annual Review of Applied Linguistics* 29, 51–63.
- Lenz, P and Schneider, G (2004) Introduction to the bank of descriptors for self-assessment in European Language Portfolios, Strasbourg: Council of Europe, retrieved from: www.coe.int/T/DG4/Portfolio/documents/Introduction_descriptor.doc
- Little, D (2007) The Common European Framework of Reference for Languages: Perspectives on the Making of Supranational Language Education Policy, *The Modern Language Journal* 91/4, 645–653.

- Martyniuk, W and Noijons, J (2007) *Executive summary of results of a survey on the use of the CEFR at national level in the Council of Europe Member States*, Strasbourg: Council of Europe, retrieved from www.coe.int/T/DG4/Linguistic/Source/Survey_CEFR_2007_EN.doc
- McNamara, T and Roever, C (2006) *Language Testing: The Social Dimension*, Oxford: Blackwell.
- Mislevy, R J (1992) *Linking Educational Assessments Concepts, Issues, Methods, and Prospects*, Princeton: Educational Testing Services.
- Papp, S (2009) Development of Can-do statements for KET and PET for Schools, *Research Notes* 36, 8–12, Cambridge: Cambridge ESOL.
- Pellegrino, J W, Chodowsky, N, and Glaser, R (Eds) (2001) *Knowing what students know: The science and design of educational assessment*, Washington, DC: National Academy Press.
- Poehner, M (2007) Beyond the test: L2 dynamic assessment and the transcendence of mediated learning, *Modern Language Journal* 91/3, 323–340.
- Teasdale, A and Leung, C (2000) Teacher assessment and psychometric theory: a case of paradigm crossing? *Language Testing* 17/2, 185–214.
- Tenberg, R (2009) Linking learning and assessment: Cambridge ESOL's blended learning approach, *Research Notes* 36, 20–22, Cambridge: Cambridge ESOL.

Development of Can-do Statements for KET and PET for Schools

SZILVIA PAPP RESEARCH AND VALIDATION GROUP, CAMBRIDGE ESOL

Introduction

Cambridge ESOL has developed new exams for learners under 15 years of age, focusing particularly on the 11–14 age group learning English in school. These new exams are variants of existing exams – KET and PET – at A2 and B1 levels of the Common European Framework of Reference (CEFR) and are called KET and PET for Schools. A focus of the research activity surrounding these exams has been the construction and validation of a set of can-do descriptors that are age-appropriate and that can be linked to CEFR levels through salient features of performance shared between learners aged 11–14 and adults. Can-do statements are empirically validated performance indicators.

In relation to these new exams, the long-term purposes of can-do descriptors for large examination boards such as Cambridge ESOL include:

- To inform item writing at task level for exams developed for candidates aged 14 and under.
- To act as a research tool to link exams for young learners up to the age of 14 to the CEFR and to other exams within Cambridge ESOL's suite of exams.
- To assist in providing validity evidence for exams developed for this age group.
- To create a link between summative and formative assessment.

The long-term purposes of can-do descriptors for candidates, teachers and other stakeholders include:

- To support learners and teachers in the ongoing monitoring of progress of learning towards exams developed for candidates aged 14 and under.
- To offer parents, schools, education authorities, ministries and other stakeholders transparent descriptions of the likely proficiency levels of candidates who have taken Cambridge ESOL's exams developed for this age group.

Context and rationale

Two general trends in education and in language testing have contributed to the development of criterion-based, action-oriented can-do statements as performance indicators in language teaching and assessment. In education, there has been a general, continuing movement towards criterion-based goals and interpretations around the turn of the millennium. This trend towards criterion-based educational goals is exemplified by the introduction of standards-based education in the USA, such as the No Child Left Behind policy (NCLB). In Europe, a similar trend is exemplified by the rise of the CEFR to prominence after it was widely distributed (Council of Europe 2001). In addition, in language assessment, the evolution of the concept of validity as a unitary concept contributed to an even deeper probing into the purposes, uses and interpretation of language test results. These trends together have led to the following questions and concerns:

- Does a particular test support inference to a candidate's performance in some "real world"? (relating to *outcomes* of learning)
- What happens in classrooms during the process of learning? (relating to emerging abilities and competences as a result of *inputs* to learning)
- How to relate these two aspects (inputs and outcomes)?

Can-do statements as performance indicators are seen to contribute to an understanding of these questions both in external assessment carried out by large examination boards such as Cambridge ESOL and in teacher assessment in the classroom.

In the development of can-do statements, the first task is to decide on the main purposes and audience of the performance descriptors. The methodology for development and validation of can-do statements depends on this initial decision. Once the main purpose and audience of can-do descriptors have been identified, a theoretical framework of

language learning and assessment needs to be identified. Cambridge ESOL has used the *socio-cognitive framework* (Weir 2005) for test development and evaluation purposes, as this framework offers a useful conceptualisation of language learning, use and assessment. It attempts to incorporate and link both the social and cognitive dimensions of language acquisition and learning. This is especially relevant in the case of the development of performance descriptors for 11–14-year-olds. This framework is seen as preferable to a simple *needs analysis* paradigm, as it does not restrict the underlying construct of language abilities that any test for a particular group of learners is intended to assess.

As a prerequisite to the development of performance descriptors for younger candidates, it was important to define the construct of 11–14-year-old candidates' language knowledge and use. In particular, consideration was given to:

- the identification of plausible target language use situations for young learners (social context)
- the interaction of cognitive (age-related) development and language proficiency level (cognitive dimension)
- the difference between natural development and typical formal learning routes (educational context).

For the definition of the construct of 11–14-year-old candidates' language knowledge and use, Cambridge ESOL has carried out a survey of the literature for descriptions of child and adolescent development from the cognitive, social, emotional and linguistic perspectives (Papp 2007, 2008a). Also, a study investigating the skills profiles across age groups was carried out in order to identify differences in performance in different age groups (8–9, 10–11, 12–14 and 15+ year-olds; Papp 2008b).

The next step was to consider and specify aspects of language competence laid out in the CEFR as suggested by its authors (North and Schneider 1998, North 2000, 2002) for the purposes of developing descriptors for 11–14-year-olds. The scales and descriptors listed in the CEFR were scrutinised for *relevance* (especially the domains, situations, themes, topics, notions, purposes/functions, and activities) and for *difficulty* (especially the conditions, constraints, mental contexts, activities, tasks, texts, processes, skills, strategies, and competences) to check their appropriacy for the age group.

The CEFR's model of language use incorporates:

'The actions performed by persons who as individuals and as social agents develop a range of *competences*, both *general* and in particular *communicative language competences*. They draw on the competences at their disposal in various contexts under various *conditions* and *constraints* to engage in *language activities* involving *language processes* to produce and/or receive texts in relation to *themes* in specific *domains*, activating those *strategies* which seem most appropriate for carrying out the tasks to be accomplished ...'

(Council of Europe 2001:9, emphasis in original)

This quote is relevant for the appreciation of the aim of the CEFR. All the perceived shortcomings of the CEFR, such as lacking a theory of language development (a *cognitive dimension*); being too imprecise for test construction purposes; having an imbalanced focus, a narrow scope and

narrow applicability; and its potential for misuse, are misplaced when one considers the suggestion by its authors that 'Users of the framework may wish to consider and where appropriate state' each of the following for the particular test purpose and target group they are working with (Council of Europe 2001):

- *context of language use* including domains, situations, conditions and constraints, mental contexts of participants
- *communication themes* including thematic areas/topics, sub-areas and specific notions the learners will need in selected domains
- *communicative tasks and purposes* the learners may need to perform in the various domains
- *communicative language processes* that may be required including planning, execution, reception and monitoring
- *texts* including media, text types
- *communicative language competences* including linguistic, sociolinguistic and pragmatic competences.

Cambridge ESOL has considered all the above aspects for the intended test purpose (English language tests to be used in the educational sector within schools) and target candidature (11–14 year-old children).

Methodology

Thus, given the background and rationale for developing can-do statements for young learners aged 11–14-years-old, the specific approach Cambridge ESOL has taken is the following. In order to address all audiences (test constructors, teachers, learners and other stakeholders), the methodology has included a number of stages:

1. Build on the CEFR (Council of Europe 2001) and the *Waystage* (van Ek and Trim 1998a) and *Threshold* (van Ek and Trim 1998b) learning objectives.
2. Adapt relevant qualitatively developed Bergen Can-do statements (Hasselgreen 2003) and validated ALTE and EAQUALS/ALTE Can-do statements (Jones 2002).
3. Select anchors from the Asset Languages project, i.e. seven descriptors for reading and listening that were found to be the most stable and reliable (Asset Languages 2008).
4. Complement the above with can-do statements from European Language Portfolios developed for young learners (Council of Europe 2009).
5. Supplement them with performance indicators related to syllabus content, learning skills, and language strategies adapted from benchmarking and standards frameworks such as the English Language Proficiency Benchmarks for primary learners in Ireland (IILT 2003a,b) or WIDA in the US (WIDA Consortium 2007).
6. Validate them on learners and teachers (get them endorsed by users), similar to what has been done for Asset Languages in their teacher assessment and self-assessment schemes.
7. Scale those that are scalable for common language proficiency scales, such as the CEFR or the Cambridge Common Scale of ability.

The original Council of Europe learning objectives have been revisited, that is, the coverage of *Waystage* and *Threshold* developed in the 1970s (van Ek and Trim 1998a, 1998b) was checked against the new exams developed for this age group. The advantage of these original learning objectives is that they contain more extensive lists of examples and exponents of performance than the more recent CEFR which built on them. However, the *Threshold* level (B1) was originally developed ‘to specify what kind of language an immigrant or visitor needed to operate effectively in society’ (North 2006:8), so the target group was different from 11–14-year-old children learning English in a school context. Also, as Alderson et al. (2004) found, there are minimal differences between *Waystage* (A2) and *Threshold* (B1) in terms of grammar, communicative activities, texts, and functions. Therefore, Cambridge ESOL carried out a mapping exercise to check coverage of the functions, notions and topics from *Waystage* and *Threshold* in the KET and PET exams to see what, if anything, should be changed for the new versions of these exams, KET and PET for Schools.

In addition to the CEFR scales and the original Council of Europe learning objectives *Waystage* and *Threshold*, available European Language Portfolios (ELPs) were surveyed. ELPs take account of aspects of formative assessment relevant to the age group under consideration. As Lenz and Schneider (2004) pointed out, different kinds of can-do descriptors have been used in ELPs:

- Scaled CEFR descriptors
- Adaptations of CEFR scaled descriptors for particular learner groups, e.g. young learners
- Context-specific achievement-oriented descriptors, e.g. related to syllabus content
- Descriptors which do not relate directly to a language proficiency scale but rather to independent categories such as learning strategies or descriptions of cultural or intercultural experiences.

These descriptors relate to the content to be learned, the skills to deal with learning, and the targeted outcomes in terms of capacity for language use. Lenz and Schneider (2004) recognise the relevance of context-specific, achievement- and syllabus-oriented descriptors for this age group:

‘Being so closely related to actual classroom practice, these descriptors may be tremendously useful in improving learners’ self-assessment skills and, if linked to an ELP in some way, enhance the perceived usefulness of that ELP. It is not surprising that descriptors of this type were produced in large numbers for use within or in combination with ELPs for children ...’ (Lenz and Schneider 2004)

However, when developing their own can-do statements, Cambridge ESOL heeded North’s advice on using existing statements: ‘Interpreted negatively one could say that in this way conventions and clichés get copied from scale to scale without an empirical basis’ (North 2000:182).

In order to validate the draft can-do statements derived through the process described above, the statements were trialled on a sample of 11–14-year-old KET and PET candidates and correlated with the candidates’ self-ratings on the set of can-dos (presented to them in English and

translated in their L1) and with their KET or PET exam grade. The third stage was to link the can-do self-ratings with teachers’ ratings for a typical student at the lower end, in the middle, and at the top end of their classroom, i.e. to come up with a list of Actually Do statements.

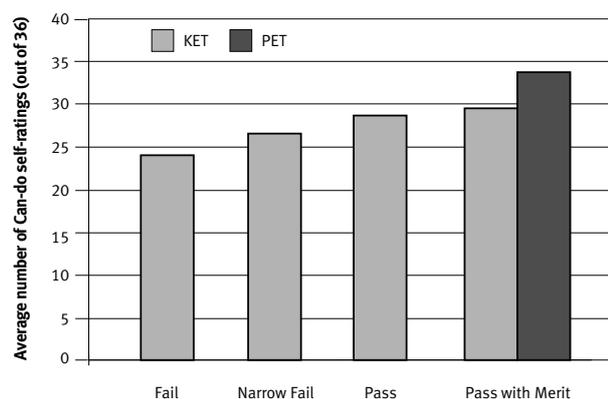
Trialling

Can-do responses were collected from live KET and PET candidates in May and June 2008. Out of a pool of over a 1,200 descriptors, thirty six can-do statements were selected for piloting with appropriate rewording including the seven ‘anchor’ statements from the Asset Languages project which were found to be the most stable indicators of level. The statements were translated into the mother tongue of the participating candidates (Argentinean and Mexican Spanish, Cantonese, Greek, and Italian). A total of 26 teacher assessments and a total of 827 candidate self-assessments were sent back. Responses were received from 19 teachers of KET and 7 teachers of PET and self-assessments were received from 609 KET candidates and 218 PET candidates. Performance data, teacher rankings and candidate self-assessments were available for a total of 195 students (190 KET and 5 PET).

Relation between exam grade and candidate self-assessment

Figure 1 shows the mean self-assessment of candidates grouped by the exam grade achieved; a clear relationship is evident between the two measures. For PET we only have 4 candidates who passed with Merit, and one achieved a Pass grade. This last candidate is not represented in the figure as his self-assessment cannot be taken as representative of all PET candidates. However, the mean self-rating of the 4 PET students who achieved Merit is indicated for reference in Figure 1.

Figure 1: Mean self-assessment of candidates by exam grade



Selected Can-do statements

Based on the results above, Tables 1 and 2 contain the can-do statements that were finally chosen to go in the KET and PET for Schools Handbooks and on the back of certificates. These are a mixture of existing can-do statements for KET and PET exams, deemed suitable for both age groups (adults and 11–14-year-old candidates),

Table 1: Handbook Can-do statements for KET for Schools

Typical abilities	Listening and Speaking	Reading and Writing
Overall general ability	CAN understand simple questions and instructions. CAN express simple opinions or requirements in a familiar context.	CAN understand straightforward information within a known area. CAN complete forms and write short simple letters or postcards related to personal information.
Social and Leisure	CAN have short conversations with friends about interesting topics. CAN make simple plans with people, such as what to do, where to go, and when to meet. CAN express likes and dislikes in familiar contexts using simple language.	CAN understand short simple messages from people who share his/her interests, for example emails, postcards or short letters from pen-friends. CAN write a very simple personal letter, note or email, for example accepting or offering an invitation, thanking someone for something, apologising.
School and Study	CAN understand basic instructions on class times, dates and room numbers. CAN ask the person to repeat what they said, when he/she does not understand something. CAN express simple opinions using expressions such as 'I don't agree'.	CAN understand the general meaning of a simplified text book or story, reading very slowly. CAN write about his/her daily life in simple phrases and sentences, for example family, school, hobbies, holidays, likes and dislikes.

Table 2: Handbook Can-do statements for PET for Schools

Typical abilities	Listening and Speaking	Reading and Writing
Overall general ability	CAN understand straightforward instructions or public announcements. CAN express simple opinions on abstract/cultural matters in a limited way or offer advice within a known area.	CAN understand routine information and articles. CAN write letters or make notes on familiar or predictable matters.
Social and Leisure	CAN understand the main points of TV programmes on familiar topics. CAN talk about things such as films and music and describe his/her reactions to them.	CAN understand factual articles in magazines and letters from friends expressing personal opinions. CAN write to his/her friends about the books, music and films that he/she likes.
School and Study	CAN understand instructions on classes and homework given by a teacher or lecturer. CAN repeat back what people say to check that he/she has understood. CAN give detailed practical instructions on how to do something he/she knows well.	CAN understand most information of a factual nature in his/her school subjects. CAN write a description of an event, for example a school trip. CAN take basic notes in a lesson.

and a subset of the new can-do statements developed and validated during the project described in this article.

Conclusion

In the future, the list of validated Can-do Statements for KET and PET for Schools will be added to, in order to develop a comprehensive set of can-do statements for an expanded age group to include 7–14-year-old candidates, the age range Cambridge ESOL's Young Learners English Tests (YLE) cater for.

The future development of can-do statements for younger candidates will need to use a different elicitation format as this much younger age group cannot reliably carry out self-assessment. Elicitation formats for YLE candidates may therefore include one-to-one elicitation, spoken rather than written elicitation, real-life demonstration of ability, or some other means of elicitation, including visuals, example tasks, etc. Existing ELPs designed for a very young age group, such as the Irish ELP (11.2001 rev.2004) and the CILT (70.2006) Portfolio for junior learners in Key Stage 2, as well as benchmarking frameworks such as the Irish *English language proficiency benchmarks for non-English-speaking pupils at primary level* and the WIDA *English language proficiency standards for English language learners in kindergarten through grade 12* will need to be consulted for young children. Ultimately, the aim is that Cambridge ESOL's can-do statements for Young Learners (including YLE, KET and PET for Schools candidates) will be developed in electronic format for flexibility, practicality and usefulness for candidates and teachers. That way the list of can-do statements can be expanded and modified according to Cambridge ESOL's evolving purposes. A learning management system seems to be the most suitable platform in the future for this development (see Tenberg 2009).

In order for this project to inform the construct to be tested underpinning KET and PET for Schools (for 11–14-year-old learners) at A2 and B1 levels within the CEFR as well as YLE which is targeted at 7–12 year-old children, the following research questions still need to be investigated:

1. What is the natural developmental growth (rate, route and ultimate attainment) in L2 English language proficiency in children aged 7–14? How can this be expressed in terms of can-do statements?
2. What are the developmental levels of language ability of children aged 7–14 in educational contexts as currently expressed in can-do statements?
3. Is the route of proficiency progression for young learners the same as for their adult counterparts as currently expressed in can-do statements?
4. What are the salient features of performance at each level of the CEFR (A1, A2, B1) that are shared between the group of candidates aged 14 and under and adults as currently expressed in sets of related can-do statements?

These, and other, questions will be explored in further studies which will continue to investigate the construct of younger learners' communicative competence and related impacts on learners, teachers and educational systems more widely.

References and further reading

- Alderson, J C, Figueras, N, Kuijper, H, Nold, G, Takala, S and Tardieu, C (2004) *The development of specifications for item development and classification within the common European framework of reference for languages: learning, teaching, assessment. Reading and listening. The final report of the Dutch CEF Construct Project*, Project Report, Lancaster University, Lancaster, UK, retrieved from <http://eprints.lancs.ac.uk/44>
- Asset Languages (2008) *Language Ladder 'Can Do' statements*, retrieved from: www.assetlanguages.org.uk/UserFiles/File/specialists/CanDoStatements2008.pdf
- Council of Europe (2001) *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*, Cambridge, Cambridge University Press.
- (2009) *European Language Portfolio*, retrieved from: www.coe.int/portfolio
- Hasselgren, A (2003) *Bergen "Can-do" Project*, Strasbourg: Council of Europe.
- Jones, N (2002) Relating the ALTE Framework to the Common European Framework of Reference, in Council of Europe (Eds) *Case Studies on the Use of the Common European Framework of Reference*, Cambridge: Cambridge University Press, 167–83.
- Lenz, P and Schneider, G (2004) *Introduction to the bank of descriptors for self-assessment in European Language Portfolios*, Strasbourg: Council of Europe, retrieved from: www.coe.int/T/DG4/Portfolio/documents/Introduction_descriptor.doc
- IILT (2003a) *English Language Proficiency Benchmarks for non-English-speaking pupils at primary level*, Dublin: Integrate Ireland Language and Training.
- (2003b) *English Language Proficiency Benchmarks for non-English-speaking students at post-primary level*, Dublin: Integrate Ireland Language and Training.
- (2004a) *European Language Portfolio: Learning the language of the host community, primary*, Dublin: Integrate Ireland Language and Training.
- (2004b) *European Language Portfolio: Learning the language of the host community, post-primary*, Dublin: Integrate Ireland Language and Training.
- North, B (2000) *The development of a common framework scale of language proficiency*, New York: Peter Lang.
- (2002) Developing descriptor scales of language proficiency for the CEF common reference levels, in Alderson, J C (Ed.) *Case Studies in applying the Common European Framework*, Strasbourg, Council of Europe, 87–105.
- North, B and Schneider, G (1998) Scaling descriptors for language proficiency scales, *Language Testing* 15/2, 217–262.
- Papp, S (2007) *The Cambridge YLE tests in the light of cognitive, linguistic, educational, social-psychological and cultural aspects of children's L2 development and the CEFR*, Cambridge ESOL internal Research and Validation report.
- (2008a) *Factors influencing L2 development and use in the 8–14 age group – towards defining the construct*, Cambridge ESOL internal Research and Validation report.
- (2008b) *Skills profiles by age in KET and PET*, Cambridge ESOL internal Research and Validation report.
- Tenberg, R (2009) Linking learning and assessment: Cambridge ESOL's blended learning approach, *Research Notes* 36, 20–22, Cambridge: Cambridge ESOL.
- van Ek, J A and Trim, J L M (1998a) *Threshold 1990*, Cambridge, Cambridge University Press.
- (1998b) *Waystage 1990*, Cambridge, Cambridge University Press.
- Weir, C (2005) *Language testing and validation*, Basingstoke, Palgrave Macmillan.
- WIDA Consortium (2007) *WIDA's CAN DO Descriptors*, retrieved from: www.wida.us/standards/CAN_DOs/index.aspx

Adapting testing materials for younger learners: developing KET and PET for Schools exams

ED HACKETT ASSESSMENT AND OPERATIONS GROUP, CAMBRIDGE ESOL

Introduction

In March 2009, KET for Schools and PET for Schools examinations were launched. These versions of KET and PET, aimed at a younger candidature, are identical in format to the existing KET and PET exams, but have content deemed more appropriate for this particular age group. This article outlines the processes involved in adapting the materials in KET and PET tasks for younger learners.

The need for a version of KET and PET for younger learners

The candidature for KET and PET has grown rapidly over the last decade. Since 2000, the growth in KET has been over 100% and PET candidature has grown by nearly 75%. In addition to this growth, there has been a gradual change in the nature of the candidature, with a year-on-year trend

of younger and younger candidates taking these exams. In 2008, nearly 75% of KET candidates were aged 14 or under and 28% of PET candidates were also in this age bracket. Improvements in the teaching of English worldwide have led to similar trends in other examinations. Whilst a significant proportion of KET and PET candidates may be under the age of 15, there are still many late teenage and adult candidates. PET is particularly popular in certain countries in the university sector, counting towards admission or matriculation. KET and PET are also used in key industries in other countries, for example by oil companies in Libya. Given this wide range of ages, it has been increasingly difficult to construct question papers with topics that relate to the interests and experience of candidates at both ends of this spectrum. In 2007, Cambridge ESOL set up a review group to look at the needs of the 8 to 14 age group, with particular reference to our provision of Young Learners of English (YLE) exams

(Starters, Movers and Flyers) and KET and PET. This review entailed extensive consultation, involving centres, supplier schools and other key stakeholders. The resulting findings outlined a clear need for the development of versions of KET and PET aimed at a younger audience, whilst retaining the existing format for older candidates.

Prior to developing materials for the new versions of KET and PET, certain key questions had to be looked into:

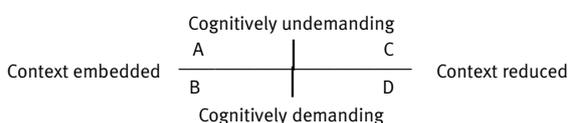
1. What age range should the new versions be targeted at?
2. Is the current format suitable for younger learners?
3. How could topics be adapted to appeal more to this age group?
4. Are any other changes to the question papers desirable, e.g. graphics and layout?

Appropriate age range

A literature review was carried out in 2007–8 (Papp 2008a) and the various factors affecting the cognitive development of young learners were investigated. Inhelder and Piaget (1958) suggested that the beginning of the development of formal operational thought begins towards the end of primary schooling, this being the final in four stages of a child's cognitive development. Whilst other studies (Bruner 1966, Vygotsky 1978) challenged the idea that these stages were purely biological and argued that children could learn from instruction and tailored appropriate support in interaction, there is a general belief that from the age of 11, children are 'beginning to develop the ability to "manipulate" thoughts and ideas' (McKay 2006:7). McKay (ibid.) goes on to state that between the ages of 11–13 children's 'use of language has expanded to enable them to predict, hypothesize and classify' and that by this age 'children are able to read a variety of fiction and non-fiction and, importantly, to develop critical literacy skills'. This ability to handle a variety of texts would be crucial to enable learners to cope with the text types employed in KET and PET.

In addition to the literature review, Cambridge ESOL looked into the current performance of KET and PET candidates in the 10–14 age group, compared to their older counterparts. This analysis (Papp 2008b), based on performance in 2007 KET and PET sessions, revealed that the younger candidates generally outperformed the older cohort in the Listening and Speaking components, with similar performance levels in Reading and Writing. This, however, should not be a surprising statistic, given that many of those taking KET and PET at this younger age were probably high performers as regards language learning ability. So, whilst these younger learners could obviously cope with the task types they were presented with, it is possible that they may have performed even better if presented with texts and content more appropriate to their age group. The need for 'scaffolding' as outlined by Cummins (2001:144), shown in Figure 1, is thought to be

Figure 1: Scaffolding strategies based on Cummins (2001)



crucial to successful performance in language based tasks, especially for younger learners.

Where tasks are cognitively demanding, it is essential that there is sufficient embedding of the context. Many tasks in KET, and some in PET, contain examples as guidance to the completion of tasks. The computer-based variants of these exams contain a short tutorial which shows candidates how to answer the questions in each part. The rubric is also a key element in providing support to candidates and in setting the scene. Not only does the rubric explain to the candidate how to answer a set of questions, it also presents a life-like context, for example, 'Listen to Nick talking to a friend about his birthday presents. What present did each person give him?' (Listening Part 2, Cambridge ESOL 2008).

Following the above research and analysis, it was decided that no changes were necessary to the format of the KET and PET examinations, but that the content of the new KET for Schools and PET for Schools exams would be targeted primarily at the 11–14 age group. However, as children develop at different speeds, no lower or upper age restrictions would apply to the new exams. Preparation for both versions of the exams (KET and KET for Schools or PET and PET for Schools) could thus be the same and the decision as to which exam to enter for would be left to the candidates, parents and schools.

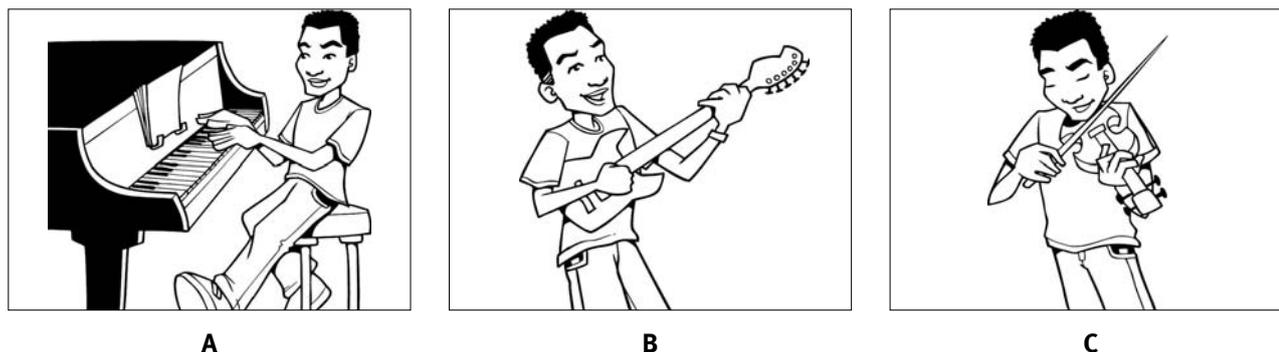
Adapting content for younger learners

The KET and PET exams have strong links to the learning objectives outlined in the Council of Europe's *Waystage* and *Threshold* documents (van Ek and Trim 1998a, 1998b). *Waystage* is at Level A2 in the Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR; Council of Europe 2001) and *Threshold* is at Level B1. Cambridge ESOL publishes language specifications for both exams based on *Waystage* and *Threshold* in the relevant handbooks for teachers (Cambridge ESOL 2008a, 2008b). Papp (2008c) investigated the relationship between the learning objectives in *Waystage* and *Threshold* and those in KET and PET and concluded that '... very few *Waystage* and *Threshold* functions, notions or topics were found inappropriate or unsuitable for KET and PET for Schools.' It is not the topics themselves that might be unsuitable but a particular aspect of that topic that might have an adult focus. If we look at the topic of food and drink, we would not expect younger candidates to be paying the bills in restaurants or arranging dinner parties, but they will probably have experience of buying drinks and snacks, and be able to express their likes, dislikes and preferences. It is therefore important that the topics lend themselves to being seen through the eyes and from the experiences of younger learners.

Once the target age group for KET for Schools and PET for Schools had been identified, consultants with experience of writing items for KET and PET were engaged to produce a list of topics for each 'for Schools' exam paper, outlining suitable and unsuitable aspects of each topic and suggesting a range of sources for locating suitable texts e.g. teen magazines, adverts, radio and TV programmes etc. The

Figure 2: Example of new graphics style for PET for Schools Listening Part 1

4 Which instrument does the boy play now?



suggested lists were then edited by the Chairs of each item writing team and added as appendices to the item writing guidelines. Training sessions were then held for item writers to assist them in locating and adapting texts for future commissions.

In addition to the sourcing and adaptation of texts for KET for Schools and PET for Schools, work was also carried out on designing the look and feel of the new exams. Whilst it was important to maintain the same rubric and general layout of the tasks, it was felt necessary to identify a style of graphic that was more appealing to a younger candidature. After several iterations of designs, it was possible to identify a style of graphic that was more appropriate for this younger audience, yet avoided alienating them by appearing too juvenile in style. It was important to find a style that sat between the current design for our Young Learners of English (YLE) exams (Starters, Movers and Flyers) and that of the existing versions of KET and PET. Once the style was agreed upon, it was trialled alongside sample materials for the new exams in April and May 2008. Trialling was carried out via questionnaires and focus groups for teachers and students in Argentina, China, Greece, Italy and Mexico. Feedback from trialling (Mayes 2008) was very positive, with the majority feeling that 'the overall design was either very good or good, and their reaction to the pictures was very positive or positive', rating each aspect 1 or 2 on a five point scale.

Pretesting and trialling of newly commissioned tasks were then carried out ahead of test construction and question paper preparation activities. Objective tasks were pretested alongside anchor tasks common to both versions of the exams to ensure comparability in the difficulty of the two formats. Work was also carried out to develop new Can-do statements more appropriate to a younger candidature (see Papp 2009, this issue).

Conclusion

Candidates in and around the target age group of 11 to 14 years should be able to cope with the format and tasks in the new KET for Schools and PET for Schools exams. Whilst the content matter has been adapted to suit the interests

and experiences of this younger age group, the difficulty of the exams is comparable to that of the existing KET and PET exams and candidates should be able to carry out the vast majority of language functions associated with *Waystage* and *Threshold* levels. KET for Schools and PET for Schools were launched worldwide in March 2009.

References and further reading

- Bruner, J S (1966) *Toward a Theory of Instruction*, Cambridge, MA: Belkapp Press.
- Cambridge ESOL (2008a) *Key English Test for Schools – Handbook for Teachers*, Cambridge: Cambridge ESOL.
- (2008b) *Preliminary English Test for Schools – Handbook for Teachers*, Cambridge: Cambridge ESOL.
- Council of Europe (2001) *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*, Cambridge: Cambridge University Press.
- Cummins, J (2001) Tests, achievement and bilingual students, in Baker, C and Hornberger, N (Eds) *An Introductory Reader to the Writings of Jim Cummins*, Clevedon, Avon: Multilingual Matters.
- Inhelder, B and Piaget, J (1958) *The growth of Logical Thinking from Childhood to Adolescence: An essay on the construction of formal operational structures*, London: Routledge and Kegan Paul.
- Mayes, H (2008) *Research into the appearance and design of the question paper materials for KET and PET for schools*. Cambridge ESOL internal report.
- McKay, P (2006) *Assessing Young Language Learners*, Cambridge: Cambridge University Press.
- Papp, S (2008a) *Factors influencing L2 development and use in the 8–14 age group – towards defining the construct*, Cambridge ESOL internal Research and Validation report.
- (2008b) *Skills profiles by age in KET and PET*, Cambridge ESOL internal Research and Validation report.
- (2008c) *Coverage of Waystage and Threshold learning objectives within KET and PET and KET and PET for Schools*, Cambridge ESOL internal Research and Validation report.
- (2009) Development of Can-do Statements for KET and PET for Schools, *Research Notes* 36, 8–12, Cambridge: Cambridge ESOL.
- van Ek, J A and Trim, J L M (1998a) *Waystage 1990*, Cambridge: Cambridge University Press.
- (1998b) *Threshold 1990*, Cambridge: Cambridge University Press.
- Vygotsky, L S (1978) *Mind in society*, Cambridge, MA: Harvard University Press.

Validating a worldwide placement test for German

ARDESHIR GERANPAYEH RESEARCH AND VALIDATION GROUP, CAMBRIDGE ESOL

SIBYLLE BOLTON CONSULTANT TO GOETHE-INSTITUTS

Background

The Goethe-Institut is a non-profit-making, publicly funded organisation with its head office in Munich. It was founded in 1951 to promote a wider knowledge of the German language abroad and to foster cultural co-operation with other countries. As an organisation with over 140 centres in 80 countries serving over 170,000 students, the Goethe-Institut plays an important role in providing access to German language and culture all over the world.

Besides organising cultural events and offering language courses, the Goethe-Institut offers German language examinations ranging from levels A1 to C2 of the *Common European Framework of Reference* (CEFR; Council of Europe 2001, see Europarat 2001 for German translation). All examinations are produced centrally at the head office of the Goethe-Institut in Munich. The first examinations to be developed were examinations for advanced learners and over the last forty years new examinations have been introduced and existing examinations have been revised on a regular basis. In 1990 the Goethe-Institut became a member of ALTE (the Association of Language Testers in Europe) and the discussions within ALTE on test development and quality management have considerably influenced this continuous revision process.

Need for a new placement test

Whereas the examinations were always developed centrally, the placement tests for the language courses in the Goethe-Institutes worldwide were developed locally. The reason for this was that the structure of the language courses differed from country to country which did not allow for a centrally developed placement test. After the publication of the German translation of the *Common European Framework of Reference* in 2001, the Goethe-Institut decided to harmonise the structure of its language courses and to align them worldwide to the levels of the CEFR. As a consequence the existing placement tests became obsolete and the colleagues in head office responsible for test development were asked to develop a new placement test which would place learners on the levels A1 to C2. This placement test was to be offered in two modes: a paper-based test (PB) and a computer-based test (CB).

The Goethe-Institute decided to develop this new placement test together with Cambridge ESOL, because since becoming a member of ALTE the two organisations have participated in several testing projects such as the development of the computer adaptive placement test (CommuniCAT) in several major European languages. Cambridge ESOL won the European Academic Software Award for the latter in 2000. Cambridge ESOL holds the German test items in its Local Item Banking System (LIBS). It was decided to draw on calibrated German items from

LIBS for the development of the new German Placement test.

Test design and construct

A placement test should be short and be able to be scored quickly and reliably. It was therefore decided to develop a test with 60 items testing grammar/vocabulary and reading comprehension. After the trial, however, it was decided to increase the PB test length to 70 items to achieve higher discriminatory power for some of the band levels. For reasons of practicality we did not include tasks on listening comprehension in the paper-based version of the placement test, because on the days of the course inscriptions at many Goethe-Institutes hundreds of new learners have to take the PB placement test and offering a listening section as part of the test would not be possible technically. Listening comprehension is tested – together with speaking – in an interview which is an obligatory part of the placement procedure. For reasons of ease and speed of scoring we also did not include a writing task. The placement test (without the interview) lasts 40 minutes and the scoring of the 70 items takes five minutes.

The choice of the items for the paper-based test from LIBS was influenced by two factors: the difficulty (and other statistical properties) of the item and the suitability of its content (topic, grammar, lexis), based on the content of the language courses in the Goethe-Institutes.

The paper-based placement test is offered in three versions, the third version being a mix of the first two versions. The computer-based version is an adaptive placement test on CD-ROM which draws on the item bank in LIBS. Besides testing reading comprehension and grammar/vocabulary, it also tests listening comprehension.

The placement test consists of different types of multiple-choice questions as shown in Figure 1.

Table 1 shows the span of item difficulties and item types for each of the six levels from A1 to C1.

Methodology

Five inter-related, quality assurance issues had to be addressed before the release of the GPT. Firstly, the reliability of the tests had to be determined to achieve a satisfactory level of reliability (as indicated by an Alpha of above 0.85). Next, the parallel use of the two modes of delivery had to be investigated carefully if piloting indicated that the CB and PB versions rank-ordered candidates differently. This would have impacted on the administrative procedures to be adopted. Thirdly, score reporting had to provide maximally useful information to Goethe-Institut centre teachers so that the Council of Europe 6-level system (CEFR) could be defined alongside the then ALTE 5-level

Figure 1: Three sample test items

1.	Bitte teilen Sie uns telefonisch (089/ 2714977) oder per beiliegender Antwortkarte bis zum 9. Oktober mit, ob Sie an diesem Kurs teilnehmen möchten.	A Der Kurs findet am 9. Oktober statt. B Man kann sich nur telefonisch anmelden. C Man muss bis zum 9. Oktober antworten.
...

2. Neues Sportzentrum

Schon lange haben die Berliner auf ihr neues Sportzentrum gewartet. Nun ist (6) endlich soweit. Mit (7) Tag der offenen Tür feiert der Sportverein Charlottenburg seine neuen Turnhallen und Tennisplätze. Am Sonntag (8) 9,30 Uhr haben alle interessierten Bürger die Möglichkeit, die Hallen und Tennisplätze (9) . Wer möchte, kann auch selbst auf den neuen (10) spielen.

6	A das	B es	C sie
7	A ein	B einem	C einer
8	A ab	B seit	C von
9	A anschauen	B anschaut	C anzuschauen
10	A Tennisplätze	B Tennisplätzen	C Tennisplatz

3.

68 Leider waren die Ferien schon zu Ende, sonst ich mit den Kindern noch länger geblieben.
A hätte B wäre C werde D würde

69 In den letzten 50 Jahren haben wir mehr Wissen und Informationen angesammelt als im ganzen Jahrhundert
A damals B davor C einst D früher

70 Die Eröffnung des neuen Messegeländes ist auf den 1. Oktober worden.
A fertiggestellt B festgelegt C festgesagt D festgestellt

Table 1: Item types and item difficulties for each level of the German Placement Test

Level	Item Type	Difficulty*
A1	5 texts + 1 MCQ each	25-35
A2	3 gapped texts + 5 MCQ each	Average diff. of texts 40-49
B1	20 MCQ	45-67
B2	10 MCQ, 1 gapped text + 5MCQ	67-77; Average diff. 70
C1	10 MCQ, 1 gapped text + 5 MCQ	72-80; Average diff. 74

*difficulties are reported on a scale of 0-100 linked to the CEFR levels MCQ = multiple-choice question

system. In addition a scaled score had to be devised to indicate the relative position of candidates within the same level. The error term for both the CB and PB tests would have to be identified prior to the release of the test. Lastly, the effects of first language and country of origin were also to be identified through piloting so that suitable recommendations could be made as to how to interpret scores for specific language groups.

To address the above issues a validation project was set up where a number of Goethe-Institut centres around the world piloted three parallel versions of a paper-based German Placement Test and a CB version. Three versions of the paper-based test and the adaptive version of the computer-based test were administered to over 90 candidates in 12 centres around the world. Trialling of the PB versions involved 1963 candidates from 15 centres, almost half of whom (44%) were in Italy and Estonia. Figure 2 shows the spread of GPT trial candidates who took the PB test by country.

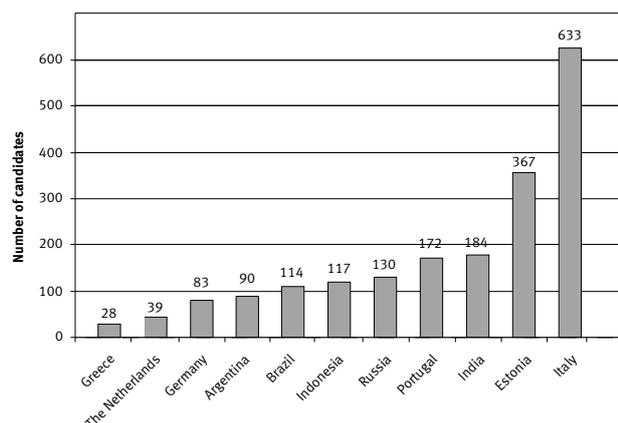


Figure 2: Distribution of PB candidates by country (N = 1957)

Both PB versions 1 and 2 were taken by more than 800 candidates each; 452 candidates did both versions 1 and 2 whilst Version 3 was taken by 298 candidates (see Table 2). Note that all PB versions were given out in both a 40-item and a 60-item version; candidates did one or both of these versions.

Table 2: Candidates by centre for PB version 3

Centre	No. cands
Milan	86
Bangalore	184
Athens	28
Total	298

The CB version was taken by 87 candidates from 3 centres in Italy, India and Greece, all of whom also took PB Version 3, as shown in Table 3.

Table 3: Candidates by centre for the CB version

Centre	No. cands
Milan	25
Bangalore	32
Athens	30
Total	87

Results and discussion

Tables 4 and 5 show the reliability and standard error of measurement for both PB and CB versions that were trialled. The figures reported here are all respectable and are well above the minimum standard set out in the research design.

Table 4: Reliability (Alpha) and SEM for PB versions

	Reliability			SEM		
	V1	V2	V3	V1	V2	V3
40-item version	0.92	0.92	0.94	2.58	2.46	2.57
No. candidates	863	800	300	863	800	300
60-item version	0.90	0.91	0.91	3.15	3.02	3.00
No. candidates	324	313	43	324	313	43

It is worth noting that the paper-based test is administered in two phases. The candidates first receive the 40-item test. If they score above a certain level, they are allowed to do the second part of the test which is progressively more difficult. This is to ensure that low level candidates are not unduly exposed to a long difficult test which may not provide additional useful information about such candidates.

The reliability for the CB version has been calculated using a Rasch reliability estimate (an internal consistency measure, analogous to Cronbach's alpha), see Table 5.

Table 5: Scale statistics and reliability estimates (Rasch) of the CB version

	Total
Average Length	28
SEM (RMSE)	0.44
Adj. SD	1.91
Separability	4.34
Reliability	0.95

SEM: Standard error of measurement

Adj. SD: Standard Deviation of candidate's ability adjusted

Separability: the degree of separability between ability levels

The following terminology will be used with reference to the CB scores: Test Score refers to CB test score (0–100), Band Score refers to CEFR band levels (A1–C2), and Ability Level refers to candidates' ability as estimated by Rasch model (Logit).

It is important to mention that test scores on the CB version do not refer to the raw scores; they are actually ability estimates derived from a latent trait (Rasch) analysis, converted into CB scores by means of a scaling procedure. The items in the CB test were taken from one item bank with pre-calibrated item difficulties.

Table 6 and Figure 3 show how closely the two modes of placement tests functioned.

The high correlation between the two modes of test delivery allows us to be relatively confident that the tests can be used interchangeably. Table 7 shows the conversion table for comparing the two test results.

Table 6: CB and PB score correlation with outliers removed

Correlation Total Score	0.84
Correlation Band levels	0.86

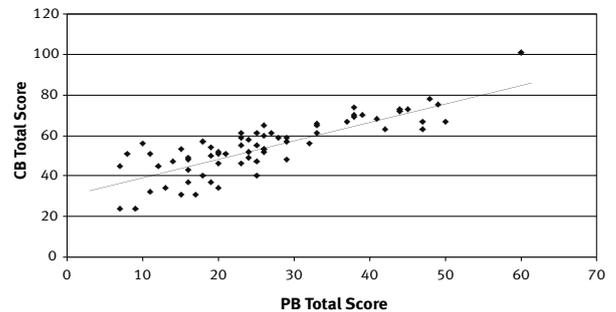


Figure 3: CB and PB total score comparisons with outliers removed

Table 7: Lookup table for the 70-item German Placement Test

ALTE Stufen	Niveau-bezeichnung	Teil 1 (40 Item)	Teil 1+2 (70 Item)	Europäisches Referenzniveau
---	Anfänger	0-10	0-10	---
0	Breakthrough	11-18	11-19	A1
1	Grundstufe I	19-28	20-34	A2
2	Grundstufe III	29-34	35-45	B1
3	Mittelstufe II	(35-37)	46-54	B2
4	Mittelstufe III	Teilnehmer, die nur Teil 1 des Tests gelöst haben und die 29 Punkte oder mehr erreicht haben, sollten auch Teil 2 lösen (die Punkte werden dann in der Spalte 4 abgelesen).	55-62	C1
5	Oberstufe		63-70	C2

Conclusion

In this short article we reviewed the need for a standardised new German placement test for the Goethe-Institut centres. We explained how paper-based and computer-based versions of a two-tier test of 40 and 60 items were constructed to report scores on CEFR levels. Based on the test analysis and recommendation to improve test accuracy for certain band levels the final version of the test length increased to 70 items. The statistics reported in this article for both CB and PB versions of the test provide evidence of the accuracy of the placement test for measuring candidates' German language proficiency prior to taking a language course. It appears that the model of a two-tier paper-based test proposed in this report would be appropriate for a number of language schools which have to run a quick placement test for all the candidates who enrol on their classes without exposing the low level candidates to unnecessarily difficult items which may not provide more useful information about each candidate's proficiency level. We have also demonstrated that an adaptive computer-based placement test could complement the paper-based test and provide a very accurate picture of a candidate's language ability whereby only one test may need to be

administered. The high correlation between the scores of the two modes in this report provides evidence that the two tests are comparable and can be used alongside each other. This would allow much more flexibility and consistency in the administration of such tests where for various practical reasons the administration technology may not be uniform around the world. We hope to be able to report on the use of the tests in future issues of *Research Notes*.

References

- Council of Europe (2001) *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*, Cambridge: Cambridge University Press.
- Europarat (2001) *Gemeinsamer europäischer Referenzrahmen für Sprachen: Lernen, lehren, beurteilen*, Berlin: Langenscheidt

Linking learning and assessment: Cambridge ESOL's blended learning approach

REINHARD TENBERG PROGRAMME OFFICE, CAMBRIDGE ESOL

Introduction

Cambridge ESOL's core business is to provide tests and examinations. However, differences between formative and summative assessment are increasingly being eroded. The market for e-learning platforms and the growth of e-learning and blended learning are blurring the boundaries between formative and summative assessment. The recently coined term *Assessment for Learning* (AfL) which is now often used for formative assessment in the British educational context reflects this and has acquired a programmatic and political dimension:

'A clear distinction should be made between *assessment of learning* for the purposes of grading and reporting ... and *assessment for learning* which calls for different priorities, new procedures and a new commitment. In the recent past, policy priorities have arguably resulted in too much attention being given to finding reliable ways of comparing children, teachers and schools.' (The Assessment Reform Group 2002:2)

The literature on AfL is extensive and this area has been the subject of a number of research studies.¹ Black et al. (2003:10) define formative assessment or AfL as 'any assessment for which the first priority in its design and practice is to serve the purpose of promoting students' learning ... an assessment activity [is formative] if it provides information to be used as feedback by teachers and their students'. It has been shown that learning takes place more effectively when:

- All learners are actively involved in the learning process.
- Learners understand their learning objectives.
- Assessment emphasises progress and achievement rather than failure; comparison with higher achievers is unlikely to motivate learners.
- Dialogue between teachers and students is encouraged and facilitated.
- Careful attention is paid to learners' motivation through constructive feedback and support.
- Learners are enabled to see their progress, thus building confidence and self-esteem.

- Learners are encouraged to engage in self- and peer-assessment with the aim of becoming reflective and autonomous learners.
- Learners are made aware of their own learning styles.

Formative assessment can take place in the classroom but can equally be promoted through e-learning. From 2009 onwards, Cambridge ESOL is complementing its core business with a range of online language courses in a blended learning format in order to assist students and teachers in preparing for their examinations. In this article we describe the approach which Cambridge ESOL has taken towards the development of blended learning courses which complement and support our examination products. We analyse how some of the above key characteristics are reflected in the complex interface between a learning platform and a bespoke blended learning programme.

What is blended learning?

The term 'blended learning' is a relatively new one in the education and training market. It combines traditional classroom learning with a media-rich, interactive online technology in order to create an optimum training programme for a specific audience. Blended learning recognises that we all have different learning styles and preferences regarding how and when we want to learn by providing an online course component which is available independent of time and space. It also encourages learner autonomy and allows students to focus on their perceived weaknesses. For example, whilst a learner may have achieved an advanced level of proficiency in their receptive skills they may want to spend more time on practising their productive skills. In an English for Specific Purposes (ESP) course, such as Financial English, the learner may choose to focus on topics which they perceive to be difficult when preparing for an examination and choose their learning path accordingly. Although the majority of e-learning programmes may be structured chronologically, the learning platform renders the courses menu-driven, thus allowing the learner to choose their own pathway through the course.

1. See, for example, *The 10 principles: Assessment for Learning*, www.qca.org.uk/qca_4336.aspx

Cambridge ESOL's learning platform

Today's Learning Management Systems (LMS), sometimes also referred to as VLE (Virtual Learning Environments) or MLE (Managed Learning Environments), all share a range of similar tools and functionality. These include three main components, described below:

Communication and collaboration tools

- asynchronous tools such as email, messaging, forums or discussion groups
- synchronous tools including VoIP (Voice over Internet Protocol, i.e. communication on the Internet in real time), whiteboards, application share, video conferencing – all of which enable virtual classroom teaching in real time and allow teachers and students to emulate all aspects of a face-to-face classroom environment including peer assessment, group work and the building of a learner community.

Content management tools

These enable:

- teachers and students to store and share their materials online
- course authors to use the system as a repository for multimedia resources and store reusable course content
- administrators to authorise or restrict access to courses and repositories.

Training or course tools

These enable users to:

- author and maintain a course using proprietary authoring tools

- upload externally produced courses²
- publish and run a course
- set, send and mark students' assignments online
- assess and monitor performance.

One of the advantages of using Cambridge ESOL's sophisticated LMS is the ability to configure the very broad range of tools according to the end user's requirements. Teachers, administrators and students all access the course for different reasons, therefore they will require access to different tools and functions. The student's view of the course is shown here which will differ from a teacher's view where one would find additional tools such as a Teacher Resources folder with face-to-face materials or a Portfolio tool for monitoring students' work. See Figure 1 for options available to students.

The Cambridge Financial English course

The Cambridge Financial English course (CFE) is a blended learning course, 75% of which is delivered online and 25% through face-to-face teaching, although other combinations can be tailored by the tuition provider to the needs of their students or employees. The course provides learners with an interactive learning environment to help them develop their Financial English language skills. It consists of 20 thematically-organised modules, each of which is further sub-divided into five units. The introductory unit (overview) provides the context for the topic, both in terms of setting

2. These must be SCORM 2004 (Sharable Content Object Reference Model) compliant, which means they can be shared and reused on different LMS platforms (see Advanced Distributed Learning website).

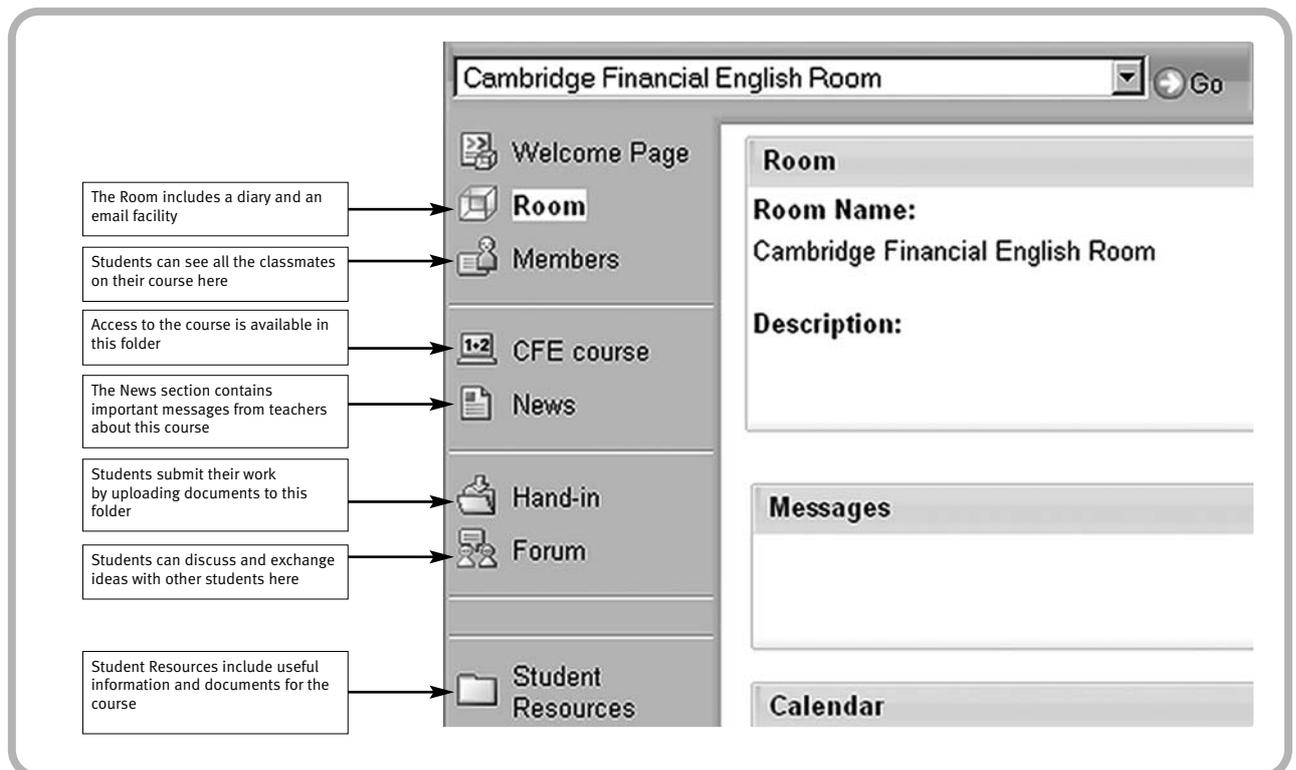


Figure 1: Cambridge Financial English course VLE options

the scene for a case study and introducing the learner to key language and vocabulary:

'The case study for this module focuses on an accounting firm that is recruiting a new accountant. Learn all about the background to the firm by clicking *Problem background* and *Problem strategy* activities. Watch the news to learn what's happening in the job market. To understand more about what different people in finance and accounting do, read the *Introduction* text. *Introductory vocabulary* and *Introductory language* activities will give you the English you'll need before you work through the module activities.' (Module 1, CFE)

Having introduced the learner to the topic of recruitment in Unit 1, the remainder of the units then focus on each of the four skills. For example, the writing unit looks at constructing and proof-reading CVs and in the listening unit the learner watches and evaluates job interviews.

The learning target of CFE is level C1 on the CEFR, and therefore the assumption is that the learner will be entering the course at a level around B2 making sufficient progress over the course to pass Cambridge ESOL's International Certificate in Financial English (ICFE) at the end of it.

Common criticism levelled against 100% online courses and blended learning courses alike are that:

- They do not clearly state objectives of individual learning modules, thus making it difficult for the learner to decide on an individual learning path.
- They provide meaningless and decontextualised feedback on learners' errors and progress.
- They do not sufficiently motivate and support the autonomous learner.
- In the so-called 'blended learning' programmes the face-to-face element is a loose bolt-on rather than an integrated part of the course.

This appraisal may be a little harsh but the CFE course offers examples of how some of these problems may be avoided. We discuss below how the CFE measures up against some of the main criteria for effective learning

referred to in the AFL literature. We will focus on three aspects of the course:

1. Objectives and navigation through the system
2. Interaction, feedback and support
3. Integration of online and face-to-face elements.

Objectives and navigation through CFE

If the learner wants to decide which part(s) of the course will best meet their own aims and objectives they will need to be guided at the outset by an overall map of the course in an analogous way to the contents list of a printed book. The learner will furthermore require an overview of each module and a stated objective at the micro-level of each unit. The overall aims and objectives of CFE are clearly stated on the Welcome page. From there an 'About this course' tab provides the user with access to a complete map of the course detailing both online and classroom elements of the course, shown in Figure 2.

Once the learner has started the course they will be able to judge the personal relevance of each module by the detailed introductory material. Furthermore, each unit clearly states its particular objective for each of the practised skills. For example, in Unit 5, 'Teamwork and managing people', the objective for the writing skills unit reads:

'In this section, you focus on improving your writing skills. You'll concentrate on one type of text per Module, getting practice writing the same documents that finance and accounting professionals must produce every day: financial reports, letters and emails to clients, etc. To do that, you'll also focus on very specific ways to improve your writing. Certain sections deal with spelling, punctuation, and vocabulary. Other sections look at sentence construction, so you can develop a more varied, fluid writing style. And other sections deal with the composition of texts as a whole, so you'll study the process of writing, from pre-writing to outlining, proof-reading and revision.' (Module 4 Unit 5, CFE)

The screenshot shows a PDF document titled 'Map of the course' from the CFE syllabus. It contains two tables. The first table is for 'Module 1: Jobs in finance' and is split into 'Online' and 'Classroom' sessions. The 'Online' session includes a case study about Baker Duncan and various activities for listening, speaking, reading, writing, and language. The 'Classroom' session includes a case study about selecting a new accountant and several communication tasks. The second table is for 'Module 2: Teamwork and managing people' and includes a case study about a company and several communication tasks.

Module	Session	Case study	Listening	Speaking	Reading	Writing	Language
1 Jobs in finance	Online	Baker Duncan, an accounting firm, needs to recruit a new accountant	The job market (video 1) Job interview (video 2) Professional skills (audio)	Using stress Question words Do/Did you ...? Question tags Short answers	Students' career guide: Finance and accounting Corporate structures Job adverts Careers in finance	Proofreading an application letter Punctuation Covering letter: Structure Task: A covering letter	Verbs in finance Questions Proofreading Punctuation marks
	Classroom	Selecting Baker Duncan's new accountant	1 Discussion: What makes a good CV/resume 2 Pair work: Forming and asking job interview questions 3 Discussion: Qualities and skills of an accountant 4 Role play: Job interview questions 5 Pair work: Case study 6 Presentation: Your company 7 Discussion: Job prospects for accountants 8 Test				Question forms Interview questions Describing a company Professional qualities

Figure 2:
CFE 'About this course' extract



Figure 3:
Hierarchical arrangement
of content in the CFE
course

Apart from allowing the learner to align their personal learning goals with the course objectives and empowering them to make autonomous choices, they will also need to have a clear understanding of how to navigate the learning landscape. Navigation concerns the way in which the user finds their way through the course, it provides them with an 'Ariadne thread' to prevent them from getting lost in hyperspace. At any stage in the course the learner must be able to navigate away from the current page and jump to a different activity, unit or module. The CFE model is a hierarchical arrangement: modules – units – activities, which is visible (and can be hidden) on the left hand panel of the screen. The screenshot in Figure 3 shows that the learner is currently studying activity 2.3 of the reading unit in Module 4 on cash flow. The learner can navigate to another activity, unit or module of the course as he wishes, i.e. the SCORM player enables random access to any part of the course.

Interaction, feedback and support

At the heart of didactic considerations on e-learning is the premise that all forms of learning, quite irrespective of the media concerned, must be established within a 'conversational framework', that is to say they require a discursive structure which allows the learner to participate actively in the learning process (Laurillard 1993:98). This is

not an easy task for any online course. As a 'warm-up' task and introduction to a new module CFE encourages the learner to reflect on their prior knowledge of the topic presented and then compare their notes with the information given in a short video clip or a reading text. For example:

'You are going to read an article explaining several ways that large accounting firms try to manage their people. Before you read, think about why it might be difficult for big companies to retain good workers. Do you have any ideas about how companies can try to keep their best staff, besides paying higher salaries? Read the article and see if any of your ideas are mentioned.' (Module 5, CFE)

The learner is then presented with different structured ways to engage with texts in a multi-media format. The exercises span a wide variety of activities ranging from receptive listening comprehension to free-text entry and productive speaking skills engaging the learner to respond to questions on content and language alike.

Correct or model answers are provided for all activities, and in some instances context-sensitive tips are given in response to an incorrect answer in order to encourage the learner to click the 'improve result' button, see Figure 4. These features can be seen as an important way to implement the notion of support or scaffolding.

In addition, a range of support features are provided within the learning environment.

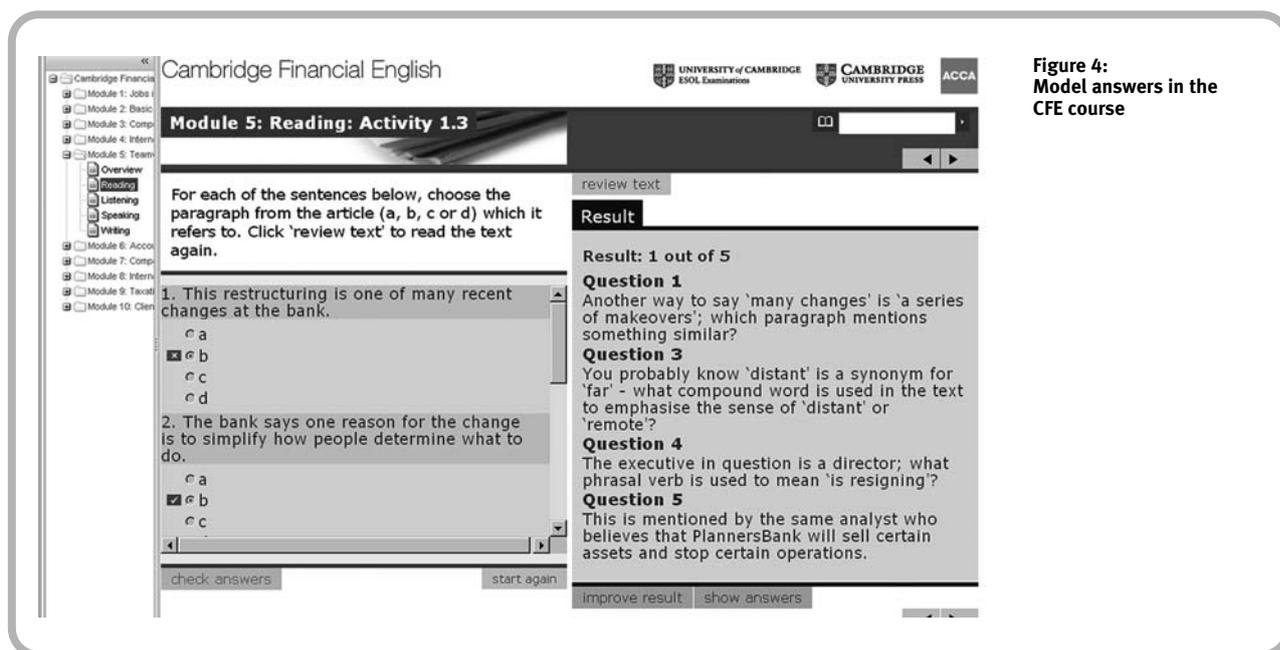


Figure 4:
Model answers in the
CFE course

There are two types of support features, firstly, those which are course-intrinsic and have to be built in by the course designer:

- a show transcript button allowing the learner to read the text of a sound file or a video clip
- a review button which allows the learner to look at the questions and a video clip, a sound or text file at the same time
- an online dictionary
- a print text button to avoid having to read longer text on screen.

Secondly, there are external features which are part of the Learning Management System, such as:

- a hand-in tool for sending written assignments to a tutor
- a portfolio tool for students' marked work and grades
- communication tools such as email, messaging, forums or discussion groups.

Integration of online and face-to-face elements

As Figure 2 demonstrates, the design of the CFE course presupposes a close link between the online and the face-to-face elements of the course. The latter expands on the theme of the online module, giving further opportunity to develop relevant lexis and structures in authentic, communicative speaking tasks. In each module the learner's attention is drawn to how the online elements of the course link up with the face-to-face classes and what they need to do to prepare for these. For example:

'In the face-to-face session with your teacher and the rest of the class, you are going to work in a small group with other students to discuss how to form the teams at the new regional centre for Planners Bank. To do that, you will have to understand how different people work together and how management can help. The strategy shown below will help you prepare for this.' (Module 5, CFE)

CFE includes a complete set of printable and photocopiable Teacher Notes and classroom activities per module. These materials allow for up to three hours of face-to-face tuition per module, offering the teacher a set of task-based, structured, communicative activities, including pair work, role plays, games and discussion ideas enabling students to put what they learn into practice. The Teacher Notes and classroom activities have been designed to offer maximum flexibility with lesson planning and typically contain:

- an introduction specifying how to blend the classroom session and the online work, stating approximate timings for all tasks
- two or three 10–20 minute discussion/personalisation tasks to introduce or consolidate the module topic

- a functional role play (30–45 minutes)
- a case study speaking task (45–60 minutes)
- one or two short extension discussion tasks (15–30 minutes)
- a multiple-choice review test.

This design is in stark contrast to a 'blended learning' concept where the learner can book the occasional on-demand session with a virtual tutor. It has also been shown that the course attrition rate for e-learning courses is sharply reduced where online and face-to-face tuition are fully integrated as in this course.

Conclusion

A fundamental problem and the greatest challenge for any e-learning programme is to provide context-sensitive and adaptive feedback to the learner that links their work and responses to their current proficiency and target level. This is what the formative assessment model strives to achieve as it would enable the learner to progress on an individual learning path based on their subject knowledge and language proficiency.

This is both a challenge for the content developer and the system architect working with sequencing and navigation in SCORM 2004. As in adaptive computer-based testing the content developer would need to develop a testbed for a formative model with a large item bank of calibrated language material. The system architect would need to develop more sophisticated options in sequencing and navigation storing and analysing each of the learner's responses in order to advise them on the optimal navigation path through their course.

The Cambridge Financial English course is one of the first online courses provided by Cambridge ESOL and the early interest in its use shows that there is a great need and potential for courses of this kind. Visit the CFE website for further information: www.financialenglish.org/cfe

References and further reading

- Advanced Distributed Learning website, www.adlnet.gov/Technologies/scorm
- Black, P, Harrison, C, Lee, C, Marshall, B and William, D (2003) *Assessment for Learning: Putting it into practice*, Maidenhead, UK: Open University Press.
- Laurillard, D (1993) *Rethinking university teaching: a Conversational Framework for the Effective Use of Learning Technologies*, London: Routledge.
- QCA (2007) *The 10 principles: Assessment for Learning*, retrieved from www.qca.org.uk/qca_4336.aspx
- The Assessment Reform Group (2002) *Assessment for Learning: 10 Principles: research-based principles to guide assessment practice*, retrieved from www.aaia.org.uk/pdf/AFL_10principlesARG.pdf

Impact of a Blended Learning course: observation and stakeholder views

ROGER HAWKEY CAMBRIDGE ESOL CONSULTANT

Introduction

This article summarises key findings from an impact study of an online blended English course (B1OLC) designed for students at Italian universities required to achieve a CEFR B1 (Threshold) level qualification in English. The focus of the study is the course taken by students at the University of Ferrara in Emilia Romagna, northern Italy, in 2008.

In this article the term *washback* refers to a course's influences on teaching, teachers, learning, curriculum and materials (see Alderson and Wall 1993, Hawkey 2006). The superordinate term *impact* is interpreted as 'the total effect of a test [or course] on the educational process and on the wider community' (McNamara 2000:133). Washback is thus part of impact (see Green and Hawkey 2004, Hamp-Lyons 1998, McNamara 1996, 2000, Shohamy 2001). Note also that we do not evaluate the impact of a course only directly, for example from before and after language proficiency measures or from participant responses to questions on how the course affected them (see sections below). We also seek insights into impact less directly, by probing what various stakeholders *think about* the course in the contexts, narrow and broader, in which it is taking place. Stakeholders' opinions of a course are often an indirect reflection of the ways in which it is impacting on them.

Background to the B1 online blended learning course

Cambridge ESOL has been working with Cambridge University Press and the Italian Universities Vice-Chancellor's Conference (*Conferenza dei Rettori delle Università Italiane*, CRUI) with particular reference to B1 level English teaching assessment since the 2001 University reform bill required that students could only graduate with a "minimum knowledge" of both computing and one other European language. Most universities interpreted the minimal foreign language level for graduation as CEFR B1.

As a *blended* learning course B1OLC is designed for students to undertake a combination of 75 hours of online self-study and 25 hours of face-to-face (ftf) teaching. Publicity for the course points out that it requires no software installations for the student or the university. Students can learn at their own pace by logging in from any multimedia PC, whether it be at home or at their university. Teachers are supported with classroom material and lesson plans for the ftf elements of all 24 course modules. Summary reports of student progress are available online so that teachers can monitor student progress.

The Ferrara University Impact Study

The course which was the focus of the impact study took place from February to May 2008 at the Ferrara University Language Centre. Like the first B1OLC course at the University of Bergamo (January to May 2007), the Ferrara University course was indeed run for students needing to acquire their B1 qualification in order to graduate in their various fields of study.

The data collection instruments (available with the formal report, Hawkey and Beresford-Knox 2009) were adapted for the Ferrara impact study from validated instrumentation used on the IELTS and Progetto Lingue 2000 impact studies (see Hawkey 2006, Saville 2009) and piloted in the University of Bergamo study. They included:

- student impact study (IS) questionnaires
- course teacher questionnaires
- interviews with the course teachers, academic co-ordinator, course administrator and technician
- course face-to-face lesson observations and analyses
- course online tracking data
- exam performance information.

Although the Ferrara University impact study was relatively small-scale, based on two visits to the campus and regular other contacts with key participants, the data collected were extensive. The analysis of impact and washback is summarised here according to stakeholder groups, that is the students, the teachers, and the management and administrative teams.

Course structure, participation and institutional support

Ferrara University had devoted significant resources to raising English language standards. For the 2008 March to May session of B1OLC, the University financed student participation although next time around the students themselves might be asked to contribute. A B1OLC Co-ordinator, appointed by the University to manage the B1OLC, had, in fact, helped with the piloting of the course and attended the Cambridge ESOL/Cambridge University Press Italy marketing meeting. One hundred and seventy students had taken a test designed by the Co-ordinator, which he described as a single version, cross-department instrument aimed in the main at A2 and B1 levels. Ninety B1OLC places in four blended learning course groups were available. Two English teachers, an administrator/liaison officer and a technician completed the B1OLC team. Each teacher taught two B1OLC course groups, at a rate of two course modules and two 1.5 hour face-to-face sessions a

week, which was agreed to represent the right amount and intensity of teaching. The course administrator, who had attended a Cambridge University Press B1OLC briefing before the course, was available for consultation throughout the course (which she herself also attended as a student).

A technician assisted and advised the students on any technical problems, including issues with their laptop connections.

About 75 of the original 92 students who started the course were regular participants on the February to May 2008 course.

The B1OLC learners

Information and opinions were collected from the students taking the B1OLC at Ferrara University through the Impact Study questionnaire which was completed online in its Italian-language version, by 35 of the students, around half of those still attending the ftf sessions in May 2008. The impact study principle observed in the data summarised in Table 1 is that we need to know about the background of the students, in particular their language learning and use, in order to interpret subsequent impact data.

Table 1: Key Ferrara University B1OLC student base data

Gender		Students' use of English outside the B1OLC	
Female 21	Male 14	Hours per week (hpw) reading for their field of study?	Hpw access to the media in English?
Age		0–2 hours	25
20 or less: 3	21–25: 24	3–5 hours	9
26–30: 6	30+: 2	6+ hours	1
Fields of Study		24	5
Architecture	2	6	2
Communication	2	Student self-assessment of level of computer expertise	
Economics	7	Beginner	2
Engineering	6	Intermediate	14
Languages	1	Advanced	17
Law	2	Very advanced	2
Lit & Philosophy	1		
Science	12		
Other	2		

The course co-ordinator and the administrator confirmed that this sub-group of students was typical of the whole B1OLC cohort in terms of age, academic and outside use of English and computer literacy. The slight preponderance of science and technology students was also, it appeared, fairly characteristic of English for Specific Purposes (ESP) groups at Italian universities. The higher proportion of female students among our respondents is probably slightly biased although Giacometti (2007) indicates that 55% of the first-degree Italian university enrolments were female.

Motivation and attitudinal factors

Table 2 summarises motivation and attitudinal data from the student questionnaire section derived from the Language Learning Questionnaire (LLQ; Purpura 1999) as adapted and validated through the Cambridge ESOL IELTS, Progetto Lingue 2000 and Florence Language Learning Gain projects

(Hawkey 2003). The responses here show a typical mixture of motivations and attitudes (see Dörnyei 2003, Hawkey 1982, Masgoret and Gardner 2003). For our Ferrara student group, the two strongest motivations for learning English through B1OLC appear to be its potential usefulness for getting a good job and 'because English is useful when travelling abroad'. The reasons 'to satisfy University regulations' and 'to help me pass an English language exam at B1 level' were not particularly strongly specified by this group. There could be important implications for university English as an Additional Language (EAL) course design here. If typical learners are expecting English for travel and contact as well as for job purposes, the impact of their EAL courses might be stronger if they attempted to cater for both purposes (see further in Table 2).

Table 2: Students' reasons for learning English

(N=35)		I strongly disagree			I strongly agree	
		1	2	3	4	5
1	Because it will be useful for getting a good job.	1	1	3	6	24
2	To satisfy University requirements.	4	11	12	6	2
3	To communicate with other people in my field of study.	3	4	7	12	9
4	Because studying languages is fun.	4	5	14	7	5
5	To help me in my plan to study abroad.	3	6	8	8	10
6	Because I have English-speaking friends.	8	9	6	5	7
7	To help me pass an English exam at B1 level.	3	2	8	10	12
8	Because English is useful when travelling abroad.	1	1	2	1	30

The broader motivation questions in section one of the student questionnaire are reinforced by responses to the 32 items (again from the LLQ) probing students' feelings and perceptions as they learn English using the blended learning course. A 60% or more selection of the top or lower two Likert scale categories, i.e. (*strongly*) agree or (*strongly*) disagree, is taken here as evidence of clear preferences in our characterisation of the student group. The selected items are ordered in Table 3 according to the strength of feeling expressed.

Of particular interest are responses indicating the particular learning and performance strategies favoured by the students (numbers 1–8). Here we have programme washback evidence which could helpfully inform further development of the B1OLC and its implementation in stakeholder universities.

Perceived course qualities

Table 4 summarises the opinions of the 35 student post-course questionnaire respondents on B1OLC coverage of the four language macro-skills; its perceived level across the language elements; its topics and themes, its navigation system and the match between the blend of online and face-to-face components.

Table 3: How I feel about, and learn, English

(N=35)	I strongly disagree					I strongly agree				
	1	2	3	4	5					
Feeling towards or approach to learning English										
<i>r/o The top eight most strongly agreed English language learning strategies:</i>										
1	I try to improve my speaking in English by including words and expressions I have heard.				28					
2	I try to improve my speaking in English by pronouncing new sounds until I can say them like a native speaker.				27					
3	I try to improve my writing in English by including words and expressions I have read.				26					
4	I try to improve my reading in English by trying to understand without looking up every new word.				23					
5=	I enjoy taking tests because it gives me a chance to show how much I know.				22					
5=	I enjoy learning English.				22					
5=	I make notes of the mistakes I make in English so that I can learn from them.				22					
8	When I am learning new material in English, I make sure to summarise information that I hear.				22					
<i>The top two most strongly disagreed English language learning strategies:</i>										
9	I try to improve my writing in English by analysing how writers organise their paragraphs.				21					
10	When I begin learning a new language, I decide which proficiency level I really intend to aim at.				23					

The following main inferences may be drawn from the responses in Table 4:

- Some adjustment may be considered to the balance between the reading and listening and the writing and speaking skills in B1OLC.
- On the proficiency level of B1OLC, responses are generally positive.
- Also encouraging, and supported by other impact study data, is the very positive student response on the interest and usefulness of the B1OLC topics.
- The student view of the B1OLC navigation system is 67% positive.

Student perceptions of course washback

Most, twenty of our 35 questionnaire respondents (57%), claimed to have achieved quite a lot of improvement in their English, a further 20% a lot of improvement. This encouraging view is further detailed in Table 5, with the students' view on B1OLC washback across their language skills. More students than for any other skill (46%) perceived their listening as their most improved skill, followed by reading (29%), which, however, was seen as only the third most improved by 34% of the student questionnaire respondents. Clearer-cut perhaps is the 40% of the group who felt that writing was their least improved

Table 4: How the Student Questionnaire Group perceive B1OLC

Adequacy of B1OLC skills activities				
	Reading	Listening	Writing	Speaking
More than enough	23	18	12	8
Enough	10	15	17	16
Not enough	2	2	4	8
Not nearly enough	0	0	2	3
Appropriacy of level of language elements				
	Grammar	Vocab	Task types	Instructional language
About the right level	21	25	18	26
Too difficult	7	1	4	2
Too easy	1	3	3	1
Depends on the Module	6	6	10	6
Online navigation systems				
Clear what to do				
	Always	Usually	Sometimes	Never
Moving between the units in a module	24	10	0	1
Moving between the three parts of a unit	24	10	0	1
Moving between activities	22	11	1	1
Checking answers during activities	24	5	5	1
Topics and themes				
Interesting and useful.				28
Interesting but not very useful.				4
Useful but not very interesting.				3
Neither interesting nor useful.				0
Fit between online and face-to-face classroom sessions				
Online course activities well linked to the classroom sessions.				21
Classroom sessions did not seem to link sufficiently with online work.				4
Not enough time for classroom sessions.				8
Classroom sessions not always available.				2

Table 5: Washback on language skills and elements

Perceived improvement across the four skills				
	Reading	Listening	Writing	Speaking
Most improved	10	16	2	7
Second most	7	11	8	9
Third most	12	2	11	10
Least improved	6	6	14	9
Perceived improvements in language elements				
	Grammar	Vocabulary	Pronunciation	Fluency
A lot	7	6	5	7
Quite a lot	2	26	19	18
A little	16	2	10	9
Very little	10	1	1	1

skill. Most of the students (74%) see, in addition, quite a lot of improvement in their vocabulary, 54% and 51% respectively in their pronunciation and fluency. Most of the group (74%) see only a little or very little improvement in

their grammar, a point that we shall return to below, in our analysis of face-to-face classroom lesson observations.

Blended learning courses, with their emphasis on student self-marking and correction online and monitoring by the teacher through the LMS are often seen as in tune with tendencies towards formative assessment (e.g. Rea-Dickins et al. 2008:61). From the student questionnaire responses we may make certain inferences on the B1OLC approach to assessment, noting that, broadly speaking, the students (54%) approved of the amount of testing in B1OLC. It is interesting, however, especially from a course washback perspective, that a fairly substantial 31% of the group thought that the tests on the B1OLC should resemble “the external exams”, for example a B1 level exam such as Cambridge ESOL’s Preliminary English Test (PET).

It is clearly of interest to users and potential users of the B1OLC what kind of washback the experience of a blended learning course *model* has on the Ferrara University student users (and, of course, on other stakeholders such as the teachers, administrators and so on, see below). The message from the questionnaires, albeit from interested parties, is that the 35 responding students mostly (69%) prefer the blended learning course model, especially over an online-only or a self-study book-oriented course. Responses to an item seeking insights into the washback of the online learning mode indicate that a higher proportion of students are both more enthusiastic and familiar with this mode (71% in all) than unaffected (29%) or less enthusiastic (0%).

The final question on the Student Questionnaire looks at the possible longer-term impact from B1OLC, which, after all, is intended as a course to help university students to achieve their qualifying B1 proficiency level in English. A good proportion of the students foresee a lot or quite a lot of impact from the course on their main courses at University (62%), on their English outside the University (80%), on their future work (83%) and on their attitudes to the English language (a very high 34 of the 35 participants (97%).

Target language performance

For the Ferrara University B1OLC students the most significant assessment was of their performance on the end-of-course English test designed by their Course Co-ordinator and approved as a qualification for their *required* B1 proficiency level certification in English. Information on this provided by the Course Co-ordinator was that:

- Fifty three of the original 90 B1OLC students took the test in May 2008, others would take the test, or another B1 exam, later. Some said they intended to take an international B1 exam such as PET rather than or in addition to the internal B1 level test.
- All those who took the final test scored more than they had on the parallel Ferrara University B1OLC selection test (described above). The Course Co-ordinator found this encouraging though he had not ascertained empirically how the levels of the selection test and the final assessment compared.
- The Course Co-ordinator’s summary claimed ‘a range of “improvement” of between 3 and 29 per cent (based on

scores), with a little cluster at 10 per cent and another, bigger cluster from 15 to 20 per cent’.

Although the data we received on test performance were insufficient to measure precise improvement for the group here, the picture of target level proficiency was generally positive.

The B1OLC teachers

In their questionnaire responses, the two teachers considered their students’ English language proficiency had improved, Teacher 1 adding that she found students ‘at the end noticeably more confident at speaking’, their ‘overall written work of a higher standard’. Both teachers saw reading and then writing as the two most improved skills. Could it be that the washback of B1OLC is stronger for the skills which one would have thought are more the focus of the online activities than those, namely listening and speaking, which feature strongly in the ftf sessions? Yet note that Teacher 1 also sees some of her students improving and becoming more confident in their speaking because the course can be intensive, every day, for them, thanks to the online speaking practice facility.

Both teachers see a lot of improvement in the students’ grammar and vocabulary, and “quite a lot” of improvement in pronunciation. Is there an interesting difference of perception here between the teachers and their students (see Table 5)? The students (26 out of 35 or 74%) saw only a little or very little improvement in their grammar. This apparent contradiction is discussed further below.

Like their students, the teachers feel that B1OLC hits the right language level in its coverage of grammar and vocabulary. This suggests a major positive impact of the course, achieved perhaps because of the significant research, trialling and orientation for B1OLC referred to in their interviews by the Ferrara Course Co-ordinator and his administrator (see also Tenberg 2009, this issue).

Equally heartening, for the Cambridge ESOL, Cambridge University Press and CRUI stakeholders in B1OLC was the teachers’ approval of the course’s topics and themes, which 80% of the students viewed as interesting and useful. Teacher 2 praised the fact that ‘even though the students came from a variety of university faculties with a wide variety of interests ... [most of them found] the topics and themes interesting and some of them particularly useful, e.g. (for students on the Erasmus programme) preparing a covering letter’.

The B1OLC navigation system permits the teacher to monitor student use of the self-study units through its four main functions: moving between the units in a module, moving between the three parts of a unit, moving between activities and checking answers during activities. Table 6 records the teachers’ complete closed and open-ended responses regarding navigation through the LMS.

Through the reports of the B1 LMS, teachers can call up and monitor at-a-glance student rates of online materials covered to date, and time spent online. Table 7 summarises, from the LMS, students’ activity levels over the February to May 2008 duration of the Ferrara B1OLC.

Table 6: Teacher views on B1OLC Learning Management System

	Monitoring student use of self-study units			
	Very	Quite	Not very	Not at all
Easy	T2	T1		
Useful	T2	T1		
Quick	T2		T1	

T1: LMS an excellent idea but correcting student written work takes c. 6 hpw, the ftf lessons 8 hpw. Limited time left for monitoring the students “in detail”. Students checked “every few weeks for rate of progress over the modules, but not much time to “look at individual results”, which involves clicking on each student then checking on each module: time consuming with 40 students to check.

T2: Approximately 1 hpw monitoring student performance.

From both teachers:

LMS clear and easy to use. Formatting tools would be useful for correcting written work to return to students or for messages to students, e.g. *coloured fonts, bold, highlight, underline, bullet points*; also would welcome a facility to put marked written exercises into separate folders (c.f. email inbox) for each student.

Table 7: B1OLC course Learning Management System data

Class (no. students)	% material covered to date		Hrs. online work	
	Ave.	Max.	Ave.	Max.
A (21)	46%	83%	31	85
B (24)	37%	95%	33	89
C (19)	64%	93%	44	102
D (18)	57%	84%	38	79

Table 7 reveals the variation across groups and individuals that is to be expected given the freedom, inherent in a blended learning programme, for participants to choose how much time to spend online.

The teachers felt that the online course activities are well-linked to the classroom sessions, a crucial feature for teachers on blended learning courses. Both had their main focus on speaking, followed, in order of priority, by listening, reading and, lastly, writing. Teacher 1 felt that the students came ‘to the [ftf] lessons to practise “real” conversation’ to supplement the online listening and recording of their voices.

Teacher 2 makes a further crucial point on a matter inherent, one would have thought, in the blended learning course format. After a few weeks some students had fallen behind schedule with the online modules and so were not familiar with the material covered in the related face-to-face lessons. But, as Teacher 2 noted, the blended format allowed such students to catch up through their online work, which is, of course, subject to their own pacing.

The two teachers rated the B1OLC Teacher’s Notes and materials in support of the ftf class sessions “not bad” to “good”. Teacher 2 had been impressed by how B1OLC had ‘kept the majority of students motivated to continue following’, and how “interesting and stimulating” the material had been from a teacher’s point of view. Both teachers were in broad agreement that B1OLC includes the right amount of testing and self-assessment.

Main findings from the directly washback (and impact) related items on the Teacher Questionnaire were that both

teachers felt that their experience with B1OLC would make them more enthusiastic about online language courses. They also agreed that B1OLC had had quite a lot of influence on their own teaching materials, testing and, significantly, on their beliefs as teachers. B1OLC had “some” influence on their lesson planning, teaching methods and skills. Following the B1OLC experience, Teacher 1 felt that the blended format was better for groups from a range of different fields of study (as in the case of B1OLC Ferrara). Teacher 2’s preferred choice would be a blended learning course or face-to-face small group, followed by face-to-face classroom or individual student courses. Both teachers would be happy to teach a blended learning course again in the future. Teacher 1 expected the blended model to permit students to study English at the appropriate level, who would ‘not otherwise have had time’ whilst Teacher 2 felt that University administrators would be keen to see this kind of course used again.

Lesson observation

B1OLC Impact Study team members were fortunate enough to be permitted to observe, audio-record and analyse two ftf lessons, each with a different student group and with a different teacher. The lessons were analysed by two observers independently, both using the impact study classroom analysis form used in the Cambridge ESOL IELTS, Progetto Lingue 2000 and Florence Language Learning Gain impact studies (see Hawkey 2006, Saville 2009).

The main observed B1OLC lesson features agreed by the observers and double checked by additional team members were as follows:

- Both lessons were mainly listening: speaking focused though with some student note-taking and a little writing.
- The teachers adopted a mainly communicative language teaching approach, encouraging the tapping of learner-authentic information and views, but responded to the students’ need and desire for accuracy-based grammar and lexis remedial work.
- The teachers combined handouts from the B1OLC *Teacher’s Notes and materials* with their own materials.
- The ratio of teacher: student-led activity in the two lessons was around 65:35% and 45:55% respectively. Most of the students entered into their group and pair work assignment with apparent interest and some enthusiasm.
- The ambiance was very pleasant and relaxed in both classes.

Other stakeholders

Course Co-ordinator

The Course Co-ordinator expressed the following main views on B1OLC at Ferrara University:

- Course construct and content represented quality, “second-generation” online material, with relevant, challenging content, the online audio and video materials good, though requiring plenty of teacher support.
- B1OLC represented a “hardline” (B1) PET level.
- Online learner: teacher contact can be *more* rather than

less personal because the online correcting and marking can be more immediate, the criticism more private than in class.

- The fff element was important not only for language learning reasons but also in terms of student commitment; the more fff, the more student 'loyalty'.
- The English teachers felt that their B1OLC experience could prove an extra string to their professional bow and facilitate more flexible teaching hours.
- A good, innovative English language programme such as the B1OLC, with positive measures of good student progress would strengthen the University's competitiveness.
- With participating students covering most of the course units in their own time and space the experience of the Ferrara B1OLC confirmed that blended learning courses should require fewer face-to-face lessons than wholly classroom-based programmes.

Course administrator and technician

The course administrator had been busy initially responding to B1OLC student emails as they familiarised themselves with the course format. She remained available to assist B1OLC users, but her help was called for much less often once the learners had familiarised themselves with the course.

The technician assigned to B1OLC support had experience of other online courses; he considered B1OLC of good quality overall, technically and in terms of design, although he thought that some students might have benefitted from clearer technical information early on.

Conclusion

The data from the impact study of the B1OLC course across a wide range of relevant variables suggest that this group of learners, with their varied motivations and strategies for learning English, have mainly positive views on the level, content, approaches, assessments and technical quality of the B1 online blended learning course. They also seem to have benefitted from B1OLC in terms of their language proficiency. The positive washback *expectations* of the students also appear in general to have been fulfilled.

Impact-related feedback from other key stakeholders, the teachers and management and administrative staff, suggest that the blended learning model is indeed appropriate for use with English for (mainly) academic purposes students

at Italian universities who have to meet a B1 level qualification in the language. The B1OLC construct may well be adaptable to other EAL contexts as suggested by Tenberg's article in this issue of *Research Notes*.

References and further reading

- Alderson, C J and Wall, D (1993) Does washback exist? *Applied Linguistics* 14, 115–129.
- Dörnyei, Z (2005) *The psychology of the language learner: Individual differences in second language acquisition*, Mahwah, NJ: Lawrence Erlbaum.
- Giocometti, M (2007) *Gender Equality and Higher Education*, paper presented at the Conference of the European Association in Science, Engineering and Technology, Vilnius, 27–28 November, retrieved from: www.esib.org/index.php/Publications/conference-documents/45-board-meetings/316-esus-53rd-board-meeting
- Green, A and Hawkey, R (2004) Test washback and impact, *Modern English Teacher* 13, 66–70.
- Hamp-Lyons, L (1998) Ethical test preparation practice: the case of the TOEFL, *TESOL Quarterly* 32/2, 329–337.
- Hawkey, R (1982) *An investigation of cognitive/affective and social factors and language learning*, unpublished PhD thesis, Institute of Education, London University.
- (2003) *Florence Language Learning Growth Project: Developments since October 2002, and Details of the Dataset*, internal Cambridge ESOL report.
- (2006) *Impact Theory and Practice: Studies of the IELTS test and Progetto Lingue 2000*, Studies in Language Testing vol. 24, Cambridge: UCLES/Cambridge University Press.
- Masgoret, A and Gardner, R (2003) Attitudes, motivation, and second language learning: a meta-analysis of studies conducted by Gardner and associates, *Language Learning* 53/1, 123–163.
- McNamara, T (1996) *Measuring Second Language Performance*, Harlow: Longman.
- (2000) *Language Testing*, Oxford: Oxford University Press.
- Purpura, J (1999) *Learning Strategy Use and Performance on Language Tests: A structural equation modelling approach*, Studies in Language Testing vol. 8, Cambridge: UCLES/Cambridge University Press.
- Rea-Dickins, P, Poehner, M, Leung, C, Taylor, L and Shohamy, E (2008) *Breaking the Mould: evaluating validity from a situated perspective*, paper presented at 3rd International Conference of the Association of Language Testers in Europe (ALTE), Cambridge, UK, 10–12 April.
- Saville, N (2009) *Developing a Model for Investigating the Impact of Language Assessment within Educational Contexts by a Public Examination Provider*, unpublished PhD thesis, Centre for English Language Learning and Assessment, University of Bedfordshire.
- Shohamy, E (2001) *The power of tests: a critical perspective on the Use of Language tests*, Harlow: Pearson Education.

Profile of Skills for Life candidature in 2007–8

IVANA VIDAKOVIĆ RESEARCH AND VALIDATION GROUP, CAMBRIDGE ESOL

Introduction

Cambridge ESOL's Certificates in ESOL Skills for Life (SfL) are part of the Government's strategy for improving adult literacy, numeracy and ESOL skills in England and

are based on the Adult ESOL Core Curriculum.¹ They are available at five levels (Entry 1, Entry 2, Entry 3, Level 1

1. Details of the Adult ESOL Core Curriculum can be found at: www.excellencegateway.org.uk/sflcurriculum

and Level 2) on the National Qualifications Framework.²

SfL was designed as a modular exam to allow for the fact that many candidates have varying levels of ability in different skills. This means that candidates may choose to be assessed in different modes (Reading, Writing, Listening and Speaking) at different levels in the same or different exam sessions. They can also enter for single modes, as their skills develop, and thus build up a portfolio of achievement over a period of time (Skills for Life Handbook, Cambridge ESOL 2005).

Candidature

The target Skills for Life candidature consists of adult learners whose first language is not English and who are living or trying to settle in England. These ESOL learners may include refugees or asylum seekers, migrant workers, people from settled communities and partners or spouses of people who have been settled in the UK for a number of years (Cambridge ESOL 2005). Since the learners' educational and employment backgrounds are diverse, as well as their literacy levels and language learning skills, SfL exam papers are designed to reflect this diversity in the exam materials, topics and task types.

The present article will provide an insight into the characteristics of 2007–8 SfL candidature in terms of age, gender, first language and length of time spent studying English. It is, therefore, a follow-up to an earlier article which described 2005–6 Skills for Life candidates, covering a year after the exam launch (Novaković 2006). The general aim of these descriptions of SfL candidates is monitoring changes in the test taking population which can lead to test developments and revisions.

The results which will be presented in this article are based on 41,475 candidates who took a SfL Writing exam between October 2007 and October 2008. Candidates were asked to fill out a Candidate Information Sheet (CIS) at each SfL Writing examination and thus provided information on their age, gender, their first language and number of years they have studied English. This information is confidential and it does not affect candidates' exam scores.

Age and gender

Figure 1 represents the percentages of candidates belonging to different age groups. The largest age group among SfL candidates is 21–30 (43%). This group is closely followed, percentage-wise, by candidates between 31 and 50 years of age (38%). The remaining candidates are between 16 and 20 years of age (15%), and over 50 (4%). This pattern is present across all SfL levels, apart from Entry 1. The largest age group at Entry 1 level is the older one, 31–50 (42%), which is followed by a group of candidates who are between 21 and 30 years old (35%). It could be hypothesised that the deviation of the Entry 1 candidates from this general age pattern occurred because younger SfL candidates (21–30) have a better knowledge of English than other age groups and are, therefore, more likely to take exams at a level higher than Entry 1.

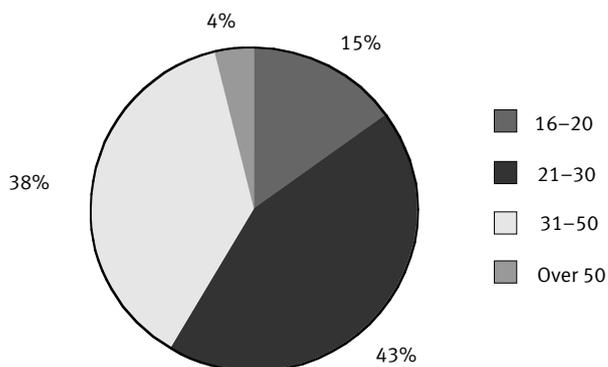


Figure 1: Age of SfL candidates in 2007–8

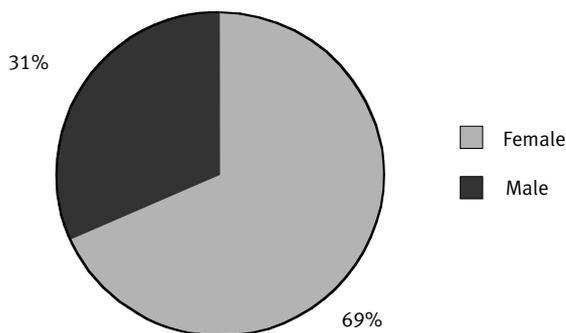


Figure 2: Gender of SfL candidates in 2007–8

If the current data are compared to those from 2005–6, it is evident that the SfL candidature has become older. In 2005–6, almost half of the candidates were between 16 and 20 years old, 34% were between 21 and 30, while those who belonged to 31–50 age group constituted only 14% of SfL candidature (Novaković 2006).

As far as gender is concerned, Figure 2 reveals that most candidates are female (69%), while 31% are male. This is true for all entry levels and is in line with 2005–6 findings (Novaković 2006).

First language

Fifty five languages were identified as first languages (L1s) of the 2007–8 SfL candidature. For practical purposes, only the ten most frequently chosen L1s are represented in Figure 3. As this figure shows, the majority of candidates

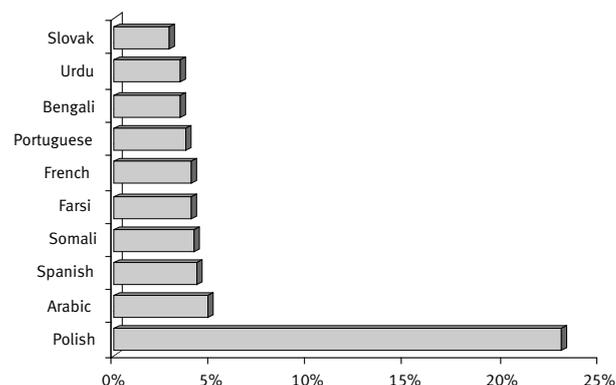


Figure 3: Ten most frequently chosen L1s by SfL candidates in 2007–8

2. Details of the National Qualifications Framework can be found at: www.qca.org.uk/qca_5967.aspx

are L1 speakers of Polish (23%), followed by Arabic, Spanish, Somali and Farsi speakers. The Polish candidates are most numerous across all five entry levels, which is in line with 2005–6 data, however the percentage of Polish candidates in the Sfl candidature as a whole has increased from 16% in 2005–6 to 23% in 2007–8.

Years of learning English as a foreign language

At the time of sitting the exam, most candidates had spent between 6 months and 2 years learning English (50%), as shown in Figure 4. Candidates who had been learning English between 2 and 5 years constituted 28% of the

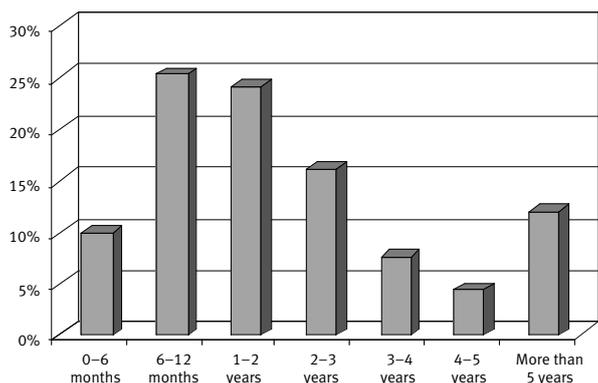


Figure 4: Time spent studying English by Sfl candidates

candidature. These are followed by candidates who had learnt English for more than 5 years (12%) and less than 6 months (10%). These tendencies closely reflect those observed in 2005–6.

Conclusion

The investigation of 2007–8 data has shown that Sfl candidates are mainly above twenty years of age. The largest age group consists of candidates who are between 21 and 30 years of age, and it is closely followed by 31–50 age group. The majority of candidates are female, and they had spent between six months and two years studying English before taking a Sfl exam. The comparison of 2007–8 data with those from 2005–6 has revealed plenty of similarities, but it has also shown that Sfl candidature has become older. Cambridge ESOL will continue monitoring the profile of Sfl candidates with the aim to flag changes in the test taking population which could result in future developments and revisions of the existing exam formats and topics.

References

- Cambridge ESOL (2005) *Skills for Life Handbook*, retrieved from www.cambridgeesol.org/assets/pdf/resources/sfl/sfl_handbook.pdf
- Novaković, N (2006) Profile of Skills for Life Candidature, *Research Notes* 25, 14–15, Cambridge: Cambridge ESOL.

IELTS candidate performance in 2008

More than 1 million candidates sit an IELTS test each year. We report here on the performance of candidates on both Academic and General Training Modules in 2008.

Band score information

IELTS is assessed on a 9-band scale and reports scores both overall and by individual skill. Overall band scores for Academic and General Training candidates in 2008 are shown here together with scores for individual skills according to a variety of classifications. These figures are broadly in line with statistics for previous years.

Academic and General Training candidates

Table 1 shows the split between the Academic and General Training candidature in 2008. Three quarters of the candidature sat an Academic IELTS test with the rest taking the General Training Module.

Gender

Table 2 shows the mean band scores by skill and overall by gender. For both male and female candidates, those taking the General Training option gained a marginally higher band score overall than for the Academic Module. Female

Table 1: Proportion of candidates taking each module

2008	
Academic	75.4%
General Training	24.6%

Table 2: Mean band scores by gender and skill

	Female				
	Listening	Reading	Writing	Speaking	Overall
Academic	6.01	5.92	5.53	5.87	5.89
General Training	6.20	5.80	5.90	6.28	6.11
	Male				
	Listening	Reading	Writing	Speaking	Overall
Academic	5.83	5.69	5.37	5.69	5.71
General Training	6.05	5.66	5.73	6.17	5.97

candidates earned higher average band scores than males for both Academic and General Training Modules. Looking at the skills in more detail, female candidates taking the Academic module gained the highest band score on

Listening, followed closely by Reading, Speaking and then Writing, whilst Male candidates performed similarly on the Academic Module, with Listening, followed by Reading and Speaking together, then Writing. On the General Training Module, both male and female candidates gained the highest band score on Speaking, followed by Listening, Writing, then Reading.

Reason for taking IELTS

Table 3 shows the reasons given by candidates for taking IELTS distributed by band score for the two Modules. For example, of the candidates who gave 'For application to Medical Council' as their main reason for taking IELTS Academic, 55% of these gained a band 7 or above on the Reading paper. For General Training candidates, whose reasons for taking IELTS differ, of the candidates who gave 'For employment' as their main reason for taking IELTS, 54% gained a band 5 or above on the Reading paper.

Place of origin

Tables 4 and 5 show the mean overall and individual band scores achieved by 2008 Academic and General Training candidates according to their place of origin. Note that for place of origin (and first language) the following tables show the top 40 places and languages, listed alphabetically, not in order of the size of the candidature.

First language

Tables 6 and 7 show the mean overall and individual band scores achieved by 2008 Academic and General Training candidates from the top 40 first language backgrounds. This candidate performance data, including that for previous years, can be found on the IELTS website: www.IELTS.org

Table 4: Academic mean band scores by most frequent countries or regions of origin

Country/Region of origin	Listening	Reading	Writing	Speaking	Overall
Bangladesh	5.60	5.25	5.33	5.66	5.52
Brazil	6.64	6.81	6.12	6.80	6.65
China (PRC)	5.50	5.70	5.12	5.25	5.46
Colombia	6.36	6.48	5.78	6.52	6.35
Cyprus	6.18	5.87	5.61	6.33	6.06
Egypt	6.23	6.09	5.87	6.34	6.20
France	6.73	7.00	6.12	6.43	6.63
Germany	7.36	7.19	6.59	7.24	7.16
Greece	7.10	6.82	6.07	6.45	6.67
Hong Kong	6.65	6.64	5.77	5.93	6.31
India	6.04	5.58	5.51	5.77	5.79
Indonesia	6.10	6.24	5.50	5.86	5.99
Iran	5.96	5.81	5.82	6.44	6.07
Italy	5.87	6.56	5.59	5.98	6.06
Japan	5.89	5.86	5.35	5.80	5.79
Jordan	5.59	5.46	5.13	5.82	5.57
Kazakhstan	5.88	5.85	5.52	5.94	5.87
Kenya	6.53	6.26	6.30	6.88	6.56
Korea, South	5.92	5.89	5.29	5.60	5.74
Kuwait	5.43	5.11	4.91	5.63	5.33
Malaysia	7.04	6.92	6.17	6.45	6.71
Mexico	6.67	6.91	6.03	6.69	6.64
Myanmar	6.09	6.04	5.61	5.82	5.95
Nepal	5.96	5.43	5.42	5.61	5.67
Nigeria	6.07	6.08	6.14	7.06	6.40
Oman	5.11	5.03	4.95	5.65	5.25
Pakistan	5.73	5.41	5.45	5.79	5.66
Philippines	6.94	6.51	6.25	6.81	6.69
Poland	6.90	6.69	6.11	6.86	6.70
Qatar	4.70	4.55	4.46	5.26	4.81
Russia	6.59	6.50	5.95	6.66	6.48
Saudi Arabia	5.17	4.97	4.83	5.81	5.26
South Africa	7.72	7.40	7.18	8.33	7.72
Spain	6.63	6.90	6.02	6.45	6.56
Sri Lanka	6.39	5.88	5.78	6.28	6.15
Taiwan	5.68	5.80	5.24	5.66	5.66
Thailand	5.84	5.80	5.29	5.67	5.71
Turkey	5.99	5.83	5.43	5.97	5.87
United Arab Emirates	4.87	4.80	4.73	5.33	5.00
Vietnam	5.68	5.99	5.56	5.72	5.80

Table 3: Distribution of overall band scores by reason for taking IELTS and module

	Below 4	4	4.5	5	5.5	6	6.5	7	7.5	8	8.5	9
Academic												
For application to Medical Council	0%	0%	1%	2%	6%	13%	21%	23%	17%	11%	4%	0%
For employment	1%	1%	3%	6%	13%	22%	23%	17%	9%	4%	1%	0%
For higher education extended course	2%	4%	10%	17%	21%	19%	13%	7%	4%	2%	0%	0%
For higher education short course	3%	6%	13%	19%	19%	16%	11%	7%	3%	2%	0%	0%
For immigration	1%	1%	3%	6%	12%	19%	19%	15%	11%	8%	4%	1%
For professional registration	1%	1%	4%	6%	11%	16%	21%	18%	11%	7%	3%	0%
For registration as a dentist	0%	1%	2%	3%	6%	16%	25%	21%	15%	9%	2%	0%
For registration as a nurse (including CGFNS*)	0%	0%	1%	4%	12%	23%	25%	18%	10%	4%	1%	0%
General Training												
For employment	8%	9%	14%	15%	15%	14%	11%	7%	4%	2%	1%	0%
For higher education extended course	5%	8%	16%	18%	17%	16%	11%	6%	2%	1%	0%	0%
For higher education short course	7%	10%	17%	18%	18%	15%	10%	4%	1%	1%	0%	0%
For immigration	2%	2%	5%	10%	15%	19%	18%	13%	8%	5%	2%	0%
For personal reasons	2%	3%	7%	14%	18%	20%	17%	11%	5%	2%	1%	0%
For professional registration	2%	3%	6%	9%	14%	23%	19%	13%	7%	4%	1%	0%
For training or work experience	6%	10%	19%	20%	15%	14%	9%	5%	2%	1%	0%	0%

*CGFNS International (formerly the Commission on Graduates of Foreign Nursing Schools), see www.cgfns.org

Table 5: General Training mean band scores by most frequent countries or regions of origin

Country/Region of origin	Listening	Reading	Writing	Speaking	Overall
Bangladesh	5.84	5.28	5.58	6.08	5.76
Brazil	6.27	6.25	6.05	6.60	6.37
China (PRC)	6.06	5.95	5.64	5.84	5.94
Colombia	5.89	5.88	5.70	6.20	5.98
Egypt	5.79	5.45	5.67	6.19	5.84
Fiji	6.17	5.68	6.02	6.68	6.20
France	6.88	6.67	6.42	6.89	6.78
Germany	6.77	6.39	6.36	7.04	6.71
Hong Kong	6.37	6.16	5.82	6.13	6.19
India	6.24	5.57	5.79	6.21	6.02
Indonesia	6.40	6.08	5.79	6.16	6.17
Iran	5.74	5.40	5.88	6.34	5.90
Iraq	5.67	5.29	5.50	6.42	5.77
Italy	5.61	5.76	5.52	5.99	5.79
Japan	5.84	5.59	5.37	5.85	5.73
Jordan	5.72	5.33	5.42	6.10	5.70
Kenya	7.01	6.46	6.82	7.41	6.98
Korea, South	5.43	5.27	5.08	5.28	5.33
Lebanon	5.91	5.44	5.72	6.38	5.95
Malaysia	7.14	6.69	6.51	6.92	6.88
Mauritius	6.52	6.05	6.37	6.74	6.48
Mexico	6.22	6.24	6.03	6.56	6.33
Nepal	6.33	5.65	5.90	6.21	6.08
Nigeria	5.75	5.79	6.35	7.07	6.30
Pakistan	5.94	5.43	5.85	6.31	5.95
Philippines	5.75	5.19	5.66	5.91	5.69
Poland	6.28	6.04	5.94	6.65	6.29
Russia	6.27	6.20	5.93	6.33	6.25
Singapore	7.45	6.86	6.84	7.40	7.20
South Africa	7.38	6.91	7.23	8.28	7.51
Sri Lanka	6.15	5.66	5.82	6.29	6.04
Taiwan	5.87	5.74	5.58	6.03	5.87
Thailand	5.41	5.19	5.15	5.53	5.39
Turkey	5.82	5.60	5.48	5.98	5.79
Ukraine	6.01	5.88	5.72	6.08	6.00
United Arab Emirates	4.52	3.91	4.40	5.11	4.55
United Kingdom	7.97	7.45	7.89	8.83	8.10
Venezuela	5.97	6.09	6.03	6.38	6.18
Vietnam	5.21	5.30	5.38	5.46	5.40
Zimbabwe	6.80	6.33	6.96	7.76	7.03

Table 6: Academic mean band scores for most common first languages

	Listening	Reading	Writing	Speaking	Overall
Afrikaans	7.27	7.05	6.78	7.77	7.28
Arabic	5.22	5.09	4.97	5.69	5.31
Bengali	5.67	5.31	5.38	5.71	5.58
Burmese	6.09	6.04	5.61	5.82	5.95
Chinese	5.64	5.81	5.20	5.35	5.57
Farsi	5.97	5.81	5.82	6.44	6.07
French	6.52	6.70	6.04	6.41	6.48
German	7.38	7.21	6.59	7.24	7.17
Greek	6.74	6.44	5.88	6.42	6.44
Gujarati	5.74	5.28	5.38	5.58	5.56
Hindi	6.36	5.85	5.69	6.04	6.05
Ibo/Igbo	5.89	5.99	6.14	7.02	6.32
Indonesian	6.10	6.23	5.49	5.85	5.98
Italian	5.86	6.55	5.58	5.98	6.05
Japanese	5.89	5.86	5.36	5.80	5.79
Kannada	6.96	6.42	6.25	6.78	6.67
Kazakh	5.80	5.78	5.46	5.87	5.79
Korean	5.92	5.89	5.29	5.60	5.74
Lithuanian	6.92	6.79	5.88	6.40	6.56
Malay	6.83	6.66	6.00	6.31	6.51
Malayalam	6.53	6.12	5.94	6.28	6.28
Marathi	6.85	6.29	6.17	6.53	6.52
Nepali	5.96	5.44	5.43	5.62	5.67
Pashto	5.32	5.18	5.36	5.76	5.47
Polish	6.90	6.69	6.11	6.85	6.70
Portuguese	6.61	6.73	6.07	6.71	6.59
Punjabi	5.52	5.14	5.09	5.23	5.31
Romanian	7.14	7.08	6.39	6.93	6.95
Russian	6.51	6.42	5.90	6.55	6.41
Singhalese	6.43	5.91	5.83	6.33	6.19
Spanish	6.49	6.69	5.95	6.58	6.49
Swahili	6.11	6.00	6.15	6.78	6.32
Tagalog	6.94	6.51	6.25	6.80	6.69
Tamil	6.61	6.12	5.87	6.33	6.29
Telugu	6.15	5.55	5.60	5.89	5.86
Thai	5.84	5.80	5.28	5.66	5.71
Turkish	6.00	5.83	5.46	5.99	5.89
Urdu	5.82	5.44	5.47	5.81	5.70
Vietnamese	5.68	5.99	5.56	5.72	5.80
Yoruba	6.16	6.27	6.21	7.03	6.48

Table 7: General Training mean band scores for most common first languages

	Listening	Reading	Writing	Speaking	Overall		Listening	Reading	Writing	Speaking	Overall
Afrikaans	7.13	6.74	6.96	7.97	7.26	Marathi	6.50	5.83	6.09	6.48	6.29
Arabic	5.18	4.73	5.04	5.73	5.24	Nepali	6.33	5.64	5.91	6.22	6.09
Bengali	5.90	5.35	5.64	6.14	5.82	Pashto	5.87	5.35	5.94	6.39	5.94
Chinese	6.12	5.99	5.68	5.90	5.99	Polish	6.28	6.04	5.94	6.65	6.28
Creole	6.34	5.89	6.22	6.60	6.33	Portuguese	6.24	6.21	6.01	6.57	6.33
Dutch	7.37	6.99	6.67	7.51	7.20	Punjabi	5.96	5.26	5.53	5.87	5.72
Farsi	5.74	5.40	5.88	6.34	5.90	Romanian	6.05	5.90	5.78	6.17	6.04
French	6.49	6.29	6.29	6.69	6.50	Russian	6.22	6.15	5.91	6.30	6.22
German	6.75	6.40	6.33	6.96	6.68	Shona	6.60	6.22	6.85	7.50	6.85
Gujarati	6.06	5.40	5.61	6.01	5.83	Sindhi	4.83	4.83	5.39	5.52	5.20
Hebrew	6.45	6.24	5.99	7.12	6.52	Singhalese	6.15	5.66	5.82	6.28	6.04
Hindi	6.52	5.84	6.01	6.52	6.29	Spanish	5.97	6.01	5.88	6.32	6.11
Hungarian	6.12	5.82	5.76	6.28	6.06	Tagalog	5.73	5.17	5.65	5.89	5.68
Indonesian	6.40	6.08	5.79	6.16	6.17	Tamil	5.99	5.48	5.63	6.04	5.85
Italian	5.58	5.74	5.49	5.97	5.77	Telugu	6.35	5.62	5.83	6.31	6.09
Japanese	5.83	5.59	5.36	5.85	5.73	Thai	5.40	5.18	5.15	5.52	5.38
Kannada	6.29	5.61	5.93	6.34	6.11	Turkish	5.85	5.62	5.49	6.00	5.81
Korean	5.43	5.27	5.08	5.28	5.33	Ukrainian	5.78	5.59	5.53	5.99	5.80
Malay	7.08	6.55	6.46	6.95	6.82	Urdu	6.12	5.54	5.90	6.41	6.06
Malayalam	6.26	5.77	5.99	6.27	6.14	Vietnamese	5.21	5.30	5.38	5.46	5.40

Conference reports

Staff from the Research and Validation Group have participated in various international events over the past few months, presenting plenaries, symposia and papers with internal and external colleagues on a wide variety of topics, reported on below by Fiona Barker and Ardeshir Geranpayeh.

30th GERAS Conference, Rennes

Fiona Barker gave a plenary paper at the 30th GERAS Conference (Groupe d'Étude et de Recherche en Anglais de Spécialité) from 12–14 March at the University of Rennes 1, France. The conference theme was *From non-specialised to specialised: routes towards specialised English* which included two plenary papers and around a hundred individual papers and posters on the unity and diversity of English for Specific Purposes (ESP) teaching and learning in a variety of contexts.

The first plenary speaker, Nicolas Ballier (Université Paris 13 Villetaneuse) spoke on *La langue des linguistes: essai de caractérisation d'un anglais de spécialité*. Fiona Barker's plenary was on the topic of *Exploring unity and diversity within non-native English using learner corpora* in which she explored the relatively recent development of learner corpora, together with their implications for studying ESP, drawing on her experience of developing and using corpora, particularly those that include domain-specific learner output, and applying corpus related techniques within language assessment.

12th International TESOL Arabia Conference and Exhibition, Dubai

Hanan Khalifa presented a paper at the 12th TESOL Arabia Conference which took place in Dubai from 12–14 March 2009 on the theme of *English in Learning, Learning in English*. Hanan's paper was entitled *Assessment of washback tool: beneficial or harmful?* in which she explored how washback can be viewed on a continuum moving from negative through neutral and into a positive dimension. Hanan also examined features of teaching and learning that are subject to exam influence and provided an account of how educators can foster beneficial washback before presenting washback as an integral part of a framework for developing and validating tests.

31st Language Testing Research Colloquium, Denver

The 31st Annual LTRC – the annual conference of the International Language Testing Association (ILTA) – was held on March 17–20 in Denver, Colorado, USA. Several members of Cambridge ESOL's Research and Validation Group participated in this event, which included the presentation of the 2008 IELTS Masters Award to Susan Clarke (see details of her winning dissertation in issue 35).

The theme of the 2009 LTRC was *Reflecting on 30 Years: Learning, Transparency, Responsibility and Collaboration*.

Nick Saville and Szilvia Papp presented a paper in the symposium *Investigating the impact of assessment for migration purposes*, jointly organised by Elana Shohamy (Tel Aviv University) and Nick Saville, with discussant Tim McNamara (University of Melbourne) and with contributions from colleagues from the University of Ghent, Tilburg University, Università per Stranieri, Perugia and Tel Aviv University. Evelina Galaczi presented a paper on *The role of quantitative and qualitative methodologies in the development of rating scales for speaking*, whilst Lynda Taylor gave a paper on *Telling our story: Reflections on the place of learning, transparency, responsibility and collaboration in the language testing narrative*. Lynda also presented jointly with Pauline Rea-Dickins (University of Bristol) Matt Poehner (Pennsylvania State University), Constant Leung (King's College London) and Elana Shohamy (Tel Aviv University) on *From the Periphery to the Centre in Applied Linguistics: the case for situated language assessment*.

American Association of Applied Linguists Conference, Denver

Cambridge ESOL staff also attended the annual AAAL conference in Denver on 21–24 March. Nick Saville (Cambridge ESO) and Mike McCarthy (Universities of Nottingham and Limerick) presented on *Profiling English in the real world: what learners and teachers can tell us about what they know*. This paper outlined the English Profile (EP) programme and showed how large-scale data-gathering of spoken and written data across countries and learning contexts is contributing to a precise understanding of what differentiates learners at different levels of the Common European Framework of Reference. To illustrate the approach they described the development of new spoken corpora using spoken learner data, looking at how some of the corpus linguistic work on learners' writing might be replicated with spoken data, as well as giving examples of how the spoken investigations will need to differ radically.

Evelina Galaczi and Mick Ashton presented a paper entitled: *Developing tests for teachers: Insights from the TKT (Teaching Knowledge Test) suite of examinations* in which they explored what knowledge for teaching L2 teachers need and how this can be tested in an objective test format. They explored these questions in the light of Cambridge ESOL's TKT suite of examinations, with a focus on the underlying constructs and validity evidence gathered through trialling and stakeholders' feedback.

Lynda Taylor's paper took the theme: *Test wiseness reconceptualised: developing 'assessment literacy' within the applied linguistics community*. In this paper Lynda considered the growing importance of 'assessment literacy' within the applied linguistics community and discussed the contribution of the language testing and assessment

community in ensuring that the knowledge, skills and experience needed to underpin good quality assessment are shared as widely as possible for the benefit of all.

BALEAP 2009, Reading

The British Association of Lecturers in English for Academic Purposes (BALEAP) conference was held at the University of Reading in April 2009. BALEAP is a professional organisation whose members are providers of English for academic purposes in higher education institutions. The 2009 event was attended by about 250 participants.

From Cambridge ESOL, Ivana Vidaković and Hugh Bateman presented a paper on *ILEC: testing Speaking in a legal context* in which they discussed the Speaking component of a test of English language skills in a legal context within a socio-cognitive framework of test development and validation. They focused on different ways of ensuring and monitoring the scoring validity of the Speaking paper since the launch of ILEC in May 2006.

National Council on Measurement in Education Annual Meeting, San Diego

Ardeshir Geranpayeh presented a paper in the co-ordinated session (symposium) on *Current practices in licensure and certification testing* alongside two presenters from Alpine Testing Solutions at the annual NCME meeting in April 2009 in San Diego, USA.

This symposium highlighted some examples of current test development and validation practices in licensure and certification testing (generally, credentialing). Although practices are often similar to those seen in K-12 educational assessment programmes, validity frameworks for credentialing programmes may prioritise different sources of evidence. The symposium illustrated how defining the intended use and interpretation of scores within credentialing testing programmes often necessitate more diverse studies that inform policy and practice.

Ardeshir's paper was on the *Use of language testing as a companion for a credentialing examination programme*.

The paper argued that the shortage of professional skills in various subject areas has attracted many internationally trained candidates to seek employment in the US and Canada. This has introduced a new challenge for professional licensing institutions to expand their traditional measurement of professional competency to include some element of English language competence. There has been an increasing demand for the professional licensing bodies to adopt well-established widely available international English language proficiency test batteries as part of assessing the English language competence of their international candidates. The paper therefore reported on the use of the International English Language Testing System (IELTS) as a companion for credentialing examination programmes in North America. It reported on how IELTS has been used to represent a common language literacy construct differently for different purposes through its two variants (General Training and Academic modules) and how its use as a companion for credentialing examination programmes has led to the revision of its General Training module.



From left to right, Chad Buckendahl (Alpine Testing Solutions), Susan Davis (Alpine Testing Solutions) and Ardeshir Geranpayeh (Cambridge ESOL)

Recent publications of interest

Studies in Language Testing

February 2009 saw the publication of another title in the *Studies in Language Testing* series, published jointly by Cambridge ESOL and Cambridge University Press. Volume 28, by Roger Hawkey, is entitled *Examining FCE and CAE: Key issues and recurring themes in developing the First Certificate in English and Certificate in Advanced English exams*.

This latest volume examines the historical development of two of Cambridge ESOL's most well-known examinations – the First Certificate in English (FCE) and the Certificate in Advanced English (CAE), and it offers readers a unique account of their evolution over more than 75 years. The story begins with the introduction of the pre-cursor of FCE

(the Lower Certificate in English) in 1939. The book goes on to trace subsequent developments, including the introduction of FCE in 1975 and of CAE in 1991, as well as the regular projects to modify and update both tests over time. Roger Hawkey describes some of the key issues and recurring themes down the decades: test constructs; proficiency levels; principles and practice in test development, validation and revision; organisation and management; and test stakeholders and partnerships.

Key features of the book include: a clear and detailed description of the different social, educational and linguistic factors that shaped language test specification, development, validation, production, and management from the 1930s to the present day; a reflective and critical account of test development activity within a specific

organisational context, informed by access to UCLES archives and interviews with key figures involved; a unique set of facsimile copies of FCE and CAE test versions, from the original tests in 1939 and 1991 through various revision projects to the updated formats of 2008.

With its wide-ranging and accessible coverage of matters relating to examination development, this volume will be of interest to language testing researchers, academic lecturers,

postgraduate students, and educational policy-makers. It will also be a useful reference for teachers, directors of studies, school-owners and other stakeholders involved in preparing students for the Cambridge exams, particularly FCE and CAE. This title complements previous historical volumes in the series on CPE, BEC, CELS and IELTS.

More information is available at:
www.cambridgeesol.org/what-we-do/research/silt.html

Call for IELTS Joint-funded Research Program and IELTS Masters Award proposals

All IELTS research activities are co-ordinated as part of a comprehensive and coherent framework for research and validation of the test. A major component of this framework is the IELTS Joint-funded Research Program which is sponsored jointly by British Council and IELTS Australia with support from University of Cambridge ESOL Examinations.

The IELTS Masters Award is awarded annually for the Masters-level dissertation or thesis in English which makes the most significant contribution to the field of language testing.

The IELTS Partners are sometimes able to provide other types of support for students and external researchers who are engaging in IELTS-related research projects (e.g. access to materials and/or test data). Enquiries about this type of assistance should be addressed to one of the partner organisations; each request will be considered in light of issues such as confidentiality, security and the resources needed to supply what is being asked for. For further information please visit the IELTS website: www.IELTS.org

The calls for proposals for the 2009 schemes are summarised below.

IELTS Joint-funded Research Program 2009–10

The IELTS partners are making available grant funding to a total value of around £120,000 (AUS\$225,000) for IELTS-related research projects to be conducted during 2010. Educational institutions and suitably qualified individuals are invited to apply for funding to undertake applied research projects in relation to the International English Language Testing System (IELTS). Financial support for individual projects selected will in principle be limited to a maximum of £15,000/AU\$36,000 though this upper limit may be reviewed in exceptional circumstances.

What areas of interest have been identified?

The IELTS Joint Research Committee is interested in supporting external researchers to conduct quality research under the three following headings. This year we would particularly encourage applications to undertake research studies which fit within these broad areas.

1. Test development and validation issues

- Studies investigating the IELTS General Training Reading

test, particularly stakeholder reactions to the revised section two.

- Investigation of the cognitive processes of IELTS test takers.
- Investigation of the process of writing IELTS test items.
- Investigation of L1 Writing and Speaking features that distinguish IELTS proficiency levels.

2. Issues relating to contexts of test use

- Studies to establish appropriate IELTS score levels in specific contexts (for access to a university department or vocational training course).
- Use of IELTS for professional registration purposes, or for purposes of migration, citizenship or employment.
- Studies investigating the use of IELTS and IELTS scores in local contexts, especially in-depth case studies focusing on individuals or small groups.
- Investigation of IELTS usefulness compared to other tests in similar contexts.

3. Issues of test impact

- Investigations of the role of interactive communication in the IELTS speaking or writing tests and the impact of this on candidates' preparedness for interactive communication in study or training contexts.
- Investigations of attitudes towards IELTS among users of test scores including admissions and academic subject staff in receiving institutions.
- Investigations of perceptions towards IELTS among test takers, teachers, learners and others engaged in preparing for IELTS (particularly in countries where there has been little or no previous IELTS related research).
- Test impact relating case studies in local contexts (e.g. teachers' experiences of working with IELTS).
- Case studies on English language progression among overseas students in undergraduate or post-graduate contexts.
- Case studies on in-session English programmes at English-speaking universities, particularly with consideration of the relationship between hours of study and typical score gains.

Consideration will also be given to other issues of current interest in relation to IELTS.

Annual timetable

30 June 2009	Applications close
July–August 2009	Preliminary review of applications
September 2009	Evaluation and selection of successful proposals
October 2009	Notification of Research Committee's decisions to applicants
January 2010	Research commences
September 2010	Research completed
December 2010	Final reports due

Proposals from researchers and institutions with established links with IELTS are welcomed. Institutions/individuals are invited to submit a written application (maximum of 10 pages, unbound) in accordance with the format and content requirements described in the *Guidelines* available online.

IELTS Masters Award

As part of the tenth anniversary of IELTS in 1999, the IELTS partners – British Council, IELTS Australia and University of Cambridge ESOL Examinations – established an annual award of £1,000 for the Masters-level dissertation or thesis in English which makes the most significant contribution to the field of language testing.

Submission and evaluation procedures

Each year the IELTS Research Committee, comprising members of the three partner organisations, reviews submissions for the award and shortlists potential winners. Submissions must be for a dissertation/thesis written in partial or total fulfilment of the requirements for a Masters

degree or its equivalent, and must be supported by a letter from the applicant's academic supervisor. The work should be language testing focused but need not be IELTS-related.

A full copy of all shortlisted dissertations/theses is then requested and a further reference may be sought. Shortlisted items are reviewed and evaluated by the IELTS Research Committee according to a set timetable and established criteria. The Committee's decision is final.

Annual timetable

30 June 2009	Deadline for submission of dissertation extracts and supervisor's reference to Cambridge ESOL
01 August 2009	Deadline for submission of full copies of shortlisted dissertations (and further references if required)
Oct–Nov 2009	Meeting of IELTS Research Committee
Nov–Dec 2009	Announcement of award

The award is normally presented in public at a major language testing event during the following year, e.g. at the annual Language Testing Research Colloquium (LTRC), and the IELTS partners sponsor the award winner's attendance at this event for this purpose.

Application forms and further information

Please note that submission details for both the IELTS Masters Award and IELTS Joint-funded Research Program may change from year to year. It is therefore important that the most current procedures are consulted – see the IELTS Masters Award and IELTS Joint-funded Research Program guidelines at www.ielts.org/researchers/grants_and_awards.aspx

Cambridge to host two major conferences

4th Cambridge Assessment Conference

Public policy experts, educationalists and assessment specialists will debate the role of the state in assessment systems at the 4th Cambridge Assessment Conference to be hosted by Cambridge Assessment on 19th October 2009. The conference theme – *Issues of control and innovation: the role of the state in assessment systems* – is a timely topic as a Bill to deliver Ofqual, a new regulator of qualifications, is currently going through UK Parliament.

Hosted by the Cambridge Assessment Network, the conference will make a major contribution to thinking in the area. The conference will include keynote sessions by renowned speakers Professor Alison Wolf (Sir Roy Griffiths Professor of Public Sector Management, King's College London), and Professor Robin Alexander (Fellow of Wolfson College at the University of Cambridge). Experts including: Professor Mary James (Institute of Education, University of London); Isabel Nesbit (Ofqual); and Dr John Allan (SQA); will lead a series of seminars on the day.

Leaders from many areas within education are invited to attend, including senior managers from schools, colleges

and universities, as well as those working for national and local education bodies, professional organisations, political parties, the media, awarding bodies and employers' organisations. For further information please visit: www.assessnet.org.uk

32nd Language Testing Research Colloquium

The 32nd LTRC will take place from 14–16 April 2010 at the University of Cambridge, UK, with pre-conference workshops on 12 and 13 April. The theme is *Crossing the threshold: investigating levels, domains and frameworks in language assessment*. This theme reflects contemporary research interests and concerns within the language testing and assessment community (at both national and international level) focusing on proficiency levels, language domains, and the role of interpretative frameworks of reference for language learning, teaching and assessment (cf. CEFR, Canadian Language Benchmarks). Please visit the conference website for further information: www.CambridgeESOL.org/LTRC2010/