Research Notes

Contents

Editorial Notes

Welcome to issue 35 of *Research Notes*, our quarterly publication reporting on matters relating to research, test development and validation within Cambridge ESOL.

This is the second of two issues that continue the theme of the ALTE 2008 conference that Cambridge Assessment hosted in April 2008. This event focused on the social and educational impact of language assessment and involved hundreds of delegates from many countries. In this issue we include contributions from Cambridge ESOL colleagues and external contributors, all of whom describe various aspects of the social impact of language assessment in a range of contexts. Alongside *Research Notes* issues 34 and 35 an ALTE 2008 Proceedings volume is being prepared for the Studies in Language Testing series which will be published in 2010.

The first three articles discuss the concept of washback, providing complementary views of this process using case studies of English language learners in different countries. Khaled El-ebyary reports an empirical study of how washback is linked to formative assessment, based on trainee teachers taking English writing courses in Egypt. He proposes a model which includes direct and indirect participants and inter- and intra washback processes. Next, Dina Tsagari reports a study that investigates the washback effect of the First Certificate in English on teaching and learning in private language schools in Greece in terms of its impact on teaching methods and contents, also on participants' feelings and attitudes. Thirdly, Nkechi Christopher describes the position of English language assessment in Nigeria within the wider context of education, by exploring various environmental factors and their impact on education and assessment.

The following pair of articles explore washback effects of tests of other languages in various contexts. Waldemar Martyniuk discusses the impact of the State Certificates in Polish as a Foreign Language on its stakeholders, reporting the results of a survey on their use. Next, Anne Gallagher and Siuán Ní Mhaonaigh explore the social impact of tests of Irish – a minority language – for adults. They report a project that involved many stakeholders, including test developers, test takers and other groups.

We next focus on relating assessment to real-world issues such as migration and language for specific purposes. Szilvia Papp and Martin Robinson report on using Cambridge ESOL's Skills for Life tests to exemplify a framework for assessing language proficiency for migration purposes. In the final article Ivana Vidakovic and Evelina Galaczi explore how the rating scale for the International Legal English Certificate (ILEC) speaking test was revised to ensure raters used the whole scale, thereby promoting fairness to all test takers and better rater practice.

We finish this issue with a number of conference reports and the announcement of the 2008 IELTS Masters Award winner.

Editorial team for issue 35: Fiona Barker and Hanan Khalifa.

Deconstructing the complexity of Washback in relation to formative assessment in Egypt

KHALED EL-EBYARY NEWCASTLE UNIVERSITY, UK AND ALEXANDRIA UNIVERSITY, EGYPT

Introduction

It has been acknowledged that producing improved learning seems to be a worthwhile aim for washback (Alderson 1993, Andrews, Fullilove and Wong 2002, Bailey 1996, Cheng and Curtis 2004, Hughes 2003, Shohamy 1996, Watanabe 1997, 2004). Put another way, the intended washback of washback research is to better understand the mechanism of this phenomenon in order to design assessments that will help improve learning. So, explication of how washback operates has been deemed important in order for assessment to maximize opportunities for enhancing positive washback and at the same time minimizing, if not eliminating, negative washback. However, it was Alderson and Wall (1993) who drew explicit attention to the complexity of the phenomenon in their rhetorically titled article 'Does Washback Exist?'. Reference to this complex nature has endured ever since (e.g. Cheng 1997, 1999, 2000, 2001, Fournier-Kowaleski 2005, Green 2007, Shohamy, Donitsa-Schmidt and Ferman 1996, Qi 2004, Watanabe 1997, 2004). In this respect, Alderson (2004:ix) again states in his foreword to Washback in Language Testing: 'I believe there is no longer any doubt that washback does indeed exist. But we now know that the phenomenon is a hugely complex matter, and very far from being a simple case of tests having negative impact on teaching'.

This article reports some of the initial findings of an empirical study of washback. We argue that the hugely complex nature of washback is due to the fact that it is an interactive multi-directional process involving a constant interplay of different degrees of complexity among the different washback components. Assessment is located in both a narrower educational and a broader social context, and the involvement of not only direct participants (teachers and students), but also indirect participants including those from the broader social context (education personnel, parents, media and the community) adds both greater importance to the washback phenomenon and different degrees of complexity (El-ebyary 2008).

In an attempt to explain the complexity of washback, we argue that two major washback processes are responsible for the washback complexity: *inter washback* and *intra washback* (El-ebyary 2008). Inter washback is viewed as a process intentionally or unintentionally initiated by one or more participants (e.g. teachers) which might have an effect on the attitudes and behaviour of other participants involved in the assessment situation (e.g. students). This would be the result of an explicit use of the power of assessment (e.g. developing a new test) to cause change to happen and/or the effects of a coded messaging scheme often used by some participants (e.g. teachers or assessors) whether intentionally or unintentionally where the sent messages get

read, misread or remain unread by other participants (e.g. students). Intra washback is a parallel process which involves an individual translating and acting upon the read, or misread, messages and the result would be a change in attitude and/or behaviour of that individual.

Inter washback and intra washback are not bound to one group of participants, but they are likely to be initiated by different participants, often at the same time, and hence the multi-directional nature of the phenomenon. Additionally, the two processes often result in washback product which can be positive, negative or even zero washback. In other words, the washback product of a certain assessment act is not necessarily an absolute negative or positive, but it can have a positive effect on certain washback areas, negative on others and yet zero washback on other areas (El-ebyary 2008). In this respect, the proposed model suggests that the washback product is likely to be dependent on the nature of the inter/intra washback process. Bailey's example of washback to the learner provides support for such a claim (Bailey 1996, 1999). According to Bailey (ibid.), washback to the learner refers to those effects of test-derived information provided to the learners as the test takers and having a direct impact on them. Students tend to display certain behaviour patterns or carry out certain actions when faced with an important test. A closer look at the sample actions provided by Bailey (ibid.) suggests that these actions might be as stereotypical as practising mock tests and memorising vocabulary and grammar or as interactive as practising interactive language (conversation). However, the argument suggested by Bailey (ibid.) is that expected washback here totally depends on which actions get selected as different students might choose different actions. We would categorise these actions as intra washback product, resulting from inter washback processes initiated by other direct or indirect participants as well as from the students' experience of previous assessment situations.

The washback model we propose here, which also provides the theoretical framework for this article, claims that washback product is the result of the constant interplay between all direct and indirect participants in the assessment process. This interplay is not only context bound, but it is most likely to vary between participants (inter washback) and within participants (intra washback). This variance can be further explained in terms of aspects of specificity, intensity, intentionality, length and value (see Watanabe 2004).

Focus and methodology

The proposed washback model provides the theoretical framework for this article where, the model claims,

washback is the result of a partnership between all *direct* and *indirect* participants whose relationships involve a constant multi-directional interplay. The particular focus of this article is the study of washback in relation to formative assessment. The theoretical framework we propose here is an attempt to identify the responsibility each participant would assume and explain why formative assessment to which assessment *for* learning is characteristic washback does not seem to attract the same washback research interest as final assessment or assessment *of* learning.

In an attempt to answer the question *Why does formative assessment have the washback effects it has?*, the study explores the mechanism of washback where potentialities of *inter* and *intra* washback are investigated at the level of direct participants (students and instructors). Questionnaires, interviews and focus groups were used to gather data from 586 EFL students and 35 instructors at English Sections in teacher-training colleges in Egypt.

Findings

In order to examine the inter/intra washback proposition, both qualitative and quantitative data were collected from students and instructors. One main concern was to examine students' awareness of any inter washback impact from other direct participants (instructors in this case) and whether or not this awareness instigated intra washback impact on the students' own attitudes and learning behaviour. While some findings revealed students' awareness of inter washback initiated by instructors in terms of attitudes and behaviour, further findings showed that students could precisely mark a relationship between the inter washback initiated and the intra washback impact on their own attitudes and learning behaviours. In addition, it was shown that this relationship did vary in terms of the *inter washback intensity* initiated by instructors compared to other participants. This article will provide a brief summary of some findings which we claim provide support for an inter/intra washback distinction. Detailed discussion of the aspects of the inter/intra washback interplay can be found in El-ebyary (in progress).

In examining the inter/intra washback distinction, responses in interviews and focus group sessions with students were highly indicative of different aspects of inter/intra washback interplay. Exploring students' perceptions of their instructors' assessment oriented behaviour was extremely informative for purposes of investigating whether or not such instructors' behaviours initiated inter washback impact on students' attitudes and/or learning behaviour. Like summative assessment, formative assessment is often seen in the target context as the sole responsibility of instructors. In fact, instructors teach, design tests, assign marking criteria and carry out anonymous (in the case of summative) paper marking. Moreover, examining the school regulations revealed that although these provided general parameters for the use of formative assessment in all subjects, these parameters were found to amount to a brief description of grade distribution. In practice, the school leaves the issue to be managed solely by instructors. However, these regulations elucidate the fact that formative assessments do count towards the final grade of students. Interestingly, during interviews and the focus group sessions it appeared that certain instructors' behaviour in relation to formative assessment was influential not only on students' own behaviour with regard to formative assessments, but on their self-image as learners on a course (writing in this case). One major concern was instructors' descriptions of the assessment criteria of formative assessment tasks in comparison to summative assessment. Generally speaking, the students' perception was that their instructors devoted much less effort to clarifying the criteria used in formative assessment than to explaining the marking criteria used in the final assessment. Further evidence was sought on students' perception of the extent to which instructors made assessment criteria for formative assessments available to them. On a three point scale, 549 students were asked to rank their perception of their instructors' behaviour in terms of the formative assessment criteria (see Figure 1).

Figure 1: Students' perception of how frequently instructors talk about formative assessment criteria
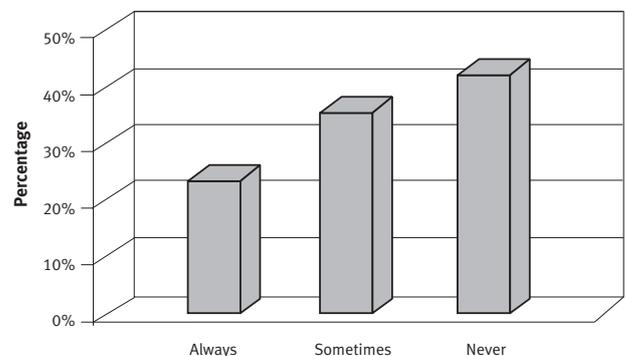
Interestingly, 42% of the subjects claimed that their instructors *never* make the formative assessment criteria clear to them. It also emerged that the fact that formative assessments do count towards the final grade of students was known by all instructors, but not necessarily by all students. In addition, although there was some evidence of instructors making students aware of the criteria used in formative assessment (18% of the students claimed that instructors *always* made them aware of the criteria, and 28% claimed that this was *sometimes* the case), many students claimed that instructors' efforts to clarify some formative assessment tasks were made by using those tasks as mock final exams. As a result, instructors' talk about the marking criteria was implicitly oriented towards the summative assessments rather than formative ones. Further findings across the four years of study were consistent in this respect (see Table 1).

Several comments also referred to the type of criteria instructors provide, and in writing courses correct grammar and vocabulary often appeared to be emphasised by instructors. Exactly what effect the students' perception of their instructors' behaviour had on their own behaviour was also a main concern here as it formulated the *intra washback impact*.

In investigating the intra washback impact of instructors' behaviours on students, it appeared that such behaviours instigated an intra washback not only in terms of students'

Table 1: Number of teachers who make formative assessment criteria known to students by year of study

Year of Study	No. of teachers who make criteria known to students			Total
	Always	**Sometimes**	**Never**	
First	2	29	42	73
Second	42	41	15	98
Third	106	105	59	270
Fourth	18	30	48	96
Total	**168**	**205**	**164**	**537**

perception of themselves as writers, but in terms of students' learning behaviour (preparing for the assessments). Many students' comments in the interviews showed some constructed association between students' self images as EFL learners in a writing course on the one hand and instructors' behaviour on the other. There appeared to be a general belief that a good writer would be one whose writing meets the instructors' marking criteria, which in this case focused only on lexical and grammatical accuracy. Examining the issue with a larger sample, 549 students were asked to rate their self-image as writers. They were also asked to explain their self-image, with no directions given as to what they should write. As expected, some students described themselves as *poor*, others as *good* and a third group were *not sure* (see Figure 2).

Figure 2: Students' perception of being a good writer in English
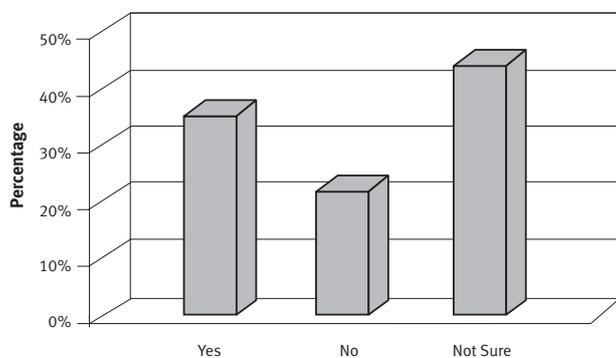
However, the way they accounted for their perceived *in/adequacies* was highly instructive as it showed the extent to which the inter washback of instructors' description of their marking criteria as well as the type of feedback provided was responsible for intra washback impact on students' perception of *good/bad* writers and consequently on their expected learning behaviours. Identifying at least one of the sources of their self-image in this way provides a clue to what actions they might have taken to maintain their positive self-image or even to improve a negative one.

Findings on the intra washback impact on students' self-image as writers show that 35% of students perceived themselves as *good* writers compared to 22% who viewed themselves as *poor* writers. The rest of the subjects (43%) were simply *not sure*. Nonetheless, an analysis of the explanations students gave for their self-

image in terms of writing indicated the possible influence of instructors' feedback practices. Some of their comments (translated from Arabic) are given below.

I am a good writer because:

- I try to learn a lot of grammar rules and vocabulary.
- I think I somewhat [sic] have a wealth of vocabulary and grammar rules.
- I pay attention to tenses and I study them over and over.

I am not a good writer because:

- I am not fully aware of English grammar and expressions which makes my English writing and the expressions I use closer to Arabic.
- I think I often have grammar mistakes.
- Poor background in grammar…I do not know punctuation till now.
- Not memorizing and retaining grammar rules for a long time, but rather studying them during the course time only.

Further examination of these self-images revealed that most of the responses students made were generally linked to instructors' assessment practices (e.g. type of feedback and assessment criteria) and this could elucidate one aspect of the interplay between *inter washback* and *intra washback*. A number of students indicated that they attended special tutoring sessions at private institutions to develop their language skills and by language skills they meant the knowledge and application of grammar rules and vocabulary.

Conclusion

This article has argued that washback is an interactive multi-directional process in which an ongoing interplay of different degrees of complexity takes place among the different elements of the process. This interplay occurs at the process level and is termed here *inter washback* and *intra washback*. However, the complexity of the phenomenon partially lies in the multi-directionality which often entails variance in inter/intra washback specificity, intensity, intentionality, length and value (Watanabe 2004). A review of washback research provides further support for the claims put forward in this article (Alderson and Hamp-Lyons 1996, Chen 2002, Cheng 1997, 1999, Cheng, Watanabe and Curtis 2004, Fournier-Kowaleski 2005, Green 2007, Scott 2007, Shohamy 1996, Shohamy et al. 1996, Wall and Alderson 1993, Wall 2000, Watanabe 1992). Similarly, a closer look at the strategies that promote positive washback, as described in the literature (Bailey 1996, Hughes 2003, Shohamy 1993) provides further support for our claims. This article also suggests that the different degrees to which these strategies, and others, are achieved, or *not achieved*, in assessments would initiate inter washback impact on students and consequently on the nature of their washback product.

References and further reading

Alderson, J C (2004) Foreword, in Cheng, L, Watanabe, Y with Curtis, A.

Alderson, J C and Hamp-Lyons, L (1996) TOEFL preparation courses: A study of washback, *Language Testing* 13/3, 280–297.

Alderson, J C and Wall, D (1993) Does washback exist?, *Applied Linguistics* 14/2, 115–129.

Andrews, S, Fullilove, J, and Wong, Y (2002) Targeting Washback: A Case Study, *System* 30/2, 207–23.

Bailey, K M (1996) Working for washback: A review of the washback concept in language testing, *Language Testing* 13/3, 257–279.

— (1999) *Washback in language testing*, Report RM–99–04, TOEFL-MS-15, Princeton, New Jersey: Educational Testing Service.

Chen, L-M (2002) Washback Of A Public Exam On English Teaching, *Educational Resources Information Center on Assessment and Evaluation (ERIC/AE) ERIC Number: ED 472167*.

Cheng, L (1997) How does washback influence teaching? Implications for Hong Kong, *Language and Education* 11/1, 38–54.

— (1999) Changing assessment: Washback on teacher perspectives and actions, *Teaching and Teacher Education* 15/3, 253–271.

Cheng, L and Curtis, A (2004) Washback or backwash: A review of the impact of testing on teaching and learning, in Cheng, L, Watanabe, Y, with Curtis, A, 3–18.

Cheng, L, Watanabe, Y and Curtis, A (2004) Concepts and Methodology of Washback, in Cheng, L, Watanabe, Y, with Curtis, A, 3–52.

Cheng, L, Watanabe, Y with Curtis, A (Eds) *Washback in Language Testing: Research Contexts and Methods*, Mahwah, NJ: Laurence Erlbaum and Associates.

El-ebyary, K (2008) Addressing the complexity of washback, paper presented at 3rd International Conference of the Association of Language Testers in Europe (ALTE), Cambridge, UK, 10–12 April.

— (in progress) *Computer-based feedback on students' written work: the washback impact on students and on formative assessment practices*, PhD thesis, Newcastle University, UK.

Fournier-Kowaleski, L A (2005) *Depicting washback in the intermediate Spanish language Classroom: a descriptive study of teachers' instructional behaviors as they relate to tests*, unpublished PhD thesis, The Ohio State University.

Green, A (2007) *IELTS Washback In Context: Preparation For Academic Writing In Higher Education*, Studies in Language Testing vol. 25, Cambridge: UCLES/Cambridge University Press.

Hughes, A (2003) *Testing for language teachers*, Cambridge: Cambridge University Press.

Scott, C (2007) Stakeholder perceptions of test impact, *Assessment in Education: Principles, Policy and Practice* 14/1, 27–49.

Shohamy, E (1993) A collaborative/diagnostic feedback model for testing foreign languages, in Douglas, D and Clapham, C (Eds), *A new decade of language testing research*, Alexandria, VA: TESOL Publications, 185–202.

— (1996) Test impact revisited: Washback effect over time, *Language Testing* 13, 298–317.

Shohamy, E, Donitsa-Schmidt, S and Ferman, I (1996) Test impact revisited: washback effect over time, *Language Testing*, 13/3.

Wall, D (2000) The impact of high-stakes testing on teaching and learning: can this be predicted or controlled? *System* 28/4, 499–509.

Wall, D and Alderson, J C (1993) Examining washback: The Sri Lankan impact study, *Language Testing* 10/1, 41–69.

Watanabe, Y (1992) Washback effects of College Entrance Examination on language learning strategies, *JACET Bulletin* 23, 175–194.

— (1997) Constructing a classroom observation scheme for the test impact research, *Sophia Linguistica* 41, 297–313.

— (2004) Methodology in Washback Studies, in Cheng, L, Watanabe, Y with Curtis, A, 19–36.

Revisiting the concept of test washback: investigating FCE in Greek language schools

DINA TSAGARI DEPARTMENT OF ENGLISH STUDIES, UNIVERSITY OF CYPRUS

Introduction

It has long been noted that high-stakes language exams exert a powerful influence on language learning and teaching, a phenomenon known in the language testing literature as the 'washback effect' (Alderson and Wall 1993). However, even though the issue of washback has long been discussed in the literature of general education (Frederiksen and Collins 1989, Haladyna, Nolen and Haas 1991, Kirkland 1971, Popham 1987), it was not until the early 1990s that washback attracted attention from writers and researchers in the field of language testing. The three basic models of washback were proposed during this decade (Alderson and Wall 1993, Hughes 1994 and Bailey 1996) coupled with research in a number of countries around the globe (see Table 1) revealed several interesting findings with regard to its nature and the ways it functions.

The research studies showed that rather than being a direct and automatic effect, washback is a complex phenomenon. Furthermore, even though studies showed that washback exists on a variety of teaching and learning areas (e.g. curriculum, methods of teaching, classroom

Table 1: Overview of the research literature on test washback

Context	Washback Researchers
Brazil	Retorta (2008)
Canada	Saif (2006)
Central/Eastern Europe	Wall and Horak (2006)
China	Qi (2004)
Hong Kong	Andrews (1994a, 1994b, 2002); Cheng (2005); Lam (1993)
Hungary	Kiss-Gulyas (2001); Glover (2006)
Israel	Shohamy (1993); Shohamy, Donitsa-Schmidt and Ferman (1996); Ferman (2004)
Japan	Watanabe (1997; 2001)
New Zealand	Read and Hayes (2003)
Romania	Gosa (2004)
Spain	Amengual and Herrera (2008)
Sri Lanka	Wall and Alderson (1993); Wall (2005)
UK	Saville and Hawkey (2004); Hawkey (2004); Scott (2007)
USA	Alderson and Hamp-Lyons (1996); Stecher, Chun and Barron (2004)

assessment, student learning, feelings and attitudes of teachers and students), it was also found that washback varied in 'form' and 'intensity' (Cheng 2005) as well as in 'specificity', 'length', 'intentionality' and 'value' (Watanabe 2004).

Washback was also found to be broad and multi-faceted and brought about through the agency of many independent and intervening variables beside the exam itself. These include teachers and students, the status of the subject being tested, resources and classroom conditions, management of practices in the schools, communication between test providers and test users and the socio-political context in which the test is put to use.

However, while theory and research so far have succeeded in formulating a general framework for the washback phenomenon that involves several components and intervening factors, there is still a need to account for the interplay between them. Actually, an important step forward would be to construct a conceptual model of washback which would provide a synthesis of its multiple components and demonstrate the interaction between them and the factors involved. This would provide a better understanding of the reasons why washback occurs and an explanation of how tests can be designed in order to engineer positive washback.

Aim of the study

The present research study set out to investigate the 'washback effect' of a high-stakes language test, namely the First Certificate in English (Cambridge ESOL), on the teaching and learning that takes place in Greek private language schools (also known as 'frontistiria').

Preliminary steps

To help resolve uncertainty of whether washback would actually result from the exam under study or not (Cheng 1998, Watanabe 2004), a preliminary study was conducted. The study addressed the following general research questions:

• Does washback from the FCE exam exist?

• What is its nature and effect on the contents and methods of teaching?

• What is its effect on the feelings and attitudes of teachers and students?

Fifteen native and non-native FCE teachers were interviewed using a specially-designed interview guide (Patton 1987). The results showed that teachers believed that there was a close relationship between the exam and the teaching and learning that went on in their classes. For example, teachers claimed that what was taught and learned was dominated by the extensive use of exam preparation materials and teaching to the test practices. The exam encouraged teachers and students to place more value on the skills and activities that were assessed in the exam and less value on those that were not.

Teachers also stressed that the exam influenced the textbooks they used in that these contained activities and materials that reflected the format and content of the exam.

There were also claims that the textbooks, and the accompanying test booklets, mediated the impact of the exam on the classroom tests.

The FCE exam was also said to have detrimental effects on students' attitudes towards learning the language as it created feelings of stress, anxiety, boredom and demotivation. Teachers also believed that students did not enjoy learning the language as such but viewed it as a means of passing the exam. The exam also appeared to have negative effects on teachers, too. Due to its high-stakes usage, success on the exam was a way to judge teachers' professional value. This made them feel accountable to a number of stakeholders and in turn led to their anxiety and stress.

Main study

Based on the results of the interviews, further research was undertaken in order to confirm teachers' claims about the washback of the exam. The research questions the main study addressed were:

• Does the FCE exam influence the nature of teaching materials teachers use in their FCE classes? If yes, what is the nature and scope of the influence?

• Does the FCE exam induce feelings of stress and anxiety in students?

• Does the FCE exam make students feel bored and demotivated?

• Does the FCE exam influence students' attitudes towards learning English?

Methodology

To answer the first question, three internationally- and two locally-produced FCE textbooks as well as two general English textbooks (accompanied by their Teachers' Books) were analysed using a specially-designed instrument based on the initial pilot version of the *Instrument for Analysis of Textbook Materials* (Bonkowski 1996). The instrument used went through rigorous stages of adaptation and piloting. In its final version, the *Checklist for Analysis of Textbook Materials for the FCE exam* (CATM-FCE) contained a 13-page checklist divided into two stages and eight parts consisting of 78 questions on various textbook features (see Tsagari 2007). The data collected from the analysis of the textbook materials were stored and analysed using Microsoft *ACCESS 2003* and *SPSS 11.5*.

To answer the remaining research questions, a diary study was set up (see Tsagari 2007) that involved 29 FCE students who used one of the exam textbooks analysed in the previous study. For the analysis of the data two qualitative programmes were used, Atlas.ti (Muhr 1997) and WordSmith (Scott 1999) and a specially-devised coding scheme that went through several stages of piloting (see Tsagari 2007).

Results of the textbook analysis

The findings of the textbook study revealed that the influence of an exam on textbook materials is a complicated phenomenon. First of all, the analysis revealed that the

exam did influence the contents of the textbooks. It seems that writers and, by extension, publishers of the FCE exam-preparation materials operationalised the exam specifications into practical textbook activities in order to meet the needs of prospective candidates and their teachers. In some cases also, the influence had positive washback as the textbooks, by adherence to exam specifications, included a wider range of sources of input and language elements compared to their equivalent general English language textbooks.

However, the findings revealed that there were differences among textbooks while, at times, authors and publishers inaccurately reflected parts of the test specifications. For example, there were instances where exam features (e.g. pronunciation and scoring criteria) were underrepresented while others such as grammar and vocabulary were overemphasized. In addition, some textbooks and individual textbook features were more tightly focused on the exam than others. In other cases, the textbooks were so faithful to the exam that they failed to represent the full range of language skills and tasks needed at this level (in terms of authenticity of texts and writing genres) creating negative washback. Finally, the exam textbooks, much like the general English textbooks, presented and sequenced their tasks in similar ways, offered test practice exercises and incorporated non-exam task types, too.

Results of the diary analysis

The analysis of the diaries revealed that the exam influence was not direct but mediated through a variety of factors. First of all, the influence of the exam was mediated through the exam-oriented textbook used. The analysis of the diaries showed that the content and sequence of classroom teaching and assessment emanated from the exam-preparation textbook materials. In fact, in the absence of any official syllabus, the syllabus of the exam textbook became the course syllabus and determined the content of teaching and classroom assessment leading to *textbook washback*.

Secondly, the study showed that the teacher shared with the exam textbook a central role in mediating the process of exam washback. Evidence in the diaries showed that the teacher reshaped the exam textbook during instruction by reinterpreting the textbook in terms of methods, for example by adding extra techniques and structuring the lessons of the day in her own way leading to *teacher washback*.

Other than the above, the results revealed that presence of 'washback to the learner' was particularly strong (Bailey 1996). The diaries showed that students experienced various levels of stress and anxiety throughout the year confirming their teachers' claims. Evidence of more intensive washback was recorded in the diaries as the date of the exam drew closer. This reached a peak in the weeks prior to its administration and was accompanied by intense physical reactions such as upset stomach, headache, and sickness.

Diaries also gave a general feeling of dissatisfaction. Students did not enjoy their classes and felt tired, disinterested and bored. This led to lack of attention, participation and discipline during their exam-preparation lessons.

However, an interesting finding with regard to the interplay between washback, motivation and anxiety was observed. In fact, despite the negative feelings experienced, students were determined to acquire the FCE certificate. The status of the exam, an important factor in the presence of washback (Alderson and Wall 1993, Shohamy et al. 1996), made them willing to expend time and effort to pass the exam and put up with the uninteresting activities, materials and methods used. In students' minds, the exam was mainly associated with instrumental benefits, such as future employment or further studies abroad. In fact, students came to consider studying for the test as language education, and passing the exam as the aim of their schooling and evidence of educational excellence. Students valued the FCE certificate so highly that they believed it would enhance not only their linguistic ability but that it would also increase their personal confidence and self-esteem.

Given the importance placed on the exam, students were also more interested in the utilitarian benefits of the language rather than the desire for communication with the speakers of English and their community. Students believed that the immediate goal of learning English was to achieve success in the FCE. What mattered more to them was the acquisition of the qualification itself rather than learning the language.

In addition to the above, the diaries provided evidence of what Bailey (1996) terms 'washback to the programme'. The presence of a number of other stakeholders in the immediate teaching context and broader environment also influenced what was happening in the classroom. For instance, the school atmosphere nurtured and supported test washback. In fact the school enforced strict rules (e.g. students had to make amends for missed classes and tests), maximised homework set and added extra intense exam preparation classes to the weekly programme as the dates of the exam drew nearer. This created an exam-like atmosphere in the classroom and, in some cases, lowered the students' self-esteem, giving them the impression that they would not succeed in the exam.

The study also revealed that the role of parents was crucial in the washback process. Washback to parents was mediated both through the school (progress reports and parents' meetings) and manifested in behaviour such as exhorting students to work hard. The findings also revealed that, very often, parents affected students' language motivation (e.g. prompting students to take up learning English at a young age) and instigated an instrumental disposition towards English. Parents' beliefs about the value of the exam and the language were actually shaped by the local context which recognised the FCE certificate as an official language qualification and placed a great deal of importance on learning/mastering the language.

Finally, there is evidence in the diaries which showed that local teaching practices and beliefs about language teaching and learning also played an important role. For instance, the prominence of grammar teaching and dictation writing revealed in the diary data was reinforced

by similar practices employed for the teaching of English as a subject in the Greek mainstream schools.

Implications of the study

The present research began with a very broad question about the effect of the FCE exam. As the research progressed it became clear that the nature of the exam washback was not as straightforward as the language teachers believed. In fact, the findings suggested that test washback is a very complex process.

In summary, the study showed that while high-stakes language tests, such as the FCE exam, could positively affect teaching and learning, they cannot have exclusive power over what is happening in the classroom. Rather other factors, in direct or indirect relation to the exam, seem to be playing a greater role or a role at least as significant as the exam in shaping the classroom teaching and learning and contributing (or not) towards positive washback.

The various factors that need to be taken into account when an attempt is made to induce positive washback of high-stakes tests include the following stakeholders: textbook writers and publishers, teachers, students, schools, parents, local educational systems and local society. It is also possible that other factors, like mass media (ELT newspapers, newsletters, TV and radio) might have interfered with the FCE constructors' efforts to bring about the intended washback effect but these were not explored in the present study.

However, the above factors, likely to enhance or interfere

with the test's intended purpose, do not function in isolation. The picture that emerged from this study is that the influence of the exam at the classroom level is a dynamic and interactive process between the exam, the textbook, the teacher and students, each influencing the other, in varying degrees of quantity and quality. This interaction is in turn shaped by the interrelation of several other related factors which operate within the immediate and broader context, such as test constructors, material writers and publishers, schools, parents and the local education system. These factors can then be seen as additional agents in helping to engineer or hamper the intended washback of an exam.

Conclusion

The present study has examined washback from a high-stakes exam, taking as its specific focus teachers, exam-preparation materials and learners. The results showed that high-stakes tests, powerful as they are, might not be efficient agents for profound changes in an educational context. While the study reiterates the complexity of investigating washback noted by previous studies, it also provides an indication as to the sources of this complexity that can be traced both inside and outside the classroom context.

Figure 1 illustrates the complex ecology of exam washback as delineated in the present study. In the model, washback is represented as an open loop process identifying the number of stakeholders involved in the

Figure 1: Proposed model of washback
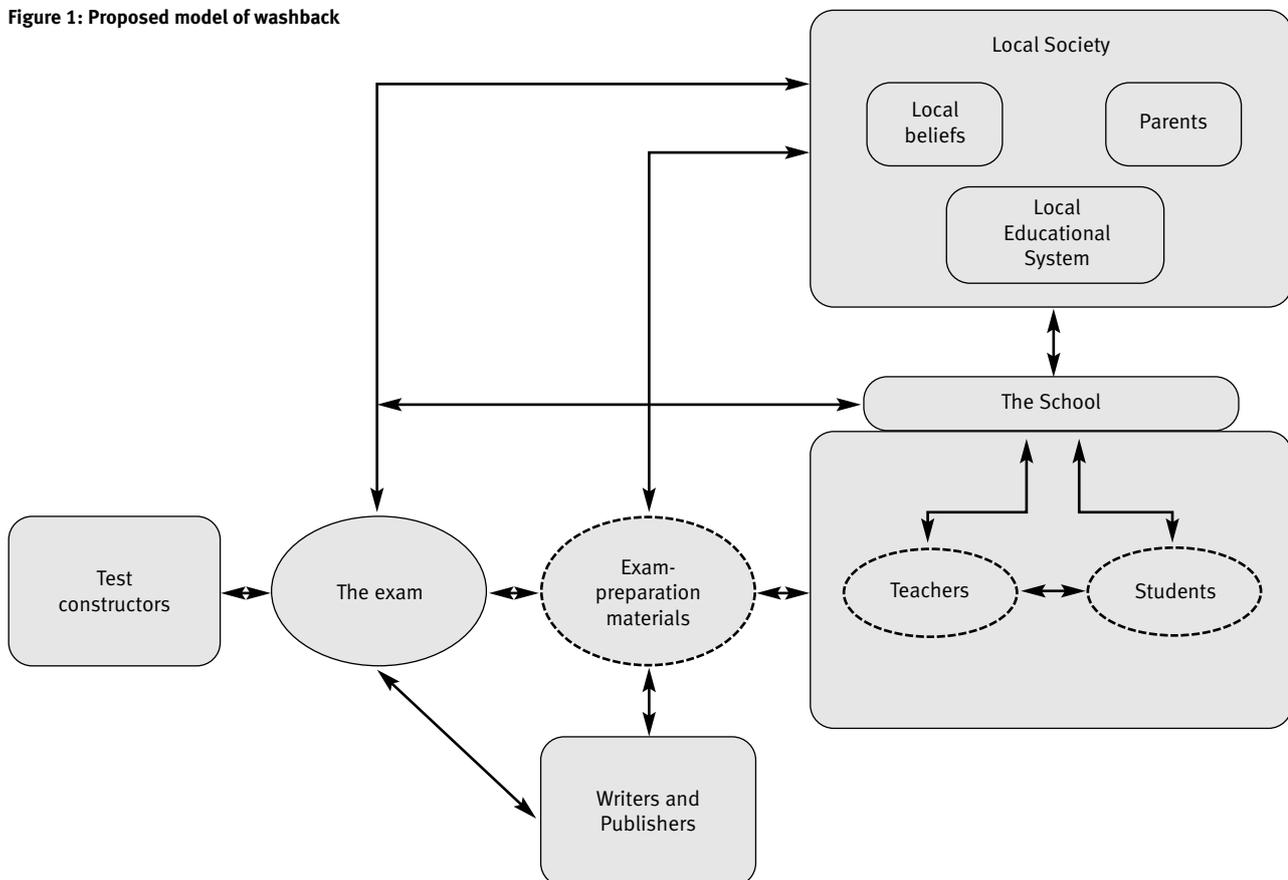
process and attempting to portray the relationship between them. However, despite it being a multi-directional relationship among stakeholders, the model, in its visual representation below, is 'simplified' to make it possible to represent it graphically.

In the model, the nature of exam washback is circuitous and interactive. Exam washback is indirectly engineered on teaching and learning that takes place in the exam-preparation classroom through the understanding of the exam requirements by various intermediary participants. First and foremost, exam washback is mediated through commercial exam-preparation materials that are shaped by the perceptions of the needs of teachers and students by writers and publishers. The exam-preparation materials mediate between the exam intentions and the exam-preparation class. The teacher's role is also crucial in the process as they mediate between material and students. Within this process, washback is also mediated by the school and strengthened by the perceptions and understanding of various other stakeholders operating in the wider local community, such as parents, as well as by the local educational system and beliefs about the exam and the language tested.

References and further reading

Alderson, J C and Hamp-Lyons, L (1996) TOEFL Preparation Courses: A Study of Washback, *Language Testing* 13, 280–297.

Alderson, J C and Wall, D (1993) Does Washback Exist? *Applied Linguistics* 14, 115–129.

Amengual, M and Herrera, H (2008) *The influence of the Spanish University Entrance Examination (SUEE) on the teaching of English*, paper presented at 3rd International Conference of the Association of Language Testers in Europe (ALTE), Cambridge, UK, 10–12 April.

Andrews, S (1994a) The Washback Effect of Examinations: Its Impact Upon Curriculum Innovation in English Language Teaching, *Curriculum Forum* 4, 44–58.

— (1994b) Washback or Washout? The Relationship between Examination Reform and Curriculum Innovation, in Nunan, D and Berry, R (Eds) *Bringing About Change in Language Education*, Hong Kong: Department of Curriculum Studies. University of Hong Kong, 67–81.

— (2004) Washback and Curriculum Innovation, in Cheng, L, Watanabe, Y with Curtis, A, 37–52.

Bailey, K M (1996) Working for Washback: A Review of the Washback Concept in Language Testing, *Language Testing* 13, 257–279.

Bonkowski, J F (1996) *Instrument for the Assessment of Teaching Materials*, unpublished manuscript, Department of Linguistics and Modern English Language, Lancaster University, Lancaster, UK.

Cheng, L (1998) Impact of a Public English Examination Change on Students' Perceptions and Attitudes toward Their English Learning, *Studies in Educational Evaluation* 24, 279–301.

— (2005) *Changing Language Teaching through Language Testing: A Washback Study*, Studies in Language Testing vol. 21, Cambridge: UCLES/Cambridge University Press.

Cheng, L, Watanabe, Y with Curtis, A (Eds) (2004) *Washback in Language Testing: Research Context and Methods*, Mahwah, NJ: Lawrence Erlbaum Associates.

Ferman, I (2004) The Washback of an EFL National Oral Matriculation Test to Teaching and Learning, in Cheng, L, Watanabe, Y with Curtis, A, 191–210.

Frederiksen, J R and Collins, A (1989) A Systems Approach to Educational Testing, *Educational Researcher* 18, 27–32.

Glover, P (2006) *Examination Influence on How Teachers Teach: A Study of Teacher Talk*, unpublished PhD thesis, Department of Linguistics and Modern English Language, Lancaster University, Lancaster, UK.

Gosa, C M C (2004) *Investigating Washback: A Case Study Using Student Diaries*, unpublished PhD thesis, Department of Linguistics and Modern English Language, Lancaster University, Lancaster, UK.

Haladyna, T M, Nolen, S B and Haas, N S (1991) Raising Standardized Achievement Tests Scores and the Origins of Test Score Pollution, *Educational Researcher* 20/5, 2–7.

Hawkey, R (2004) The CPE Textbook Washback Study, *Research Notes* 20, 19–20.

Hughes, A (1994) *Backwash and TOEFL 2000*, unpublished manuscript, commissioned by Educational Testing Service (ETS), University of Reading, UK.

Kirkland, M C (1971) The Effects of Tests on Students and Schools, *Review of Educational Research* 41, 303–350.

Kiss-Gulyas, J (2001) Experiencing the Examination Design, Content, Materials and Procedures, in Együd, J G, Gal, I A and Glover, P (Eds) *English Language Education in Hungary, Part III: Training Teachers for New Examinations*, Budapest: The British Council, 40–58.

Lam, H P (1993) *Washback-Can It Be Quantified? A Study on the Impact of English Examinations in Hong Kong*, unpublished MA dissertation, University of Leeds, UK.

Muhr, T (1997) *Atlas.Ti: The Knowledge Workbench*, Berlin: Scientific Software Development.

Patton, M Q (1987) *How to Use Qualitative Methods in Evaluation*, Newbury Park, California: Sage Publications.

Popham, W J (1987) The Merits of Measurement-Driven Instruction, *Phi Delta Kappan* 68, 679–682.

Qi, L (2004) *The Intended Washback Effect of the National Matriculation English Test in China: Intentions and Reality*, Beijing: Foreign Language Teaching and Research Press.

Read, J and Hayes, B (2003) The Impact of IELTS on Preparation for Academic Study in New Zealand, *IELTS International English Language Testing System Research Reports* 4, Canberra/IELTS Australia, 153–206.

Retorta, M S (2008) *Washback effects of entrance examinations on high school teaching/learning/assessment in Paraná, Brazil*, paper presented at 3rd International Conference of the Association of Language Testers in Europe (ALTE), Cambridge, UK, 10–12 April.

Saif, S (2006) Aiming for Positive Washback: A Case Study of International Teaching Assistants, *Language Testing* 23, 1–34.

Saville, N and Hawkey, R (2004) The IELTS Impact Study: Investigating Washback on Teaching Materials, in Cheng, L, Watanabe, Y with Curtis, A, 73–96.

Scott, C (2007) Stakeholder perceptions of test impact, *Assessment in Education* 14, 27–49.

Scott, M (1999) *Wordsmith Tools*, Oxford, UK: Oxford University Press.

Shohamy, E (1993) *The Power of Tests: The Impact of Language Testing on Teaching and Learning*, National Foreign Language Center Occasional Papers, Washington, DC: National Foreign Language Center.

Shohamy, E, Donitsa-Schmidt, S and Ferman, I (1996) Test Impact Revisited: Washback Effect over Time, *Language Testing* 13, 298–317.

Stecher, B, Chun, T, and Barron, S (2004) The Effects of Assessment-Driven Reform on the Teaching of Writing in Washington State, in Cheng, L, Watanabe, Y with Curtis, A, 53–72.

Tsagari, K (2007) *Investigating the Washback Effect of a High-Stakes EFL Exam in the Greek context: Participants' Perceptions, Material Design and Classroom Applications*, unpublished PhD thesis, Department of Linguistics and English Language, Lancaster University, UK.

Wall, D (2005) *The Impact of High-Stakes Examinations on Classroom Teaching: A Case Study Using Insights from Testing and Innovation Theory*, Cambridge: Cambridge University Press.

Wall, D and Alderson, J C (1993) Examining Washback: The Sri Lankan Impact Study, *Language Testing* 10, 41–69.

Wall, D and Horak, T (2006) *The Impact of Changes in the TOEFL® Examination on Teaching and Learning in Central and Eastern Europe: Phase 1, the Baseline Study*, TOEFL Monograph Series, report Number: RR–06–18, TOEFL-MS–34, Princeton, NJ: Educational Testing Service, retrieved from: http://www.ets.org/

Watanabe, Y (1997) *The Washback Effects of the Japanese University Entrance Examinations of English-Classroom-Based Research*, unpublished PhD thesis, Department of Linguistics and Modern English Language, Lancaster University, Lancaster, UK.

— (2001) Does the University Entrance Examination Motivate Learners? A Case Study of Learner Interviews, in Akita Association of English Studies (Ed.) *Trans-Equator Exchanges: A Collection of Academic Papers in Honour of Professor David Ingram*, Akita, Japan: Akita Association of English Studies, 100–110.

— (2004) Methodology in Washback Studies, in Cheng, L, Watanabe, Y with Curtis, A, 19–36.

Interrelation between environmental factors and language assessment in Nigeria

NKECHI M CHRISTOPHER DEPARTMENT OF COMMUNICATION AND LANGUAGE ARTS, UNIVERSITY OF IBADAN, NIGERIA

Introduction

The aim of English language teaching (ELT) in Nigeria should be to enable learners to use the language to think, learn, solve problems (Shafer 2001 quoting Ken Goodman) and function in society. Otherwise, English language teaching and assessment will fail to drive national development and unite people. According to Young and Fletcher (2000) assessment is the key factor in monitoring literacy development, and research reveals that there is a strong connection between assessment practice and learners' success (Trevisan 2002).

This article explores the relationship between some environmental factors and suboptimal performance of Nigerian learners of English as a second language in language examinations. The standard of English spoken by Nigerians, as well as examination performance, continues to fall despite increasing exposure to different communication media using English extensively, indicating that language assessment has not contributed immensely to successful language learning. Other factors contributing to poor achievement in language examinations and language use are discussed.

Socio-economic and socio-cultural backgrounds of the test taker

It is evident from research that second language acquisition and literacy development are impacted upon by learners' socio-economic background (Dunn 2001, Gonzalez 2001) and that a suitable environment enhances learners' mental ability (Braid 2007). According to Badger and Wilkinson (1998) Australian students from poor and disadvantaged communities do not perform as well as children from wealthy families in literacy tasks; thus they are placed in disadvantaged schools. Taken together, these two factors suggest that the majority of Nigerian learners of English

from poor socio-economic backgrounds (as many as 71% living below the poverty line) should expectedly be handicapped in learning literacy and language skills.

The Nigerian learning environment, as a product of a poorly managed system in which capacities required for development and growth trail behind the momentum of the time, is hostile to learning. Although Nigerians are poor, Nigeria is not poor. But due to uncoordinated design and haphazard implementation of policies in a system condoning entrenched corruption, Nigeria has failed to attain deserved heights over the years. All infrastructures – economic, social, political and physical – stagnate, recede or advance in fits and starts, and educational structures are changed on whims, giving the impression that there are no professionals in the ministry of education. Things are permanently in a state of flux, with no certainty that ministerial changes are not tantamount to a new regime. For example, some years back the government took over all schools from their corporate owners, but today people are being asked to adopt a school, as the government seems unable to run an effective basic education system all by itself. The government, in participating in a world agenda of literacy for all and the eradication of illiteracy, instituted nine years' compulsory education, which it, however, has failed to back up with proper funding and management.

Educational standards degenerate just as the existing physical school structures do (Sanyaolu 2008), and the momentum with which the country embraced education-for-all and declared free universal basic education wanes. Despite the country's wealth in oil and mineral resources, arable land and human capital, steps that signify development are often stalled, and Nigerians are yet to be given a well-rounded education. For example, before Nigerians could understand what a telecommunications satellite is and its relevance to society, it is speculated that the first satellite went missing only 18 months after it was

launched (Brown 2008, Nuhu-Koko 2008). Moreover, computers are not in use or available in schools: there is no electricity in schools, teachers are not computer literate, and the will to introduce them is still lacking. The implication is that all digital resources for language learning and teaching continue to elude the Nigerian learner.

Education is often linked with the political philosophy of a people, and is used as a means of entrenching political values. Educational policies and practices concur with people's defined lines of reasoning or philosophy, for example, the educational philosophies of Indonesia, China, Denmark and Scandinavia derive from their national ideologies (Braid 2007). Perhaps Nigeria, as a nation struggling with development and definition of national orientation should adopt, as the Social and Human Sciences Committee of the UNESCO National Commission (UNACOM) (Braid 2007), a continuing national dialogue on education and language development in order to raise consciousness and make people participate in determining their destiny and national development strategies.

Nigeria has been ambivalent about its language policy: recognising bilingual education and recommending multilingualism, but creating no efficient and effective means of ensuring either. Thus the language policy is applied as convenient, creating semilinguals, with neither the English language nor local languages being the better off for it. In spite of many parents adopting English as their children's first language, the extensive use of English in the learner's environment and the pervasiveness of mass media, the standard of English continues to fall, and performance in other subject areas with it; in the 2008 West African Senior Secondary Certificate Examination, only 13.76% achieved the minimum five credit passes inclusive of English and Mathematics (Olugbile 2008).

Many other factors in society constrain learners' literacy and educational development. The learning and home environments do not enhance intellectual development. Not only are the time and facilities available for learning limited; basic infrastructures are in shambles; the zeal for learning cannot develop sufficiently, and may in fact be suffocated by fumes and noise from generators in the cities and lack of money to purchase substitute light sources in rural and urban settings. Living costs are prohibitive and many students do not have access to even the most basic learning materials.

Additionally, test takers' home background is yet to respond to the wider developments in society. In earlier times, students intermarried the knowledge gained from home with that from school, as exemplified in the writings of renowned Nigerian authors such as Chinua Achebe, Wole Soyinka, J P Clark and Cyprian Ekwensi. Today the home environment hardly provides the child with the impetus for educational development and discipline, as parents are either busy trying to eke out a living or to amass more wealth, and the older generation that could fill this gap are in distant villages. Yet parents and guardians hold the old belief that children will get all the tutelage they require from school, whereas school monitoring and management have been compromised by government which has also reduced community participation through government take-over of schools and declaration of free education. Teaching

standards are falling, classrooms bulging, and necessary teaching resources diminishing. Shortcomings in school setup could be overcome by individual learners if there is strong home and community participation in a child's literacy development. Prior to the take-over of schools by government, communities, churches and Islamic societies were involved in school establishment and management and were able to enforce discipline and provide the infrastructures and supervision supportive of learning.

Generally, the environment of learning in Nigeria demands more individual effort in learning a language than in environments rich in language use and literacy activities. However, the independent learning skills required for such learning may be lacking. Thus Sanyaolu (2008:69) observes that:

> 'It is a miracle that public schools are still producing brilliant and highly intelligent pupils in spite of the odds. But such pupils are getting fewer by the day, as they give way to pupils from private institutions many of which are better equipped.'

Language acquisition goes with other learning in the environment, and the more a rich environment is discussed, the more the learner can express experience. Implicitly, lack of experience and/or context for expressing experience in the target language will limit learners' communication capabilities. Also, a literate background will further expose learners to higher levels of learning and its absence has been known to create disadvantaged learners (Dunn 2001). In the absence of a home environment supportive of language and literacy development, the school should provide the ideal learning environment.

While the majority of learners learn their mother tongue at home and are comfortable within their cultural experience, little or no literacy activity is carried out in it, and so there is no literacy experience to transfer to the target language as a child enters school and is being taught literacy in a new language. Perhaps the child could relate literacy learning experience with the environment if the language being taught was one he is already familiar with, rather than the English language, which is very different from the mother tongue in all linguistic respects. Although in a few cases families teach English as a first language to their children, it is rarely accompanied by any literacy activities. Most literate parents (who are in the minority) are neither aware nor know how to encourage their children to extend their language experience with literacy activities.

Books and reading as factors in language assessment

Nigeria is still very much an oracy-based society: literacy and reading are not yet significant features of the Nigerian family or culture, and people would rather spend their money on other pleasures than on books. The low book output bears credence to this (Christopher 2007).

Problems of books and readership, key factors in language learning, have persisted since Nigeria embarked on the development of her human capacity. Meanwhile, in relative terms, the new millennium Nigeria is yet to attain the education standard and book production level of the

1970s when free universal primary education was declared. In the 1970s books were generously published and liberally imported from outside the country to satisfy a growing thirst for knowledge. However, the economic recession of the 1980s saw all areas of production plummet and recede (Uwalaka 2000). For several years, locally produced textbooks were scarce and unaffordable (NERDC 1988), and imported books were exorbitantly priced due to the Nigerian currency (naira) exchange rate nose-diving against other currencies. Today, rates of access to books and readership are pathetically low, the book industry having tumbled during the critical era of SAP due to the unfriendly economic environment and policies (Uwalaka 2000). Christopher's (2007) study reveals that book ratios in Africa are inadequate to encourage the development of readership – book ratios are far below one title where other countries have tens of titles per 100,000 people, and below ten where other regions have hundreds of copies per 100 people.

Since the country's recovery from economic recession, but not poor economic management, the book industry has failed to produce books that can entice readers and that the majority of parents can afford, while book importation is still expensive. Consequently, a dangerous apathy towards books and reading is growing as generation after generation of students are denied books they need for their studies and reading pleasure, especially at the critical age when a reading habit is formed. Additionally, libraries are non-existent in many primary and secondary schools.

The potential market for books is enormous, yet publishers service only a small portion due to poor patronage by parents and governments. The growing enrolment figure for primary schooling stood at 24.6 million in 2003 (Osuji 2004), but by the 2006/07 session two out of five household name publishers could only sell 663,529 copies (about 2.7%) of primaries 1–6 English textbooks instead of about 35–40%, and 41,902 copies of supplementary readers. Obviously, access to written English text is limited and literature is not yet an important aspect of primary school ELT. Moreover, reading habits are not cultivated early in children, even in local languages which in fact attract less attention than English in terms of teaching, book provision and publishing output.

The critical problem with book provision is that in Nigeria it is about the only teaching and learning material available to teachers and learners alike; and literature should compensate for lack of access to social mores and other necessary knowledge in society. The limited number of textbooks in the classroom engenders limited work being done by teachers and learners. Additionally, books should compensate for teacher factors, such as observed by Emenanjo (1998) that 'the practising teachers are poorly motivated, ill-trained, overworked, unevenly distributed, and abysmally insufficient in numbers.' Moreover, the learners' environment does not model much proficient English language usage. However, unlike in the past when people learnt the English language from newspapers and the Bible, learners are not equipped to convert favourable contexts (such as television, movies, music, newsprint and other printed materials, etc.) into avenues for learning.

Language examinations: preparedness and malpractices

Sanyaolu states (2008:69):

'Student population soared in the past 20 or 30 years, while the number of schools and the volume of infrastructure for education remained stagnant, sometimes shrinking due to decay. In virtually all public educational institutions, be they primary secondary or tertiary levels, the classes are overcrowded. Basic amenities are lacking and where they exist, have gone out of fashion. Amenities like laboratories and school libraries have become luxuries'.

Considering that many test takers of national examinations are victims of poor teaching, lack of books, unsupportive learning environment, limited language experience, etc. examination papers are bound to be unfair to most examinees who would not have been taught all of the curriculum contents. A fair achievement test assesses learning outcomes, what has been learnt of what was taught (Hagstrom 2006), since 'the resulting scores reflect the amount the test takers have learned' (Banerjee 2001:3). Roseberry-Mckibbin and O'Hanlon (2005) assert that tests should exhibit equity and validity, and be nondiscriminatory. Since national achievement examinations of the English language fail to take cognisance of the learning environment of test takers, assessment may be unfair, inaccurate and unrepresentative of the learners' language skills (Saenz and Huer 2003).

According to McCarty (2008), language is specific to context but exam papers are too artificial and not context sensitive. Additionally, a literate background extends learners' language experience and increases their imaginative and expressive abilities (Dunn 2001, Gonzalez 2006). Language is the means of expressing experience; therefore to test language through literacy tasks outside the socio-linguistic experience of test takers could create cognitive problems for test takers. Socio-cultural contextual factors are important in cognitive development of bilingual learners (Gonzalez 2006). Since 'children learn to be literate in the social contexts of their everyday lives and their linguistic communities' (Dole et al. 2005:24), these factors should be reflected in language teaching and assessment.

It has been observed that summative assessment could have a punitive effect on learners, and where too much emphasis is laid on doing well tests can hamper progress in language learning if performance is unfavourable (Clapham 2001, Hagstrom 2006, Trevisan 2002). The orientation towards certification has led to teaching and learning for the purpose of passing examinations with much attention devoted to the best way to answer questions rather than on language use. There is too much concentration on examination aids rather than on the means of acquiring in-depth knowledge of the workings of the language. The cumulative consequence of years of certificate consciousness in the system is poor examination performance, in spite of increasing use of English in many learners' environment, because effective language use is not being emphasised. A further fall-out of much emphasis placed on certification is the incidence of exam malpractices, which have been increasing in scope and audacity as opportunities for personal advancement available to the individual shrink.

Educational standards fall, exam malpractice increases. As a growing child in the 1970s I observed that society seemed to permit people to sit an examination for a sibling. Today, many rich parents willingly employ the services of 'mercenaries' to sit examinations for their children. Racketeering in question papers is rampant, and schools arranging to obtain papers for their students in advance of an examination is not unheard of. Collins (2008) reports that 2,300 students were screened out of a university admission exercise for having used impostors to sit for an exam. Similarly, one of the May/June 2008 English papers was replaced when the West African Examination Council (WAEC) suspected that it was compromised (Uwadiae 2008).

It should be noted that the teaching and assessment methods create little room for the actualisation of the individual's linguistic capabilities, as the examinations are not geared towards testing communicative skills, and are not versatile enough to cater for individual differences, since some learners tend to perform better in some tasks than others (Clapham 2001). Nwagwu (1997:92) observes that 'the assessment of a student's performance is placed on just one examination either for admission to or for the award of a particular certificate'. The inadequacies of achievement exams for assessing one's communicative competence come to the fore when holders of good results fail to perform up to par with their certificated achievement in proficiency tests during job or admission screening.

Examination results and proficiency in English language

It is sometimes difficult to interpret results obtained in English language examinations in relation to proficiency in the language. According to Ogundipe (2004), the polytechnic students studied could not read at their appropriate educational level. Similarly, the following excerpts from examination scripts indicate that some graduates are struggling writers:

Sample 1

'It is obvious that indigenous communication is facing some problems, therefore retard its growth but I as professional communicator believe there some things that can be done to salvage the problem from extinction.'

Sample 2

'There a lot of problems endangered by African indigenous communication system, these problems needs dedicated and thorough attention for its survival'

Sample 3

'Indigenous communication has a limit in the rural, indigenous communication they are the custodian of moral, moral police in Oriki [epithet]. you will know their linage, the job they do, the work they do, what they and their dislike,'

With some effort the lecturer can make sense of these excerpts; but a piece that reads like any of the above can be difficult for the reader to understand. The question is how a system that takes assessment very seriously awarded any of the writers a credit in an English language test. Although

language assessment is crucial in fair selection of candidates for limited opportunities, ensures minimum competency (Cheng and Curtis 2004), and acts as a gatekeeping mechanism (Badger and Wilkinson 1998), candidates such as the writers of the above excerpts managed to pass through the sieves of language assessment undetected.

Nonetheless, it is observed that after spending some time outside the university environment, graduates returning for a Masters degree programme tend to exhibit less proficiency in the English language than when they were undergoing their undergraduate studies. Perhaps, the socio-linguistic environment has an impact on sustaining proficiency/competence in a language, there being less demand for appropriate use of the English language outside the academic environment. Further, few Nigerians have any reading habits, and books, which can provide a context for continued interaction in the language, are often discarded after the desired certificate is obtained (Okpaku 1975).

Consequences of failing to pass the English examination

Banerjee (2001:3) states that 'an achievement test could be demotivating for the students if it were designed to show up their deficiencies rather than to indicate how successful they had been at absorbing the material they had been taught'. Further, Hagstrom (2006) observes that summative assessment does not encourage self-assessment and learners' ability to set learning goals. Therefore, recent approaches to language teaching incorporate self-assessment and peer assessment in formative language assessment, thereby creating more learner involvement in learning, and awareness of an individual's learning strategies and progress being made. An advantage of summative assessment, however, is that most test takers want their scores to prove their proficiency in the language and therefore their suitability for a given position or opportunity. Thus, it is frustrating for one to attempt an exam several times without succeeding. Failure to pass an English language exam after several attempts eventually kills many test takers' ambitions, as in Zobia's case described below.

Zobia (not his real name) is the second born of six children and the only person yet to receive a university degree. With a keen interest in politics and contacts with renowned politicians, all he needs is a degree to enable him to fit into the political scene. However, he is yet to enrol for a degree programme because after several successful attempts of the university qualifying exam, O level English remains Zobia's Achilles heel. Zobia is an avid reader and performs well in all other subjects, achieving As in several history exams that he took while retaking the English language paper – an indication that he can communicate in the English language. It is frustrating that even though he communicates better than many postgraduate students and will definitely not produce any of the samples given above, he still cannot fathom why the English language paper has stood between him and his life ambition. Zobia has lost count of the number of times he sat for the O level English examination between 1989 and 2005. With each attempt

he felt that he had done better and should succeed; but he is able to identify the following as his problems:

• The practice essays he wrote before exams indicated remarkable improvement; he however has poor handwriting.

• His sentences are too long and some are incorrect.

• Generally, he has no problems with objective questions, except with providing words or group of words synonymous with given words.

Expectedly, Zobia thinks that the English paper is unfair to test takers, and with good reason. He believes that teachers transfer their limitations in the language to their students, as observed by Taiwo (1980). Furthermore, he views English language as a technical subject that should not be taught by just anybody that studied English (Emenanjo 1998). He suggests, therefore, that examining bodies should assess the examination process and the purpose language assessment serves, because many students frustrated by the Nigerian system succeed elsewhere. This suggests that some deficiencies exist in the Nigerian approach to language teaching and assessment.

According to Zobia, the English language paper has defined people's destiny, as it is the single subject that can stop one from pursuing tertiary education and getting good employment. Unlike many others, however, he will not do just about anything to get past the English examination hurdle by indulging in examination malpractices. For Zobia, whose future is still hanging in the balance, O level English exam has not revealed how successful he has been in learning the language (Banerjee 2001). Young and Fletcher (2000) opine that as a social practice, multiple measures should be devised for assessing literacy, to capture the range of behaviours exhibited in different contexts. This, in addition to improved ELT, could stem the incidence of miscreants roaming the streets daily with no future ambition.

Conclusion: ways forward

Language learning should correspond with an individual's wholesome development. Thus: 'over the past decade, standardised tests of language knowledge have given way to more communicative forms of individualised assessment that reflect the context in which learning is taking place' (Brindley 2001: 464). Adopting this approach would create more effective language users and perhaps better results in national language tests. Nigeria is, however, slow in responding to changes in language teaching and testing. Perhaps, Nigeria fails to re-examine its language assessment strategies and tools because of this type of scenario:

> 'Nnamdi Azikiwe University (UNISIK), Awka, Anambra State, has concluded its 2008 admission processes for fresh students, leaving out 29,000 of 33,000 applicants that qualified through the Joint Admission Board (JAMB) exams.' (Collins 2008:1)

Limited opportunities, notwithstanding, the English language should be taught and assessed to enable learners become effective language users.

What is needed is an expanded socio-economic and socio-cultural environment that creates avenues for self-actualisation. The infrastructures that encourage development will create room for people to diversify and seek personal fulfilment in language learning. Success in the classroom may depend on an environment characterised by fair, accurate and representative assessment of the learners' skills (Saenz and Huer 2003). In light of this, teachers and language testers require more training in purposeful language assessment (Trevisan 2000) aligned with the nation's language policies. Moreover, achievement tests should also be used to evaluate teaching curriculum and methods to bring about pertinent changes to the whole system (Banerjee 2001). Cognisance should also be taken of urban societies which are increasingly becoming barren of native knowledge and intelligence distilled from folktales and traditional cultural practices which are avenues for experiencing language in context, as renowned African writers did in their time. In developed countries literature and visits to museums and places relevant to life in society help to expand contexts for language use. In addition, language teaching and assessment cannot ignore English for international/intercultural communication in a globalised world.

References and further reading

Badger, L and Wilkinson, L (1998) Literacy assessment of students from poor and diverse communities: Changing the programs, changing the outcomes, *Australian Journal of Language and Literacy* 21/2, 147–158.

Banerjee, J (2001) Achievement tests, in Byram, M (Ed.), 3.

Braid, F R A (2007) Philosophy of Education for the Country, *Manila Bulletin*, 10 January 2007, retrieved from www.mb.com.ph/issues/2007/01/10/OPED2007011084219.html

Brown, P J (2008) Nigeria's Chinese-built satellite goes dark, *Asian Times Online*, 18 November, retrieved from www.atimes.com/atimes/China/JK18Ad01.html

Byram, M (Ed.) *Routledge Encyclopedia of Language Teaching and Learning*, London: Routledge.

Cheng, L and Curtis, A (2004) Washback or backwash: A review of the impact of testing on teaching and learning, in Cheng, L, Watanabe, Y and Curtis, A (Eds) *Washback in Language Testing: Research Contexts and Methods*, Mahwah, NJ: Lawrence Erlbaum Associates, 3–17.

Christopher, N M (2007) *Literacy and book development in Africa: Any relationship?* paper presented at the 5th Pan-African Reading for All Conference, Legon, 6–10 August.

Clapham, C (2001) Assessment and testing, in Byram, M, 48–53.

Collins, C (2008) Varsity rejects 29,000 students, *The Guardian*, 7 September, 1.

Dole, S, Yelland, N, Latham, G, Fehring, H, Wilks, A, Faulkner, J, Lee, L, Masters, J, Lowrie, T and Glynn, T (2005) *Project to Investigate Improving Literacy and Numeracy Outcomes of Distance Education Students in the Early Years of Schooling*, Australia: Australian Government.

Dunn, M (2001) Aboriginal literacy: Reading the tracks, *The Reading Teacher*, 54/7, 678–686.

Emenanjo, E N (1998) Languages and the national policy on education: implications and prospects, *Fafunwa Foundation Internet Journal of Education*, retrieved from http://fafunwafoundation.tripod.com/fafunwafoundation/id1.html

Gipps, V (1994) *Beyond testing: towards a theory of educational assessment*, London: Falmer Press.

Gonzalez, V (2001) The Role of Socioeconomic and Sociocultural Factors in Language Minority Children's Development: An Ecological Research View, *Bilingual Research Journal*, 25/1 and 2, 1–30.

— (2006) Profiles of Cognitive Developmental Performance in Gifted Children: Effect of Bilingualism, Monolingualism, and Socioeconomic Status Factors, *Journal of Hispanic Higher Education*, 5/2, 142–170.

Hagstrom, F (2006) Formative learning and assessment, *Communication Disorders Quarterly*, 28/1, 24–36.

McCarty, S (2008) The bilingual perspective versus the street lamp syndrome, *IATEFL Voices*, 203, 3–4.

NERDC (1988) *Reducing Cost of Textbooks in Nigeria*, Nigeria: NERDC.

Nuhu-Koko, A A (2008) Nigerian Communications Satellite-1 (NigComSat-1): Missing in action (MIA) (I), *Daily Triumph*, retrieved from www.triumphnewspapers.com/niger20112008.html

Nwagwu, C C (1997) The Environment of Crises in the Nigerian Education System, *Comparative Education*, 33/1, 87–96.

Ogundipe, A O (2004) Some reading difficulties of users of English for academic purposes in Nigerian polytechnics: The case of former Calabar Polytechnic, *Literacy and Reading in Nigeria*, 10/1, 223–227.

Okpaku, J O O (1975) Publishing and public enlightenment in Africa, in Oluwasanmi, E, McLean, E and Zell, H (Eds) *Publishing in Africa in the seventies: proceedings of an International Conference on Publishing and Book Development held at the University of Ife, Ile-Ife, Nigeria, 16–20 Dec 1973*, Ile-Ife: University of IFE Press, 233–250.

Olugbile S (2008) WAEC releases results, 83% of candidates fail, *Punch*, retrieved from www.punchng.com/Articl.aspx?theartic=Art2008092714561089

Osuji, F (2004) *The education system at the beginning of the twenty first century: An overview and assessment*, paper presented at 47th session of the International Conference on Education, Geneva, 8–11 September.

Roseberry-McKibbin, C and O'Hanlon, L (2005) Nonbiased assessment of English language learners: A tutorial, *Communication Disorders Quarterly*, 26/3, 178–185.

Saenz, T I and Huer, M B (2003) Testing strategies involving least biased language assessment of bilingual children, *Communication Disorders Quarterly*, 24/4, 184–193.

Sanyaolu, K (19 October 2008) The Collapse of Public Education, *The Guardian*, 69.

Shafer, G (2001) Religion, politics and the fear of whole language education, *The Humanist*, 61/5, 5–8.

Taiwo, C O (1980) *The Nigerian education system: Past, present, and future*, Lagos: Thomas Nelson (Nigeria) Limited.

Trevisan, M S (2002) The states' role in ensuring assessment competence, *Phi Delta Kappan*, 83/10, 766–771.

Uwadiae I (25 September 2008) Press release, retrieved from WAEC: www.waecnigeria.org/newsrelease.htm

Uwalaka, N M E (2000) *Book publishing performance in the Nigerian economic environment*, unpublished PhD dissertation, University of Ibadan.

Young, J and Fletcher, M (2000) A Queensland Perspective on Assessment in the Early Years: Teacher Perceptions and Use of the Reading Developmental Continuum, *Australian Journal of Language and Literacy*, 23/3, 212–229.

State Certificates in Polish as a Foreign Language: impact and stakeholders

WALDEMAR MARTYNIUK JAGIELLONIAN UNIVERSITY, KRAKOW, POLAND

Introduction

In the last couple of years a growing interest in learning Polish as a foreign language can be observed. An estimated 10,000 learners study Polish across Europe, about one third of them at universities and language schools in Poland. State certificates in Polish as a foreign language have been offered to a general adult audience for almost five years now. In January 2008 a survey was carried out to collect data on the use of the certificates by the different stakeholders and the impact that the introduction of the certification scheme may have on their activities. The aim of this article is to briefly present the Polish language and the certification system and summarise the results of the survey.

The Polish language

Polish belongs to the West-Slavic group of the Indo-European family of languages. Its structure represents a synthetic, highly inflected system. As a language of its own, Polish evolved in the 10th century playing an important role for the development of the Polish state after its Christianisation in the year 966. The oldest single words in Polish have been found in 12th century Latin texts. Until the end of the 14th century Polish was used mostly only in several spoken variations. As a literary, over regional language it developed first in the 15th and 16th centuries as it started to be used by a growing group of writers and scholars, who were able to create renaissance literature of a considerable quality in Polish. During the early stages of its development the Polish language was highly influenced by the neighbouring languages – German and Czech – and by Latin. Later on, in the 18th century, Polish came under quite strong influence from French. Nowadays the English language serves as the main source of borrowed words and phrases. The largest dictionaries of Polish contain up to 130,000 items whilst the number of words used in everyday communication is approximately 20,000.

The estimated number of people who use Polish as their language is well above 40 million, ninety percent of whom live within the current borders of Poland. Large groups of ethnic Poles and their descendants live outside the home country, mainly in the USA, in Canada, Australia, Brazil, Great Britain, France, and Germany, in Belarus, Ukraine, Lithuania and in Kazakhstan.

Learners of Polish have to face the fact that, for example, nouns in Polish can each have up to 14 different declension forms. They may be confused by the pronunciation of utterances like "*W Szczebrzeszynie chrząszcz brzmi w trzcinie*". To their surprise and amusement however, they find words in Polish that sound very familiar to them. German students immediately recognise words like *dach*, *blacha*, *jarmark*, *gmach* or *ratusz*. The Italians enjoy such Polish words as *fontanna*, *gracja*, *pomidor* and *parapet*. The French can immediately start using words of their own, like *bagaż*, *bilet*, *bukiet*, *romans* or *wizyta*. Speakers of English smile to words like *trener*, *spiker*, *komputer*, *biznesmen*, *folder*, *mikser*, *relaks*, *keczup*, *drybling* or *dżinsy*. All of them can easily understand front-page headlines in Polish newspapers, like: "*Racjonalna polityka ekonomiczna*", "*Katastrofalna inflacja – bank centralny interweniuje*", "*Konferencja prasowa ministra finansów*", "*Korupcja koalicji*", "*Skandal w USA – prezydent i praktykantka*", "*Strajk personelu paraliżuje Heathrow*" – all of these being genuine contemporary Polish!

State certificates in Polish as a foreign language

The system of state certificate examinations in Polish is relatively new and may be considered to be still evolving. There are three levels: Poziom podstawowy, Poziom średni ogólny, and Poziom zaawansowany. All three examinations have been developed along the same lines by a working group supervised by a team of experts representing seven Polish public universities experienced in teaching Polish as a foreign language. The expert team – *Komisja ds. Certyfikacji Znajomości Języka Polskiego jako Obcego* – was nominated by the Minister of Education in 1999 to conduct a feasibility study and perform development and pilot activities aiming at establishing a system for examining and certifying proficiency in Polish as a foreign language at the state level, thus fulfilling the obligations enacted in the Law on Polish Language adopted by the Polish Parliament in 1999. The expert team appointed a group of university practitioners in the field to elaborate standards and develop and trial pilot tests on three levels of proficiency – Threshold, First Certificate and Proficiency – following the system of levels and examinations already used at the Jagiellonian University in Kraków. The Working Group decided to base the level descriptions and the test specifications on such sources as: the Jagiellonian University system of levels and examinations; the draft Threshold Level description for Polish as a foreign language developed at the Jagiellonian University in 1998 in consultation with John Trim as a representative of the Council of Europe, and the set of global scales of language proficiency and the scales of illustrative descriptors for the levels B1, B2 and C2 included in the draft version of the

Common European Framework of Reference for Languages published by the Council of Europe in 1996. After an extended pilot phase of over three years during which both level descriptions and test specifications were broadly consulted in the field and several rounds of trial tests took place in and outside of Poland, the system was proclaimed to be ready for implementation. In November 2003 the work of the expert group was concluded and the Minister of Education nominated members for a State Commission for the Certification of Proficiency in Polish as a Foreign Language and issued regulations with regard to level descriptions and test specifications, as well as test development and administration. The first real examinations on all three levels were administered in June 2004 at three sites in Poland. The Commission decided to be flexible in scheduling the examinations, offering them at least three times a year (spring, summer, autumn) at the main university centres in Poland and 'on demand', when there is a reasonable number of candidates in schools, universities and language centres teaching Polish as a foreign language abroad. Until September 2008, a total of 1376 candidates from 60 countries have been examined at 19 different sites within and outside Poland where the State Examination Commissions were sent to administer the tests. So far 1127 certificates have been issued which means that 82% of the candidates have passed the examinations successfully. Currently, examinations of the same format for levels A2 and C1 are being developed for piloting. As the system develops and becomes more widely used and established, quality assurance is an area of special concern to all those involved. The first quality analysis was undertaken in 2003, using the ALTE Quality Assurance Checklists. Some areas of concern were identified and a range of steps to deal with them proposed (Martyniuk 2004). In Spring 2007 the ALTE Checklists were used again, this time in preparation for an audit carried out within the ALTE membership aiming at establishing a quality profile of the Polish examinations. The outcomes of the audit (concluded at the end of 2008) will serve as a useful basis for a thorough review of the system and a practical reference for any revisions to be undertaken.

Impact and stakeholders

An analysis of the candidature released recently by the State Certification Commission (2008) shows that the examinations are most popular with candidates of Polish origin who keep their Polish citizenship while living permanently outside of Poland. This group is closely followed by test takers from Ukraine, Germany, and the USA. Candidates of 20–29 years of age dominate the field (48.5%) while there is also a strong representation of young test takers of 16–19 years of age (25.5%). Female candidates outnumber their male counterparts by 9 to 5. The choices of examination level seem to be well balanced, with the middle one – *Poziom średni ogólny* – being most popular and the highest one – *Poziom zaawansowany* – the least frequently attempted. Analysing the registration forms, Szczęsna (2006) has found the following kinds of motivations for taking an exam as declared by the candidates:

General

- Need for self-assessment
- Need for 'a regular language check'
- For 'examination experience'
- 'One more asset in the CV'.

Specific

- For employment in Poland
- To negotiate a raise in salary
- To fulfil language requirements for study purposes
- To be admitted to a study programme in Poland
- For a reduced study programme fee in Poland.

Looking for language requirements that could be satisfied by the system, Szczęsna (2006) identified the following groups of potential stakeholders:

- Members of Supervising Boards (banking and insurance companies operating in Poland)
- Foreigners working for real estate companies operating in Poland (C2)
- Foreigners seeking employment within the Polish health sector (level depending on position)
- Foreigners seeking admission to Polish public universities (levels B1–B2).

She also noted that some employers require foreigners to produce a certificate in Polish when signing contracts and issuing documents for work and residence permits.

As was expected, the introduction of the state certification system triggered changes in the way Polish is taught as a foreign language. Mazur (2006) lists the following reactions observed in the field:

- Modification of curricula: more focus on development of certain skills (reading comprehension and writing);
- Development of new curricula, specifically in Polish schools in the USA and 'in the East' (former Soviet republics);
- Development of new textbooks and teaching materials, including some online;
- Modification of teacher education and in-service training programmes;
- Request for a new strategy for Polish schools abroad: towards a skills-based approach;
- Request for a language certificate at C2 level as a requirement to teach at Polish schools abroad.

The experiences with the new assessment scheme for Polish as a foreign language also contributed to the debate on teaching Polish as the mother tongue and as the language of school education in Poland. For example, Miodunka and Przechodzka (2007) let a group of final-year upper-secondary native speakers take the written part of the *Poziom zaawansowany* examination aiming at the C2 level. They found that the native speakers had quite similar difficulties with the test tasks and scored only slightly higher compared to the group of foreign language users who took the same test. These results suggest that the skill-oriented

proficiency in Polish as a mother tongue should not be taken for granted but requires constant attention and specific 'treatment' in schools – possibly training comparable to that offered on courses in Polish as a foreign language in preparation for the state certificate examinations.

The survey

In January 2008, a short questionnaire was emailed to all previous test takers in order to find out about their reason for taking the exams and how useful the certificates turned out to be in their personal and professional life. Sixty four respondents from 17 countries sent in their responses. The respondents' motivations for taking exams show the following hierarchy: work – study – personal reasons – formal confirmation of skills. A clear majority of those who responded (61%) confirmed that they had had an opportunity to produce their certificate and benefit from it. The use of the certificates turned out to be most beneficial in the following contexts:

- At work
- For admission to a Polish university
- For benefits related to a study programme
- Important qualification in the CV when seeking employment
- Additional qualification in the dossier of a freelancer
- To impress others with certified ability in 'a difficult language'.

The respondents estimated the 'value' of their Polish certificates as follows: high value (53%), medium value (39%), low (5%) and no answer (3%).

Conclusion

During the first five years of operation, the system of state certificates in Polish as a foreign language has produced quite a significant impact among its stakeholders. Influenced by the skill-oriented assessment approach used for the purpose of certification, the teaching of Polish as a foreign language has become more proficiency-oriented and focused on skills that were largely 'under-developed' – like writing and reading comprehension. Although the certificates are not being issued in order to fulfil any specific formal language requirements, they have already been noticed as a useful benchmark and proof of qualifications both in the public and in the private sectors. Test takers themselves estimate the 'market' value of their Polish certificates quite highly and find them useful. With a growing interest in the Polish language in Europe (the sixth largest language in the EU) and a steadily growing number of candidates on the one hand, and with the significant efforts for proper quality management on behalf of the State Certification Commission on the other, the future of the Polish certificates seems to be promising.

References and further reading

Martyniuk, W (2004) ALTE Quality Management System and the State Certificates in Polish as a Foreign Language, *ALTE News Spring–Summer 2004*, 1–2.

Mazur, J (2006) Wpływ certyfikacji języka polskiego na proces zmian w polskiej polityce edukacyjnej, in Tambor, J and Rytel-Kuc, D (Eds) *Polityka językowa a certyfikacja*, Katowice-Leipzig: Wydawnictwo Gnome, 132–135.

Miodunka, W, Przechodzka, G (2007) Mother tongue versus foreign language performance, in Martyniuk, W (Ed.) *Towards a common European framework of reference for languages of school education? Proceedings of a conference*, Kraków: Universitas, 301–308.

State Certification Commission (2008) www.buwiwm.edu.pl/certyfikacja

Szczęsna, D (2006) Kto? Co? Jak? Dlaczego ludzie zdają egzaminy z języka polskiego jako obcego?, in Tambor, J, Rytel-Kuc, D (Eds) *Polityka językowa a certyfikacja*, Katowice-Leipzig: Wydawnictwo Gnome, 86–94.

Washback effect of tests in Irish for adult learners

ANNE GALLAGHER LANGUAGE CENTRE, NUI MAYNOOTH

SIUÁN NÍ MHAONAIGH LANGUAGE CENTRE, NUI MAYNOOTH

Introduction

In a world in which the English language dominates as a lingua franca in economic, scientific and cultural life, few languages can face the future with complacency. Minority languages are in an even more precarious position, their very existence being threatened on an almost daily basis (see Crystal 2001, McCloskey 2001). In most European countries, the monolingual minority language speaker belongs to the past. Communication is therefore always possible in the dominant language. This presents additional challenges for minority language learners, who realise that they can always resort to the dominant language in order to get the message across.

This article will examine the educational and social impact of tests for adults in one such minority language: Irish. Test takers are from a variety of linguistic backgrounds: L1 English; L1 Indo-European or non-Indo-European; learners with up to 12 years' study of Irish behind them; native speakers wishing to improve their knowledge of the language, particularly in its written form.

In the Series Editors' note to Anthony Green's book, *IELTS washback in context: preparation for academic writing in higher education* (2007) Cyril Weir and Mike Milanovic, in reference to a number of validation projects (Burrows 1998, Cheng 2005, Wall 2005) state:

> 'It is clear that the interpretation and uses made of assessment procedures are not solely determined by testers, but depend on interactions involving the test and participants (such as test takers, teachers, administrators, materials developers and policy makers) with implications for teaching and educational policy'.

This article will demonstrate that a project which initially involved only writers of specifications and test developers gradually grew to include all of the parties above, whose relationship eventually became one of interdependency rather than one of simple interaction.

Background: the Irish language policy

Linguist Joshua Fishman (1991) lists state support as one of the desirable conditions for the survival of a minority language in modern times. From the foundation of the Irish State, efforts were made to enshrine in law the special position of the Irish language in Ireland. To quote historian Gearóid Ó Tuathaigh (2008:28):

> 'the main elements (of the policy adopted) were, the maintenance of the Irish-speaking community of the *Gaeltacht* [Irish-speaking regions]; the promotion/revival of Irish in the overwhelmingly English-speaking country at large, through the education system; ensuring basic competence in Irish from those working in the public service, and standardising and modernising the language itself.'

The Constitution of the Republic of Ireland states that 'Irish as the national language is the first official language'.[1] In recent times and due in no small part to the determination of Minister for Community, Rural and Gaeltacht Affairs, Éamon Ó Cuív, this legal position has been bolstered and even re-dynamised by several additions to official government policy. The Official Languages Act (2003) was introduced to give every citizen of the state the legal right to conduct their business with all public bodies through either of the official languages. In 2006, the Irish Government issued what it termed a 'statement on the Irish language' committing it to the development of a 20-year strategy for the support and promotion of the language. Its objectives include:

- Full implementation of the Official Languages Act and facilitation of the public's right to use Irish in dealings with the State.

- Provision of a wide range of services to parents who wish to raise their children through Irish.

- Continued development of high-quality broadcast services through Irish, particularly on TG4, RTÉ and Raidió na Gaeltachta.

- Special support for the Gaeltacht as an Irish-speaking area

Also, on 1 January 2007, Irish became an official and working language of the European Union.

On the ground, these changes in status appear to have had a positive effect. The Official Languages Act 2003 has prompted many public servants to return to Irish. Some would say that it has resulted in the development of an

---

1. Bunreacht na hÉireann/Constitution of Ireland

Irish-language industry, providing much additional work for interpreters and translators, as members of the public began to exercise their right to have certain meetings held in Irish, and public bodies began publishing annual reports in bilingual form in order to comply with the provisions of the Act.

Education

As Ó Tuathaigh (2008:29) puts it, 'In terms of specific government policies, the state's revivalist commitment was most aggressive in the education system'. Irish is a compulsory subject throughout primary and secondary education, with exemptions being granted in a minority of cases. Irish, as well as English, is also an entry requirement for four of the seven Irish universities, those which are constituent universities of the National University of Ireland. It is clear, therefore, that the impact of Irish on the lives of children is considerable and that the position of the language is underpinned by a robust legal framework. But does this translate into adult speakers of the language and what is the status of the language in the hearts and minds of the Irish people?

Irish and cultural identity

According to the 2006 census, 1,656,790 individuals claimed to be able to speak Irish. Of these, only 72,148 stated that they used the language on a daily basis outside the education system; 22,515 were residents of the Gaeltacht (Irish-speaking region) (Punch, 2008:52). The census, perhaps wisely, does not seek to ascertain how many speakers are native speakers of Irish. It is, of course, notoriously difficult to come up with a definition of the native speaker, see Davies (2003:237). Nevertheless, it is interesting to note that in response to a Eurobarometer survey in 2000, 40% of Irish respondents declared that Irish was their 'mother tongue'. It is not unusual to hear Irish people talk of learning 'their own language'. Indeed, the opening paragraph of Irish sociologist Iarfhlaith Watson's essay on 'The Irish Language and Identity' begins with the declaration that 'In 2007 (Irish-American) comedian Des Bishop took the ultimate step towards Irishness by moving to the Gaeltacht *to learn his mother tongue*' (Watson 2008). All of this is an indication that the efforts of the founding members of the state managed to convince the generations that succeeded them that the Irish language was, to quote Ó Tuathaigh (2008:28), 'the most irrefutable mark of a distinctive Irish "nation"'. There is also some anecdotal evidence to suggest that a more multilingual Ireland, which is now home to some 200 languages, has played its part in strengthening this emotional link to the language, as has increased student mobility as a result of European exchange programmes (Gallagher 2006).

Indeed, the role of the Irish language in national identity has not been lost on Ireland's newest inhabitants. It is estimated that newcomers (491,500) to Ireland now make up 12% of the population (Quarterly National Householder Survey, 2008). Some of these are simply migrant workers hoping to return to their country of origin in a few short years. Others, however, have decided to settle in Ireland and wish to participate fully in Irish society. For them, learning Irish opens another door to integration and is a way of demonstrating to the native Irish their desire to become fully-fledged Irish citizens. In 2005, two newcomers to Ireland, Dutch Irish-language journalist Alex Hijmans and Ariel Killick, an Australian translator, founded an association called *iMeasc* [among], one of whose chief aims is 'to encourage immigrants to learn Irish and interact with that aspect of Irish culture as a means of better integrating themselves, their children and their own cultural identity in society' (Ó Muirí 2005). In particular, those who settled in the strong *Gaeltacht* areas were at a social disadvantage where Irish was the principal language of communication.

Irish and the media

1996 saw the opening of TG4, Ireland's first Irish-language television station. Several community radio stations have opened, in addition to the already well established national Irish-language station, Raidió na Gaeltachta. According to Irish-language journalist, Breandán Delap, 'there are over three hundred people employed in media-related projects in the South-Conamara Gaeltacht', where the headquarters of TG4 are located (Delap 2008:152). It is also fair to say that the Irish-language entertainment industry punches well above its weight nationally and it has received many awards at reputed international festivals. The most recent illustration of this is the 2007 Irish Film and Television Awards, at which Irish-language or bilingual Irish-English productions received 42 nominations for either film or television productions, and won eight categories.

Practical implications

The additional demand for Irish language proficiency in the public service, the continuing strong role of Irish in the education system, the enhanced status of the language at EU and national level, and as a fashionable media language at home have conspired to create a demand for language classes for adults who wish to go beyond the by now clichéd, symbolic *cúpla focal* ('few words'). There are now those who are choosing to learn and live in Irish. However, until 2005, although there were a small number of individual groups involved in providing good quality courses in Irish for adult learners outside the formal education system, most classes were provided on a fairly casual basis, without a syllabus, and without any real idea of levels of achievement. National accreditation for these courses did not exist.

The European Certificate of Irish

In an attempt to address some of the gaps in Irish-language provision for adults, the Language Centre at the National University of Ireland Maynooth sought funding for a comprehensive package of initiatives, including the design of a syllabus linked to the six levels of the Common European Framework of Reference for Languages: learning, teaching, assessment (CEFR), teaching materials, teacher-training materials and a self-study pack. The request for

funding was unsuccessful. Convinced of the value of such a project, however, the university agreed to make available a small sum of seed money.

Anthony Green, in the opening sentence of his excellent book, *IELTS washback in context: preparation for academic writing in higher education*, (2007), articulates the principal objective of the work undertaken at Maynooth: to have a positive effect on teaching and learning Irish, specifically in the case of adult learners. The definitions of 'washback' are many and varied. For the purposes of this article, we will concentrate on a number of definitions discussed by Green and on which a significant number of experts appear to be in agreement. In order to ensure that the highest standard of work was produced, NUI Maynooth became an observer member of the Association of Language Testers in Europe (ALTE) in January 2005, and a full member on presentation of our first tests in January 2006. ALTE membership has been vital in guiding our work, by allowing us regular access to language-testing experts, and through the support received from other members, who have been more than generous in sharing their experience and expertise.

Teastas Eorpach na Gaeilge [European Certificate of Irish] was launched by Mary Hanafin, T.D., Minister for Education and Science, in March 2005. In the first year, specifications for A1 and A2 (Bonnleibhéal 1 and Bonnleibhéal 2) were made available. To date, syllabuses are provided for A1, A2, B1, B2 and C1.[2] The final level, C2, is currently in development and will be available in 2009.

The first examinations at A1 and A2 were held in May 2005. Since specifications had only been available for two months, just 22 candidates presented for the tests. In 2006, there were 96 candidates; in 2007, 304; and in 2008, 430. In the first year, examinations were held only at NUI Maynooth. However, in response to demand, the following year centres were established in each of the three major Gaeltachtaí [Irish-speaking regions]. In 2007, students at Charles University, Prague were examined for the first time. In 2008, examinations were held at the *Centre Culturel Irlandais* in Paris. New York University will host the examinations for the first time in 2009, and discussions are currently underway with the University of Cambridge, following a request to run the TEG examinations there in 2009. Requests have been received from many more organisations and institutions in mainland Europe and North America, but while it is our policy to cater for as many candidates as possible, development cannot compromise the integrity of our examinations in any way and will therefore be necessarily slow.

Educational impact

Learners

In taking the decision to develop a suite of tests to improve the teaching and learning of Irish to adults, the Maynooth team was keenly aware of the role played by good testing in motivation. Anecdotal evidence would suggest that this objective has largely been achieved. The possibility of taking a nationally recognised examination based on best

2. See www.teg.ie

international practice has drawn many learners. One example of this is the case mentioned by Seán Ó Daimhín, Irish Language Officer with Donegal County Council, in an article in the Irish Times in September 2007: 'Sixty members of staff attended the courses last year. There was a further 25 we couldn't cater for, but it looks like we'll have 80 at the very least in the classes this year' (Faller 2007).

This information may well tell us something of the motivation engendered by the very availability of a suite of tests, a vital starting point for any learning to take place, but it has not necessarily given us any information on the quality of that learning. To quote Green (2007), 'washback is often evaluated as positive or negative according to how far it encourages or discourages forms of teaching or learning judged to be appropriate'. Comments from the Irish Times article mentioned above appear to confirm that the TEG system has 'encouraged appropriate teaching and learning':

> The exam itself let me know what my weak parts were, and it was good to pass it as well.

> I noticed the progress every week. The syllabus is geared to bring you along.

These two short comments by learners from County Galway and County Clare respectively would seem to indicate that the syllabus [and teaching] has encouraged reflective, incremental learning, motivation, and a sense of achievement. To quote Green (2007:25): 'The greater the congruence [between test tasks and the skills required] the more likely positive washback becomes'.

Materials

Given that the TEG tests were the first of their kind, teachers were understandably anxious about putting their students forward for examinations. Although the syllabuses for each level were detailed enough to enable teachers to adapt them for use in the classroom, many teachers were reluctant to do this and it soon became evident to us that that the absence of course materials was going to impact negatively on take-up. We therefore set about developing teaching materials for A1 and A2. Materials for B1 were completed in September 2008, and we are currently seeking funding to enable us to begin work on B2 level. All materials are available free of charge on a dedicated bilingual website (www.teg.ie), which also contains information on the tests for all of the stakeholders.

Teachers

Once materials for the lower levels were completed, there followed a substantial number of requests from teachers and local education authorities of various kinds for training in the use of our teaching materials. A number of one-day training sessions were held at locations throughout the country. However, feedback from tutors pointed to the need for a dedicated teacher-training course in the teaching of Irish to adults. The *Teastas i Múineadh na Gaeilge d'Aosaigh* [Certificate in the Teaching of Irish to Adults], a one-year part-time NUI Certificate course was introduced in 2007. The second cohort of students began this course in the Munster *Gaeltacht* in September 2008.

Administrators

The Language Centre at Maynooth has also received requests from various teaching institutions to assist them in introducing TEG courses on their teaching programmes. One such institution at second level is Coláiste Bríde, an Irish-medium school which will teach B1 as part of its Transition Year programme from September 2008 (an optional year between the junior and senior cycles devoted to personal development and extra-curricular activities). At third-level, the Letterkenny Institute of Technology will include optional modules of the TEG in several of its degree programmes beginning 2008–2009.

Policy makers

In 2006, *Gaeilge 2010* was launched. This was a joint initiative between *Údarás na Gaeltachta* (the state agency responsible for the promotion of economic development in the *Gaeltacht*) and the Department of Community, Rural and Gaeltacht Affairs, whose chief aim was to stimulate 'the participation of local Gaeltacht communities in planning the conservation and extension of Irish as a community language within their own areas' (Údarás na Gaeltachta 2008). The first phase of the project saw the development of an initial 30 Irish language learning centres at which TEG courses are taught. The Údarás na Gaeltachta website (2008) includes the following information with regard to its future language policy:

> 'Sixty teachers have completed specialised in-service training on a regional basis in the teaching of Irish to the European Irish Language Certificate (TEG) level developed by the Languages Centre of NUI Maynooth. They will work through the language learning centres that are being established throughout the Gaeltacht. There are now 28 centres operational and it is expected that this will increase to 40 as additional locations are identified over the coming year'.

In March 2008, the TEG team was asked to give a presentation on our work to the National Council for Curriculum as Assessment, which is currently conducting a review of the Irish-language state examinations. As a result of our work on TEG, the State Examinations Commission has invited the Maynooth team to join them in a project which aims to relate language levels achieved at Irish state examinations to the CEFR. From 2008–2009, the Language Centre will train Fulbright Foreign Language Teaching Assistants who are preparing to spend a year teaching Irish at universities in the United States in the use of TEG teaching materials. In addition, a dedicated website offering support, additional materials, information and a discussion forum will be provided by the TEG team at Maynooth.

A new pathway

Green states that 'washback is grounded in the relationship between preparation for success on a test and preparation for success beyond the test' (Green 2007:2). A further positive consequence of the TEG system was the development of a new educational pathway for adult learners of Irish. Until the introduction of TEG, it was not possible for newcomers to Irish to access third-level Irish courses. Learners who pass the B1 examination (Meánleibhéal 1) may now join the first year of Dioplóma sa Ghaeilge at NUI Maynooth, an NUI, two-year, part-time diploma, which is recognised by the university and by teacher training colleges as equivalent to first year of the B.A. in Irish. Students may then opt to continue with a degree in Irish and, if eligible, postgraduate studies in the language.

Conclusion

The TEG examination and certification programme has been in existence for just over three years. The extent of its impact has exceeded our most ambitious expectations. Courses are now taught in locations from the most remote townlands in Ireland to cities such as Sydney, New York and Paris. Our website is accessed by an average of about 2000 people each month. In many ways, candidates who take our tests are only the tip of the learner iceberg. For every TEG candidate there are probably about 10 other learners who either do not feel ready to take the test or who live too far from our examination centres. What is more certain is that the introduction of a structured, coherent syllabus leading to high quality tests has had a very positive impact on the adult learner of Irish.

While one should never be complacent and many considerable challenges lie ahead, the team at Maynooth is satisfied that the TEG project has had considerable positive impact on the teaching of Irish to adults and on their learning experience. It has served to create an awareness of the CEFR among teachers, learners, administrators and policy makers. It has filled a gap in Irish-language provision both in Ireland and abroad and, perhaps most importantly, provided access to a language, in some cases a second chance, for hundreds of learners whose needs had been hitherto largely ignored. Below is an excerpt from an interview given (in Irish) to Raidió na Gaeltachta (2008) by one such learner, Clive Geraghty, who, having taken the TEG at Bonnleibhéal 1 (A1) in 2005 and A2 and B1 subsequently, will enter university in September 2008 to begin a BA degree in Modern and Old Irish:

> CG: I hadn't done much Irish for 30 years. But I was able to do the exam and the result gave me the confidence to continue with Irish. I'm now doing a Diploma in Irish in Maynooth and I hope to do a degree in the future.
>
> TS: You were an actor with a fine CV. Why did you want a certificate?
>
> CG: I retired from the Abbey Theatre two years ago. I was 65 years of age and I had the choice of leading a very comfortable life without doing much other than playing golf and reading the Irish Times in the morning. I didn't want that.
>
> I didn't have the opportunity to go to school when I was young. I only attended secondary school for one year. I joined the army, the Air Corps, when I was 17.
>
> I always wanted to improve my education. I didn't have the opportunity because I had a family and I had to work. But that pressure no longer exists, and I am going to seize the opportunity.

The challenges which lie ahead for the TEG are not insignificant: the need to expand the system while

maintaining standards; the development of a placement test; the introduction of tests in Languages for Specific Purposes. However, we take courage from the feedback our work has received thus far.

References and further reading

Adhmhaidin (2008) *Raidió na Gaeltachta*, 3 March.

Crystal, D (2001) *Language Death*, Cambridge: Cambridge University Press.

Davies, A (2003) (2 ed) *The Native Speaker: Myth and Reality*, Clevedon: Multilingual Matters.

Delap, B (2008) Irish and the Media, in Nic Pháidín, C, Ó Cearnaigh, S (Eds) *A New View of the Irish Language*, Dublin: Cois Life Teo.

Eurobarometer (2000) Report 52, retrieved from: http://europa.eu.int/comm/dg10/epo/

Faller, G (2007) *Irish Times*, 7 September.

Fishman, J A (1991) *Reversing Language Shift: Theoretical and Empirical Foundations of Assistance to Threatened Languages*, Clevedon: Multilingual Matters.

Gallagher, A (2006) Speaking in Tongues: the New Languages of Ireland, *Irish Times supplement, Who are we? The Irish in 2006*, 24 May.

Green, A (2007) *IELTS washback in context: preparation for academic writing in higher education*, Studies in Language Testing vol. 25, Cambridge: UCLES/Cambridge University Press.

McCloskey, J (2001) *Guthanna in éag/Voices silenced*, Dublin: Cois Life Teo.

Ó Muirí, P (2005) *Irish Times*, 6 July.

Ó Tuathaigh, G (2008) The State and the Irish Language in Nic Pháidín, C, Ó Cearnaigh, S (Eds) *A New View of the Irish Language*, Dublin: Cois Life Teo.

Punch, A (2008) Census Data on the Irish Language, in Nic Pháidín, C, Ó Cearnaigh, S (Eds) *A New View of the Irish Language*, Dublin: Cois Life Teo.

Quarterly National Household Survey (2008) Quarter 1 2008, retrieved from: http://www.cso.ie/releasespublications/document/labour_market /current/qnhs.pdf

Údarás na Gaeltachta website: www.udaras.ie/index.php/1165, accessed 28 August 2008.

Watson, I (2008) The Irish Language and Identity, in Nic Pháidín, C, Ó Cearnaigh, S (Eds) *A New View of the Irish Language*, Dublin: Cois Life Teo.

Weir, C and Milanovic, M (2007) Series Editors' Note, in Green, A *IELTS washback in context: preparation for academic writing in higher education*, Studies in Language Testing vol. 25, Cambridge: UCLES/Cambridge University Press, ix-xii.

A framework for migration and language assessment and the Skills for Life exams

SZILVIA PAPP RESEARCH AND VALIDATION GROUP, CAMBRIDGE ESOL
MARTIN ROBINSON ASSESSMENT AND OPERATIONS GROUP, CAMBRIDGE ESOL

Introduction

This article discusses the UK Government's recent policy decisions related to assessing language proficiency for migration purposes[1]. We outline the present arrangements made by the Home Office for testing for naturalisation, settlement and access purposes, including the recently introduced Points Based System and the list of English language tests that are currently accepted as proof of language proficiency based on the Home Office's criteria of acceptance. Cambridge ESOL language tests that have been used for these purposes include IELTS, Skills for Life, and ESOL for Work. We systematically evaluate Skills for Life for fitness for purpose within a framework of criteria developed by the Language Assessment for Migration and Integration subgroup of the Association of Language Testers in Europe in collaboration with the Council of Europe. The framework allows policy makers to consider issues relevant to language assessment for migration in a wider perspective, considering aspects of quality and fairness of language tests accepted and used for the purposes of migration, residency and citizenship.

1. This paper is based on one presented at the AERA 2008 conference, available online, see www.iaea2008.cambridgeassessment.org.uk

Recent history

The precursor of recent language testing for citizenship and settlement in the UK is the publication of the education White Paper entitled *Excellence in Schools* in 1997, when Citizenship became a separate subject rather than a cross-curricular theme in schools in England. Following this, the Department for Education and Skills (DfES) introduced Citizenship education into secondary schools in 2002. In the same year, the latest Nationality, Immigration and Asylum Act foreshadowed the round of renewed activities around migration. These activities are strongly reminiscent of concerns and solutions of the past 100 years (Saville 2008a, 2008b, 2009).

Life in the UK Test

Since July 2004, applicants for UK citizenship have had to demonstrate knowledge of the English language. In June 2005, a test with a language element, called the Life in the UK test was developed by Ufl/Learndirect Ltd for the Home Office (HO) and piloted at 8 test centres. All applicants for citizenship or British nationality have had to pass this test since 1 November 2005. Since 2 April 2007, applicants for settlement have had to take it as well to be granted

indefinite leave to remain. There is no sample test available, but the HO offers guidelines on their website about the content and format of the test in an online tutorial. The materials provided by the HO towards preparation for the test include a Handbook for teachers and mentors of immigrants entitled *Life in the UK: A journey to citizenship*. The original Handbook was published in Dec 2004. A revised and simplified version was published specifically for candidates in March 2007. The targeted proficiency level of the test has always been ESOL Entry Level 3 or above, B1 or above in CEFR (Common European Framework of Reference) terms. For those learners not yet at this level, learning materials for ESOL with citizenship courses were developed by the National Institute for Adult and Continuing Education (NIACE) and LLU+ (formerly the London Language and Literacy Unit) and piloted with 18 ESOL providers between September 2004 and March 2005, prior to the first launch of the test.

Points Based System

In addition to these developments in testing for citizenship and settlement, an Australian and Canadian-style Points Based System (PBS) is currently being phased in for access/entry for people wishing to come or remain in the UK to work, train or study. In the PBS, launched in the first quarter of 2008, different groups of applicants are required to demonstrate knowledge of English at different levels of proficiency. In Tier 1 of the PBS, from April 2008 onwards, highly skilled migrants wishing to come or remain in the UK to work, train or study are required to demonstrate knowledge of English at level C1 of the CEFR prior to entry. These are people in the finance, IT, medicine, science, and engineering sectors as well as entrepreneurs and investors. Tier 2 of the PBS was to include skilled workers with a job offer in a national shortage area (see schedule by HO, June 2008). Originally, this group of people were thought to come from the health service, white collar and trade sectors, e.g. work as nurses, teachers, and administrators. They were to demonstrate knowledge of English at level B2 of the CEFR prior to entry. More recently, there is more differentiation within the group: 'For Tier 2 General, Intra-Company Transfers and Sportspeople, the test will be required to be close to level A1 on the CEFR. This test should include an ability to understand and use everyday expressions and very basic phrases, to introduce themselves and others and ask and answer questions about basic personal details. For Tier 2 Ministers of Religion, the test will be required to meet level B2 on the CEFR' (HO UK BA, June 2008).

In Tier 3, being rolled out in the course of 2008–2009, limited numbers of temporary low-skilled workers may also be required to demonstrate knowledge of English. These people will typically work in the agriculture, food processing, hospitality sectors, e.g. as catering or construction workers. Tier 4 will apply to students, specialist workers, and trainees, whereas Tier 5 will include temporary workers with non-economic objectives, such as young people on youth mobility or cultural exchange programmes, young people on temporary work experience or on a working holiday, gap year students, professional musicians on tour, professional sports people, and other performers. The HO has not yet specified the level of English required of these groups. However, as the results of the recent consultation on marriage visas show, there is strong indication that even would-be spouses of British citizens or settled individuals applying for marriage visas will be required to show evidence of knowledge of English at level A1 of the CEFR prior to entry.

Home Office-approved English language tests

Migrants to the UK can demonstrate that they meet the required standard of English by providing evidence that they have passed an English language test. The HO has stated that only tests that have been assessed as having met the HO's published criteria will be accepted as evidence. The procedure for inclusion on the list of approved English language tests states that the tests have to meet the following requirements:

- Components – the test should cover the four language components of listening, speaking, writing and reading.
- Security of test – the test should be secure to prevent any possible attempt at deception during test sittings.
- Security of awards or certificates – the test awards should be secure and encompass a number of security features.
- Verification – the organisation should have a system in place to enable the UK Border Agency to verify test results where there are doubts about the test certificate.
- Fast track service – preferably the organisation should offer a service where applicants can request the fast-tracking of their test and results.

Organisations also have to provide details about the tests themselves and the reporting of results, including how the test is mapped to the CEFR; the overview, structure and levels of the test; where and when the test can be sat; any special needs arrangements and how long test results are valid. If the HO is satisfied that the information provided meets these requirements the test is accepted and details are published on their website (HO BIA, Oct 2007 and HO UK BA, June 2008). Any test included on the HO list of accepted tests can then be used by migrants wishing to demonstrate their English language ability. The question is whether these requirements are stringent enough.

An alternative framework for evaluating tests for migration purposes

Given the amount of criticism levelled against the current Life in the UK test, which has a language element, and some other English language tests on the HO's published list of accepted tests, it would be preferable to evaluate all proposed tests of English according to more stringent criteria. For instance, the following questions need to be answered:

- What purpose(s) was the test developed for?
- What does the test measure (i.e. how is the construct behind the test defined)?
- Who has been involved in developing, evaluating and revising the test?

• How best can one test language knowledge and skills needed for study, training, employment, social integration or the exercise of civic rights and responsibilities?

By having a language requirement as part of the requirements for citizenship, settlement, and permission to work, train or study in the UK, there is an underlying assumption that proficiency in the language of communication in the wider society is a democratic responsibility and a democratic right. This has implications not only for the candidates but wider society (government, policy makers, employers, colleges, schools, language course providers, etc.) as well. Also, it has implications for the development and administration of tests conceived and designed for these purposes (i.e. for professional language testers and stakeholders). Tests developed and accepted for these purposes need to be defensible. The question is whether there is a defensible argument, backed by evidence, for tests accepted for these purposes, in any of the frameworks developed within the assessment community. The framework chosen for test evaluation would reflect the primary issues and concerns and could be either Weir's (2005) socio-cognitive framework focusing on aspects of different types of test validity hence quality (see Papp and Wright 2006) or Kunnan's (2008) micro- and macro-framework of test fairness.

In the following discussion, we combine Kunnan's language test fairness criteria with those considerations set out in a guide developed by professional language testers and others representing the Language Assessment for Migration and Integration (LAMI) special interest group of the Association of Language Testers in Europe (ALTE) working in collaboration with the Council of Europe (CoE). The outline for policy makers set out by members of the ALTE LAMI group (henceforward refereed to as ALTE LAMI, 2008) includes considerations that need to be made when planning or accepting language tests for these particular purposes. The considerations include the mode of assessment, impact and consequences of the test, stages of test development and operation, including determining test purpose, linguistic demands and the appropriate level of difficulty, producing test specifications, and monitoring examiners.

According to the LAMI group, it is imperative to first decide on the purpose of the test:

> 'The first step in this process is the precise and unambiguous identification of the purpose of the test. After this is done, principled ways to determine the content and difficulty follow. Finally, the test specifications document, a document essential in later stages of test construction and review, must be developed' (ALTE LAMI, 2008).

Policy makers also need to reflect whether the purpose of the test is:

1. Motivating learners (to help them use and improve their current competence in the target language)

2. Ascertaining whether their competence is sufficient for participation in well-defined social situations (e.g. study or work and also other social situations more connected with the exercise of citizenship),

3. Making decisions which affect their legal [as well as human and civil] rights, such as their right to remain in a country or acquire citizenship of it. (ALTE LAMI, 2008)

It could be argued that only the second purpose can be tested in language tests. The first, motivational purpose represents a humanitarian perspective. The second purpose, assessing required competence for social participation, reflects an egalitarian aspect; and the last one, with the aim of making decisions which affect people's legal, human and civil rights, reflects a pragmatic, arguably cynical, legal and political purpose (cf. McNamara and Roever 2006). All of these purposes are recognised and widely discussed in the public and professional spheres in the UK. However, professional language testers can advise only on the second purpose with authority based on their expertise as well as professional codes and standards. For the other two applied objectives impact and washback studies are needed in order to ascertain the beneficial and potentially harmful effects of a test on society as a whole and its institutions, and teaching and learning within the classroom.

It is pointed out in the LAMI guidelines that as well as informing test development, a clear and explicit purpose will not only help to clarify test takers' expectations towards the type, content and marking criteria of the test (all to be specified and made public in the test specifications), thus contributing to test fairness, but will also allow other members of society to interpret and use test results appropriately (ALTE LAMI, 2008).

As a next step, 'only when the purpose has been clearly defined is it possible to identify the real-world demands that test-takers will face (e.g. the need to take part in societal processes and exercise the rights and responsibilities of citizenship) and which should be reflected in the test' (ALTE LAMI, 2008). The process of establishing the needs of candidates is termed *needs analysis*. However, needs analysis tends to be a taxonomic exercise which tends to significantly restrict the construct behind any test. The questions policy makers and test developers for migration purposes need to address are:

• What level and type of social participation are migrants supposed to aim for?

• What level of 'exercising their civil rights and responsibilities' are they expected to demonstrate – is it up to representing other people in democratic processes?

In addition to these considerations, Kunnan (2008) defines an ethical and fair test as displaying:

• comparable or equitable treatment in the testing process,

• comparability or equality in outcomes of learning and opportunity to learn,

• absence of bias in test content, language and response patterns, and

• comparability in selection and prediction.

According to the LAMI group, 'test fairness is relevant to all types of language test and for all target candidates, but is especially important in the case of tests of language for migration, residency and citizenship, due to the serious

implications for the test taker in terms of civil and human rights' (ALTE LAMI, 2008). For a fair and ethical test, a defensible argument is needed, backed by evidence collected in support of Kunnan's (2008) five test fairness qualities: validity, absence of bias, access, administration, and social consequences.

The ALTE LAMI document makes direct reference to the CEFR, as a framework of reference for language tests for migration, particularly to enable determining the difficulty level of these tests. Some scholars have argued that the CEFR may not be appropriate, suitable, achievable, or justified for these purposes (see e.g. Krumm 2007, Hulstijn 2007:665, and Alderson 2007:662). Despite these reservations, the CEFR has been successfully applied to adult refugee immigrants, young migrants and minority groups in schools (see Little 2007). In line with the ethos of the CEFR, the task is to adapt it for the relevant groups of learners and the specific purposes, i.e. taking the context into account. In the case of language assessment for migration purposes, the relevant groups are naturalistic adult and child L2 learners, studied longitudinally, for instance as in the early and highly influential SLA studies carried out in the first half of the 1990s in Germany with 5 L1 Gastarbeiter groups by Klein and Perdue (1992). The descriptors of these learners' performance need to be developed and expressed in language-specific, functional, linguistic, and socio-cultural exponents, based on learnability considerations, as well as actual language needs and language use of these groups of learners.

With regard to Kunnan's (2008) criteria of access and administration, the following considerations would need to be made:

- Do candidates have equal opportunities?

- Do they have comparable educational, financial, geographical, personal access to language courses and the test itself?

- Do they have equal opportunity to learn, to practise language learned in everyday life, to become familiar with the test, based on time engaged in high quality learning activities?

- Are the courses and the test affordable, are the accepted payment methods accessible?

- Are students able to attend classes and the test and not hindered by geographical distance?

- Are there accommodations for special needs?

- Is familiarity with test conditions, procedures and equipment assumed?

Exemplification of the framework

As seen in the previous section, the ALTE LAMI group have produced their outline for policy makers to enable them to better judge the merits of various language tests for social cohesion and citizenship. Exemplification of this framework is provided in the form of the Cambridge ESOL Skills for Life examinations. The following considerations are reviewed below for the Skills for Life tests:

- Test purpose and real-world demands on test takers

- Linguistic demands

- Appropriate level of difficulty linked to the CEFR

- Producing test specifications to cover test format, task and item types, assessment criteria, item writing, pretesting, test administration, special requirements, marking and grading

- Monitoring examiners, candidate responses and demographic information in order to be able to equate test versions and identify possible bias and to ensure test functionality and quality.

Determining test purpose and real-world demands on test takers

The Cambridge ESOL Skills for Life qualifications contribute to the Government's Skills for Life initiative by providing assessment of ESOL which is designed around the standards for adult literacy and the ESOL Core Curriculum (DfES 2003). This is designed to cover the standards and curriculum with sufficient range, breadth and depth to properly reflect the use of the English language in real world situations, appropriate to the needs of the client groups and the NQF level in question. The target users for the qualifications are settled communities, asylum seekers and settled refugees, migrant workers who work or settle in the UK for most of their lives, and partners and spouses of learners who may have been settled for a number of years. The qualifications are designed to reflect the fact that target learners' educational and employment backgrounds are often highly diverse, from people with no previous education or employment to highly educated professionals.

Cambridge ESOL Skills for Life certificates are available at Entry 1 (A1), Entry 2 (A2), Entry 3 (B1), Level 1 (B2) and Level 2 (C1) on the National Qualifications Framework. They are also available as separate modes or skills (Reading, Writing and Speaking and Listening) at each of the 5 levels. These individual qualifications make the examinations very accessible to candidates with a very broad range of individual level and skill profiles.

Determining linguistic demands

Linguistic demands were determined partly by designing the tests around the standards for adult literacy and the ESOL core curriculum. However, a thorough needs analysis was also undertaken involving specialists in the field of ESOL learning and teaching and this was then followed by detailed trialling on the target language users.

It was expected that the learners' educational and employment backgrounds will be diverse, as well as their aspirations, literacy levels and language learning skills. This diversity is reflected in the range of material selected for use in these tests and in the task types which candidates need to complete. Most of the topic areas correspond to themes contained in the ESOL Citizenship syllabus.

At Entry 1–3 levels candidates will encounter topics in the tests which are both familiar and relevant to them as learners of English. Topics may include: personal details and experiences, work, education/training, housing, family and friends, health, transport, weather, buying goods, leisure, UK society. At Levels 1 and 2, the tasks are designed to reflect the fact that the needs of ESOL learners

at these levels will be predominantly educational or professional, though broader social issues are also included in the materials where appropriate. Therefore, candidates may also encounter topics such as careers, academic study, information technology, the environment, law and order.

In line with the overall objectives of the Adult ESOL Core Curriculum, the choice of contexts, topics and texts is designed to encourage an understanding of spiritual, moral, ethical and cultural issues. It will also contribute to raising awareness of environmental issues and health and safety considerations, as well as European developments consistent with relevant international agreements.

Determining the appropriate level of difficulty linked to the CEFR

The levels of the Skills for Life examinations are demonstrably based on the Adult ESOL Core Curriculum (DfES 2003) and therefore the national standards for adult literacy and numeracy and specified at levels Entry 1 though to Level 2. The test specifications and more detailed item writer guidelines include thorough instructions on writing to these levels. Cambridge ESOL Skills for Life examinations have also been mapped to the CEFR at levels A1 to C1. Cambridge ESOL examinations are the only certificated English language exams identified in the framework as being specifically linked to it by a long term research programme. Cambridge ESOL's processes include a range of measures to ensure alignment to the CEFR.

When developing an examination, the test specifications include the CEFR level at which the exam will be set so that task types are chosen with the specific level in mind. The more detailed item writer guidelines include very detailed instructions on writing to each level of the CEFR and the writers themselves are trained to write to particular levels.

To ensure consistency of the level of the examination, test questions are pretested and calibrated using anchor items to monitor exam difficulty before they are used in live exam papers. The calibrated items are then stored in the Cambridge ESOL Local Item Banking System (LIBS) where each item has a known difficulty and accordingly test papers are constructed to a target difficulty.

Where human raters are involved, Cambridge ESOL follows a rigorous system of recruitment, induction, training, coordination, monitoring and evaluation (RITCME). Obligatory standardisation of writing examiners and general markers takes place prior to every marking session, and the writing samples used are evaluated by the most senior examiners for the paper. Oral assessors follow a similar RITCME process and are standardised using standardisation DVDs containing video recordings of Skills for Life candidates at each of the CEFR levels. Grading and post-examination review take place for each version of the test. Results of live administrations are monitored across a number of years to ensure consistent standards over time.

Producing test specifications

For Speaking and Listening, at each level, the assessment focuses on gathering evidence that the candidate can listen and respond, speak to communicate and engage in discussion. For Reading and Writing, at each level, the assessment focuses on gathering evidence that the student can operate at word, sentence and text level, while recognising that conveying meaning involves operating at these three levels simultaneously. This testing focus is then broken down into specific, individual testing points which can be assessed by using a suitable combination of test tasks and task types.

Item writing

The specifications and the more detailed item writer guidelines include very detailed instructions on writing to each level of the National Qualifications Framework and to each level of the CEFR. The items writers themselves have extensive ESOL teaching experience and are trained over a number of years to write to particular levels. Item writers work as external, independent consultants in teams on each of the Skills for Life modes and are led by very experienced Chairs or team leaders who are also external, independent consultants. The item writing process incorporates extensive monitoring, feedback and ongoing training for all item writers.

Pretesting

All test questions are pretested and calibrated using anchor items to monitor exam difficulty before they are used in live exam papers to ensure consistency of the level of the examination. Pretesting is carried out in Skills for Life teaching locations with large numbers of actual test takers, ensuring a mixture of first language, nationality, age and gender. Both teachers and students have the opportunity to feed back on the suitability and level of the questions. The calibrated items are then stored in the Cambridge ESOL Local Item Banking System (LIBS) where each item has a known difficulty and accordingly test papers are constructed to a target difficulty.

Test administration

All Cambridge ESOL exams are taken at authorised Cambridge ESOL examination centres which must meet high standards of professional integrity, security and customer service, and are subject to inspection by Cambridge ESOL. Centres running the exams can access detailed administrative information through a secure extranet, including handbooks for centres, administration instructions and regulations, and interlocutor training packs. Cambridge ESOL Skills for Life exams are on demand, enabling colleges and other learning providers to offer them at a time that suits both them and their students.

Special circumstances

Special circumstances cover three main areas: special arrangements, special consideration and malpractice. Special arrangements candidates with visual, hearing, learning and other difficulties are given the chance to gain a Skills for Life certificate by Cambridge ESOL's Special Arrangements service. Examples of the kinds of special services offered include papers in Braille, enlarged print, papers printed on different coloured paper and lip-reading

versions of Listening tests. Sample special papers are also made available so these candidates can have the same opportunity to practise as anyone else. Cambridge ESOL will give special consideration to candidates affected by adverse circumstances immediately before or during an examination for example due to illness or other unexpected events. Malpractice Committees will consider cases where candidates are suspected of copying, colluding, or breaking the examination regulations in some other way.

Marking and grading

For Reading, marking reliability is ensured through extensive trialling and pretesting of tasks leading to detailed calibration of items, and by standardisation and co-ordination of clerical markers and examiners. Grading the exams takes into account statistics from pretesting and trialling, statistics on the candidature, statistics on the overall candidate performance, statistics on individual items, advice and recommendations of examiners, and comparison with statistics on exam performance on previous test versions. For Writing and Speaking and Listening, examiners mark according to criteria based on the Adult ESOL Core Curriculum, where candidates need to show that they can ordinarily achieve the stated standards. Writing examiners are recruited from ESOL practitioners meeting appropriate professional minimum standards. An induction process takes place where examiners are familiarised both with Cambridge ESOL standard procedures and with paper-specific issues. Training and co-ordination takes place through extensive engagement with, and discussion on, a large number of sample scripts showing a range of achievement across band levels. Throughout the marking process extensive monitoring of examiner performance is followed up by both qualitative and quantitative evaluation and feedback.

Monitoring examiners and candidates

Cambridge ESOL follows a rigorous system of recruitment, induction, training, coordination, monitoring and evaluation human raters (known as RITCME). Obligatory standardisation of Writing examiners and general markers takes place prior to the marking of all new test versions, and the writing samples used are evaluated by the most senior examiners for the paper. Standardisation of oral examiners takes place regularly and the video samples of performances are rated by Cambridge ESOL's most experienced Senior Team Leaders and Team Leaders, representing a wide range of experience and familiarity with level. The marks provided are then subject to quantitative (SPSS and FACETS) and qualitative analysis before being approved for standardisation purposes.

Cambridge ESOL continuously collects and analyses both information about candidate responses and demographic information about the candidates (e.g. age, gender and nationality). This together with information collected through pretesting and the analysis of live response data helps to ensure that the test items are free from bias and that candidates receive consistent results regardless of test version. This ensures that each test version works in a way which is fair to all targeted test takers, no matter what their background and that test results accurately portray the ability of the candidate. Conclusions from such analyses are constantly fed back into the item writing, test construction and grading processes so that these processes are continually improved.

Limitations of the framework for evaluating tests for migration purposes

The first step in the suggested framework is to determine the test purpose and from this to undertake a needs analysis in order to determine the real-world demands on the test takers. However, it is unlikely that there will ever be a single test that is designed solely to test language for migration, residency and citizenship purposes. What is more likely to happen is that language tests already in existence will be approved for these purposes, too. This means, of course, that tests approved for migration purposes would have originally been designed with a different purpose in mind. This may mean, for example, that the test was designed for a different target candidature and therefore the test format and content may not be entirely appropriate. In the case of the Cambridge ESOL Skills for Life tests, they were, in fact, designed for refugees and migrant workers, a candidature that is similar in some respects to the range of learners who have more recently taken them. However, they were designed for settled refugees and migrant workers already living in the UK. This is not the same as migrants who wish to come to the UK but who are currently resident in another country. This discrepancy would not just affect the content of the exam. The Skills for Life exams, since they were designed for settled migrants and refugees (and their partners), can only be taken in the UK. Clearly, migrants who are planning to come to the UK but who do not yet have the visa status to stay in the UK need to be able to take a language test outside the UK. So, whereas the Skills for Life tests would appear to be relatively suitable in terms of test format, content, level, etc, the fact that they cannot currently be taken outside the UK is a major limitation at the moment. Cambridge ESOL is currently running a study on the impact of the Skills for Life examinations on the careers and life chances of candidates, as well as the issues and concerns other stakeholders have about using these tests for migration purposes.

Conclusion

Since 2002, the year when the policy of language testing for migration and social integration purposes started to be formulated in the UK, the Language Assessment for Migration and Integration (LAMI) subgroup within ALTE (Association of Language Testers in Europe) has been working closely with the Council of Europe and relevant UK government departments, such as the Home Office Border and Immigration Agency and the Advisory Board on Naturalisation and Integration (ABNI). Since this is an area of fast changing policy-making with real social consequences, it is very important that the initial dialogue and collaboration between professional language testers and policy makers and stakeholders is maintained and

enhanced in order to achieve successful social cohesion and integration in the UK.

By accepting more types and a wider range of evidence of knowledge of English and by differentiating specific groups of test takers, the recently introduced points based system will hopefully act as an instrument for recognition rather than a systematic means for exclusion. It will also tease apart content (knowledge of life in the UK) from language skills, and address the minimum language requirements for different purposes and different migrant groups. In the future, it might even, hopefully, profile and recognise partial competencies and recognise multicompetence which makes up migrants' linguistic capital. If it is implemented correctly and fairly, it could ultimately act as a strategic, forward-looking, overarching system for social cohesion (protection, integration, inclusion), rather than being part of the current retrospective management of migration (risk management, 'fire fighting', and inevitably, exclusion) in the UK.

Whatever policy making system is chosen, however, it is imperative that only language tests that meet rigorous standards should be approved for these purposes and it is here that the work of such special interest groups as the LAMI subgroup within ALTE have an important role to play.

References and further reading

ABNI (Nov 2006) *Progress towards integration: 2005/2006 annual report*, November 2004–April 2006, retrieved from www.abni.org.uk/publications/index.html

— (Oct 2007) *Second annual report*, April 2006–October 2007, retrieved from www.abni.org.uk/publications/index.html

— (Nov 2008) *Final report*, November 2004–November 2008, retrieved from www.abni.org.uk/publications/index.html

Alderson, J C (2007) The CEFR and the need for more research, *The Modern Language Journal* 91, 659–663.

ALTE LAMI Authoring Group (2008) Language tests for social cohesion and citizenship – an outline for policymakers, *Council of Europe Thematic Studies*, Strasbourg: Council of Europe, retrieved from www.coe.int/t/dg4/linguistic/MigrantsSemin08_ListDocs_EN.asp#TopOfPage

DeVelle, S and Papp, S (Feb 2008) *Response to HO BIA's Marriage visas: pre-entry English requirement for spouses Consultation Paper* (Dec 2007), internal Cambridge ESOL document.

DfES (2003) *Pathways to proficiency: The alignment of language proficiency scales for assessing competence in English language*, retrieved from www.dcsf.gov.uk/readwriteplus/LearningInfrastructureAccreditation

HO BIA (Oct 2007) *Points based system: Procedure for inclusion on the list of approved English language tests. Tier 1*, retrieved from www.ukba.homeoffice.gov.uk/sitecontent/documents/managingourborders/pbsdocs

— (Dec 2007) *Marriage visas: pre-entry English requirement for spouses Consultation Paper*, retrieved from www.ukba.homeoffice.gov.uk/sitecontent/documents/aboutus/consultations/closedconsultations/preentryenglishrequirement

— (June 2008) *Points based system: Procedure for inclusion on the list of approved English language tests. Tier 2*, retrieved from www.ukba.homeoffice.gov.uk/sitecontent/newsarticles/pbsenglishlanguagetest

Hulstijn, J H (2007) The shaky ground beneath the CEFR: Quantitative and qualitative dimensions of language proficiency, *The Modern Language Journal* 91, 663–667.

Klein, W and Perdue, C (1992) *Utterance Structure (Developing grammars again)*, Amsterdam: John Benjamins.

Krumm, H-J (2007) Profiles instead of levels: The CEFR and its (ab)uses in the context of migration, *The Modern Language Journal* 91, 667–669.

Kunnan, A J (2008) Towards a model of test evaluation: Using test fairness and the test context frameworks, in Taylor, L and Weir, C J (Eds) *Multilingualism and Assessment: Achieving transparency, assuring quality, sustaining diversity – Proceedings of the ALTE Berlin Conference*, May 2005, Studies in Language Testing vol. 27, Cambridge: UCLES/Cambridge University Press.

Little, D (2007) The CEFR for Languages: Perspectives on the making of supranational language education policy, *The Modern Language Journal* 91, 645–655.

McNamara, T and Roever, C (2006) *Language Testing: The social dimension*, Oxford: Blackwell.

Papp, S and Wright, S (2006) *Language tests and citizenship: A legacy of the past*, unpublished manuscript, University of Portsmouth, School of Languages and Area Studies.

Saville, N (2003) The process of test development and revision within Cambridge EFL, in Weir, C and Milanovic, M (Eds) *Continuity and Innovation: Revising the Cambridge Proficiency in English Examination 1913–2002*, Studies in Language Testing vol. 15, Cambridge: UCLES/Cambridge University Press.

— (2006) Language testing for migration and citizenship, *Research Notes* 25, 2-4.

— (2008a) *Language assessment in the context of migration and citizenship: A framework for considering issues*, colloquium presented at American Association for Applied Linguistics conference, Washington DC, 29 March–2 April.

— (2008b) *UK Migration – 100 Years on*, unpublished manuscript, Cambridge ESOL.

— (2009) Language assessment in the management of international migration: A framework for considering the issues, *Language Assessment Quarterly* 6/1, 17–29.

ILEC Speaking: revising assessment criteria and scales

IVANA VIDAKOVIC RESEARCH AND VALIDATION GROUP, CAMBRIDGE ESOL
EVELINA GALACZI RESEARCH AND VALIDATION GROUP, CAMBRIDGE ESOL

Introduction

The International Legal English Certificate (ILEC) assesses language skills in a legal context at B2 and C1 levels of the Common European Framework of Reference for Languages (CEFR). It is intended for prospective law students and legal professionals and can be used as evidence of the level of language necessary to work in an international legal context or to follow a course of legal study at university level (ILEC Handbook 2007). The exam has four papers: Reading (with a Use of English section), Listening, Speaking and Writing.

The Speaking component consists of four parts, each of which elicits a different type of interaction and discourse patterns:

1. Interview

2. Long turn (a monologue on a chosen topic)

3. Collaborative task (in which two candidates engage in discussion towards a negotiated completion of the task)

4. Discussion (in which each of the two candidates is prompted by questions to develop and discuss a topic in more depth).

The assessment is carried out by two examiners using different assessment scales: an Interlocutor, who conducts the exam and uses global scales to award a holistic mark, while an Assessor uses analytic scales to award marks for four criteria: Grammar & Vocabulary (G&V), Discourse Management (DM), Interactive Communication (IC) and Pronunciation (PR). Each scale has six bands (1.0–6.0), with the possibility to award half bands as well. This multi-task paired test format (two examiners and two candidates) enables ILEC candidates to show a variety of Speaking sub-skills in a range of contexts and also allows assessment from two different yet complementary perspectives – global and analytic.

The original ILEC Speaking assessment scales were based on the FCE and CAE Speaking assessment scales, since these two exams are set at the same levels of CEFR (B2 and C1 respectively) as ILEC. After the revision of the FCE and CAE Speaking scales (cf. Galaczi & ffrench 2007), it was decided that the descriptors of the ILEC scales should also be changed in order to bring them in line with FCE/CAE updates. The main drive behind the revision of the Main Suite (MS), and consequently, the ILEC scales, originated from a perceived tendency among raters to make little use of marks at the lower end of the scale, which resulted in truncation of the scale. In order to address this issue, several modifications of the Main Suite Speaking assessment scales were carried out and they were also adopted in the revision of the ILEC scales. The most general modification consisted

of re-wording and defining the descriptors in positive terms, but some other changes were also made, such as the provision of concrete and transparent descriptors, shortening of the descriptors to make them easier to refer to during rating and defining each descriptor without reference to any other descriptors (ibid 2007:28).

The revision of the ILEC Speaking assessment scales was conducted in several stages. Phase 1 involved the commissioning of an external consultant to provide draft scales and descriptors; phase 2 consisted of a review and revision of the draft descriptors; phase 3 focused on the trialling of the scales and descriptors. The aim of the final trialling stage was to gather validity evidence through quantitative and qualitative methods about the assessment scales prior to their live use. This places the revision of the ILEC scales within the larger context of Cambridge ESOL's ongoing commitment to *a-priori* validation in performance assessment (Saville 2003).

The overarching aim of the ILEC Speaking assessment scales revision was to provide a better instrument for measuring candidate performance, which should have a positive impact on all users of test results – test takers, as well as Universities and prospective employers. This paper discusses the trialling stage of the revision process where the focus was to determine if the revised ILEC Speaking assessment scales are functioning adequately.

Research questions

The main aim of the trial was to provide answers to the following questions:

1. What is the overall performance of the raters in terms of harshness/leniency, consistency of ratings and agreement?

2. What is the overall performance of the assessment criteria in terms of difficulty?

3. Do any of the descriptors need further revision, as suggested by the rater feedback and statistical evidence?

4. What implications for rater training does the rater feedback provide?

Methodology

Thirty four raters participated in this study. Twenty four marked Speaking test performances using the revised analytic scales. They were divided into six groups and they rated four out of six tests (eight out of twelve candidates) each. Not all raters rated all performances due to practical reasons, but we ensured that there was enough overlap

Table 1: Information on raters: position in Cambridge ESOL and gender

Raters	Position in Cambridge ESOL				Gender	
	Senior Team Leader	Regional Team Leader	Team Leader	Oral Examiner	Male	Female
'Analytic' Raters	9	2	2	11	12	12
'Global' Raters	–	2	2	6	2	8

between raters and examinees for the purposes of the analysis. The other ten raters used the revised global scales and they all rated six tests (12 candidates) each.

Both 'analytic' and 'global' raters were experienced examiners with ILEC training, including Senior Team Leaders. The raters' backgrounds were varied in terms of the position and experience in Cambridge ESOL, gender and location to ensure representativeness of the sample (cf. Table 1). At the time of the trial, most raters were based in Europe (ten countries), some in Latin America (two countries) and some in Asia (four countries).

Procedure

In order to add to the validity of the trial, it was necessary to familiarise the raters with the new scales before they could apply them to the recorded speaking performances. Therefore, all raters did a familiarisation exercise which was intended to ensure they engaged with the new scales prior to using them on ILEC candidates. After that, the raters marked the candidates' performances using the following materials:

1. the revised scales (analytic or global),
2. a rating form and
3. a DVD with six tests.

The 'analytic' raters, who marked 4/6 tests, were told which tests to mark and in which order. Finally, after the rating exercise, the raters completed a feedback form.

The raters applied the revised scales to all available ILEC standardisation videos footage (2007 and some 2008 videos), which consisted of 12 candidates' performances (six Speaking tests). There was a spread of L1s in the sample: French, Greek, Iranian, Italian, Lithuanian, Pakistani, Spanish, Russian and Ukrainian. The spread of marks was not large since there were no performances below ILEC band 3 in the available sample.

Analysis

Multi-Faceted Rasch analysis (MFRA) was carried out using the programme FACETS. MFRA was deemed appropriate for the trial because it can provide estimates of the performance of the raters, candidates, scales and assessment criteria.

Quantitative findings: MFRA of the performance of the raters and scales

This section sheds light on the performance of the revised scales and the examiners and also presents and discusses the spread of marks obtained with the new scales.

Assessment criteria

A way of getting information on the performance of the revised analytic scales is by investigating the variability in difficulty among the assessment criteria and their degree of fit. The difficulty of each analytic criterion is presented in logits in Table 2. A higher logit value indicates a lower score and, therefore, a more harshly rated criterion. A lower logit value indicates a higher score and a more leniently rated criterion. Therefore, the highest scored criterion is Interactive Communication (IC), while the lowest scored one is Pronunciation (PR). The difference between the highest and the lowest scored criterion is less than a logit (0.67 logits), which is well within the acceptable range of variability in difficulty. This indicates that despite the small differences in logits, the four criteria performed in a similar fashion.

Table 2: Difficulty and infit of analytic assessment criteria

Criteria	Measure (Logits)	Error	Infit Mean Square
GV	0.13	0.07	0.96
DM	0.02	0.07	1.02
PR	0.39	0.07	1.01
IC	-0.54	0.08	0.97

As far as degree of fit (Infit Mean Square) is concerned, there are no prescribed rules as to the acceptable range, but the lower and upper limits of .5 and 1.5 were adopted in the present study since they have previously been found useful for practical purposes (Weigle 1998, Lunz & Stahl 1990 in Weigle 1998). All analytic criteria were within the acceptable degree of fit, which means that the scores awarded for a specific criterion were neither unpredictable nor overfitting.

Raters' performance

The raters' performance in using the revised scales in terms of harshness/leniency and (in)consistency can also be used to shed light on the performance of the revised scales. Problematic levels of rater harshness/leniency and inconsistency would naturally have implications for rater training, but they can also be 'symptoms' of a problematic assessment scale. The results showed that the raters who used the revised analytic scales rated as a homogenous group as far as the degree of harshness/leniency is concerned. All but two were within the acceptable limits of severity (-1.0 and +1.0 logit, following Van Moere 2006, Galaczi 2007, DeVelle & Galaczi 2008): one rater being the most lenient (-2.81 logits) and the other the harshest (1.37 logits), as evident in Table 3. In addition, most raters were consistent in awarding marks. Only 3 raters out of 24 were found to be inconsistent, with the degree of infit higher than 1.5. Four raters had infit values lower than 0.5, which indicates that they exhibited the central tendency, i.e. mostly used the middle part of the scale. This was not seen as a cause for concern because the four raters were still very close to the lower cut-off point.

The agreement between the raters on all analytic criteria was moderately high, ranging from 0.49 to 0.68 (in point biserial coefficients). The agreement between raters on

Table 3: 'Analytic' raters' severity/leniency and consistency

All Analytic Criteria			
Rater	Measure (Logits)	Error	Infit Mean Square
1	-0.02	0.18	1.42
2	-0.57	0.19	1.03
3	-0.59	0.19	0.44
4	**-2.81**	0.29	**1.84**
5	-0.33	0.18	0.52
6	0.80	0.17	1.38
7	-0.69	0.19	0.75
8	-0.47	0.19	0.73
9	-0.35	0.19	0.49
10	-0.01	0.18	0.62
11	-0.02	0.18	0.72
12	-0.31	0.19	0.48
13	**1.37**	0.17	**1.30**
14	0.31	0.17	0.71
15	0.16	0.17	1.09
16	0.82	0.17	1.36
17	0.83	0.17	0.41
18	0.70	0.17	0.55
19	-0.33	0.18	1.06
20	0.22	0.18	**1.84**
21	0.43	0.17	0.90
22	-0.11	0.18	0.59
23	0.21	0.18	1.00
24	0.75	0.17	**2.56**

Table 4: 'Global' raters' severity/leniency and consistency

Global Achievement			
Rater	Measure (Logits)	Error	Infit Mean Square
1	-0.54	0.33	0.70
2	0.40	0.32	0.78
3	**1.60**	0.30	**2.75**
4	**1.15**	0.30	0.60
5	-0.97	0.33	0.62
6	-0.22	0.33	0.57
7	-0.22	0.33	0.29
8	**-1.08**	0.33	1.16
9	-0.11	0.33	1.03
10	-0.01	0.32	0.58

Table 5: Fair average marks – all analytic criteria

All Analytic Criteria		
Candidates	Fair Average Mark	Fair Average Mark Rounded to Nearest 0.5
1	2.85	3
2	3.82	4
3	3.95	4
4	5.31	5.5
5	4.29	4.5
6	5.65	5.5
7	4.13	4
8	5.18	5
9	3.19	3
10	5.41	5.5
11	5.82	6
12	4.30	4.5

Table 6: Fair average marks – global achievement criterion

Global Achievement		
Candidates	Fair Average Mark	Fair Average Mark Rounded to Nearest 0.5
1	2.62	2.5
2	4.09	4
3	4.39	4.5
4	5.76	6
5	4.57	4.5
6	5.71	5.5
7	3.75	4
8	4.77	5
9	2.71	2.5
10	5.50	5.5
11	5.81	6
12	4.77	5

each analytic criterion was quite high for DM (ranging from 0.73 to 0.98) and moderate to high, but mostly high, for the rest of the criteria (GV: 0.49–0.99, PR: 0.5–0.98, IC: 0.6–0.98).

Just like the raters who used the analytic scales, the majority of the raters who used the revised global scales were within the acceptable limits of severity and consistency. Only two raters were identified as harsh with a logit measure of 1.6 and 1.15, respectively, while one was slightly lenient, with a logit measure of -1.08 (cf. Table 4 below). As far as rating consistency is concerned, only one rater was flagged up as inconsistent with an infit mean square of 2.75. Another rater had a low infit measure which indicated that they predominantly used the middle part of the scale. The majority of the raters (eight out of ten) were within the acceptable parameters, and thus, rated consistently. In addition, the 'global' raters were all in high agreement with each other (0.71 to 0.97 in point biserial coefficients).

The Spread of Marks

Fair average marks were produced by FACETS by adjusting the raw scores to account for the variation in the difficulty of the criteria, the harshness of the raters and the ability of the candidate. Tables 5 and 6 display fair average marks for the analytic and global criteria. As can be seen, there is a relatively good spread of marks in the upper part of the scale (2.5 onwards) where most bands and half-bands have a representative candidate, but no means below 2.5.

As stated in the Methodology section, the spread of original marks, i.e. candidate abilities, in the available trial sample was not adequate since there were no performances below band 3. This is a limitation of this trial, but it also indicates that the problem of a limited spread of fair average marks at the bottom end of the scale is a shortcoming of the available candidate sample, rather than the revised scales. This is an important issue with implications for rater training, since it gives raters limited exposure to lower-ability candidates. The issue is being

addressed through a change in the procedures for collecting candidate samples for future standardisation videos.

Overall, the findings have yielded encouraging information about the performance of the revised analytic and global scales and the raters who used them. The low variability in difficulty among the assessment criteria and their acceptable degree of fit have revealed that the revised scales performed adequately during the trial. The vast majority of the raters rated homogenously as a group in terms of harshness/leniency and most of them were consistent in their ratings. The levels of agreement between the raters were moderately high and high. The overall performance of the raters and assessment criteria can be taken to provide positive validity evidence for the revised ILEC Speaking assessment scales.

Qualitative analysis: 'analytic' raters' feedback

In order to obtain further information on the behaviour of the revised scales, it was necessary to complement the quantitative data analysis with qualitative evidence derived from the raters' written feedback. The analytic raters were asked to state the following in the feedback form:

- How easy it was to apply the revised analytic scales (on a scale 1 'very easy' to 6 'very challenging')
- Which criterion they found easiest and most difficult to apply
- How confident (on a scale 1 'not confident' to 6 'very confident') they were about rating each analytic criterion (G&V, DM, PR and IC) with the revised scales.

Generally, the raters' feedback was positive and their answers exhibited the following trends:

- The majority of the raters (15/24) found the revised analytic scales easy to apply. Even though the application of the revised scales was easy for the majority, nine raters found it more challenging (choosing bands 4 and 5, but never 6).
- In general, the easiest criterion to apply was Pronunciation, while the most difficult criterion to apply was Grammar & Vocabulary. Unlike these two criteria which exhibit a clear trend, Discourse Management and Interactive Communication seemed to be dependent on individual preferences.
- The vast majority of the raters were confident about rating, with the revised analytic assessment scales. Most raters chose the bands representing higher degrees of confidence (4–6): 18/24 raters chose those bands for G&V, 20 for DM, 21 for PR and 19 for IC. Even though the majority of the raters were confident about rating the assessment criteria accurately with the revised scales, there were a certain number of raters who were not so confident.

The reasons for these trends were sought in the raters' comments provided in the feedback form. These comments could help identify the areas where the scales could benefit from a further revision and where the raters could benefit from further training. Some more general and frequently mentioned issues are discussed below.

Differentiation between the bands

Although differentiation between the bands was found easy by some raters, it was found difficult by others due to very subtle differences in phrasing between the bands. A few comments are provided below, beginning with the positive ones:

> 'Differences between bands are clearly outlined in the way in which descriptors are formulated.'

> 'I feel reasonably confident with the levels of this exam and distinguishing between a B2 and C1 candidate.'

> 'In general, the scales clearly set out the features distinguishing different bands…'

> 'Descriptors of criteria across levels usually depend on one word difference and this makes it hard to place candidates e.g. between 3.0 and 5.0.'

This is an issue which will be addressed during rater training, and it is natural, perhaps, to expect such comments since at the time of the trial the raters had not had much experience with using the new revised scales.

The relation of the ILEC Speaking bands to the CEFR

In terms of which band corresponds to which CEFR levels, the relation of the ILEC Speaking bands to the CEFR is not clear to all raters, as illustrated by one comment:

> 'The jump from 5.0 to 6.0 is still new… Where is 6.0 on the CEFR, is it at the top of C1 or somewhere in the middle? If it is at the top, what does this mean about band 5.0? These were the questions I was asked (by a lot of new Oral Examiners).'

As a result of comments of this type, it was decided to include CEFR bands across the top of the scales and over the scales bands (1–6), in order to make the correspondence between the ILEC Speaking scales bands and the CEFR bands transparent.

Positive and clear phrasing of the descriptors

Some raters found the positive phrasing of the descriptors an improvement because this helps to focus on positive aspects of a candidate's performance rather than only the negative ones. However, others considered the revised phrasings too positive, which could make the differentiation between the bands (especially bands 1 and 3) more difficult. In order to address these issues it was decided to carry out some further minor refinements of the wording of the descriptors. Additionally, any difficulties in using the scales which originate from positive wording could also be addressed during rater training. Here are some feedback comments:

> '(It is) an improvement to have more positive descriptors for all bands.'

> 'I feel that the use of positive criteria is a very 'positive' development, but it clearly requires OEs [Oral Examiners] to take a new approach to the task of assessing candidate performance.'

> 'Although the philosophy is perhaps to "look for the positives" rather than the negatives, it is probably easier to assess, especially candidates who are scoring less than three marks, if one looks out for phrases like "insufficiently accurate" (as under current scheme).'

> 'It is quite a challenge to adapt to more positive criteria…'

The clarity of wording of the descriptors was emphasised by several raters:

> 'I like using these scales … because the descriptors are clearly arranged and quite easy to memorise and apply.'

> 'I found them straightforward, with no fancy, overcomplicated descriptors…'

> 'All quite clear.'

ILEC glossary and legal vocabulary

Even though the descriptors were considered clear and straightforward by many of the raters, a need for a glossary was raised, which would clearly define certain terms or differences between them (e.g. discourse markers, cohesive devices). The raters' feedback also emphasised the need to clarify, exemplify or approximately quantify certain words and phrases (e.g. *degree of control*; *a range* vs. *a wide range* (in G&V); *some* vs. *very little* vs. *very little and generally with ease* (hesitation in DM). As a result of rater feedback a glossary was created, and it was decided to address the issue of further exemplification during rater training.

The feedback has also shown that some raters had difficulties with the legal aspect of ILEC Speaking. Since Cambridge ESOL raters are language rather than subject specialists, the trial revealed some issues with assessing legal vocabulary within the context of a speaking test. Some of the issues the trial participants commented on centred on the assessment of the *appropriateness* of use of legal vocabulary by the candidates. A further issue focused on the notions of 'familiar topics' and 'range of topics', which are descriptors used in the ILEC assessment scales. The feedback from the trial participants highlighted the need for focusing on the above issues during rater training and for possible fine-tuning of the above terms in the assessment scales. One possibility to increase the raters' confidence is to enrich their current training with a vocabulary list of the legal terminology which has occurred in past Speaking tasks (as suggested by the ILEC Chair and the Speaking Subject Officer).

Finally, the raters' comments also provided answers as to why Pronunciation was generally found easiest to assess, and why Grammar & Vocabulary was the hardest. Pronunciation was found easiest because it is a more concrete, apparent and stable linguistic feature than some of the other assessment criteria. As far as assessing Grammar & Vocabulary is concerned, the main difficulty lies in getting the balance right between Grammar, on the one hand, and Vocabulary, on the other. This is not only because Grammar and Vocabulary are linguistic phenomena which could be assessed separately but also because some ILEC candidates tend to have a rich (general and legal) vocabulary and much less accurate and varied grammar. In the raters' experience, this is a frequent issue in ESP testing situations. In a future revision of the ILEC scales, the analytic scales could perhaps benefit from sub-dividing Grammar & Vocabulary into two separate criteria.

In general, the raters' feedback showed that the majority of raters found the revised scales relatively easy to apply and that they were confident about rating the assessment criteria accurately with the revised analytic scales. The feedback also proved to be a useful source of information which informed the further minor revisions of the scales, as well as rater training.

Qualitative analysis: global raters' feedback

The 'global' raters were asked to answer similar questions as the 'analytic' ones:

- How easy to apply were the revised global assessment scales (on a scale 1 'very easy' to 6 'very challenging')?
- How confident (on a scale 1 'not confident' to 6 'very confident') are you about rating accurately with the revised global assessment scales?

The findings derived from the analysis of rater feedback are encouraging since they reveal that the vast majority of the 'global' raters found the revised global assessment scales easy to apply (all raters but one chose bands 1–3, with band 2 being the most frequently chosen option) and that raters are confident about rating accurately with the revised global assessment scales (they all chose bands 4–6 indicating higher degrees of confidence).

The raters' comments reveal why the revised global assessment scales were found easy to apply:

> 'The wording is far more specific for the different scales and helps focus assessment.'

> 'Very easy. Wording is clear and simple. Criteria are succinct.'

The above findings are very positive, but in order to obtain some more useful information, frequently recurring raters' comments on the revised global assessment scales are considered below.

Firstly, rater feedback shows that several raters found it difficult to differentiate between bands 5 and 6, with the suggestion that 'Descriptors 5 and 6 need some expansion enabling interlocutors to clearly differentiate between the two.'

The comments on the positive phrasing of the descriptors are not as numerous among global raters as they are among the analytic ones. This is probably because global assessment scales' descriptors are less positively worded than the analytic scales' ones since they occasionally contain negative phrasings. Still, two raters found that the introduction of positive wording enhanced the precision of the global assessment scales, especially for accurately rating lower ability candidates.

Similarly to the analytic raters, some global raters needed pointers as to what 'familiar topics' and 'a range of topics' refer to in the context of the ILEC exam. Again, in line with analytic raters, global raters noted that certain words like 'some', 'generally' and 'occasionally' were vague. It was decided to address this issue during rater training when exemplification of specific candidate performances is provided.

The feedback of global raters showed that the vast majority found the revised global scales easy to apply and that they were confident about rating accurately with the revised scales. It also revealed some of the issues which needed addressing either through rater training or through a further revision of the scales.

Conclusion

This article focused on the *a priori* validation of the revised assessment scales for ILEC Speaking, which went live in December 2008. Two complementary sources of evidence – quantitative data and qualitative data – have both provided encouraging validity evidence on the adequacy of the revised scales. The rater feedback was a useful source for informing further minor refinements of the ILEC scales, as well as training of ILEC raters. Since the ILEC scales were revised in order to improve the assessment of Speaking, it is expected that they will have a positive impact on ILEC candidates and the universities and employers who make decisions based on their results.

References and further reading

Cambridge ESOL (2007) *International Legal English Certificate: Handbook for teachers*, Cambridge: Cambridge ESOL.

DeVelle, S, Galaczi, E (2007) *Revising the IELTS Pronunciation Scales 2007: Phase 1*, Cambridge ESOL internal report.

Galaczi, E, ffrench, A (2007) Developing revised assessment scales for Main Suite and BEC Speaking tests, *Research Notes* 30, 28–31.

Saville, N (2003) The process of test development and revision within UCLES EFL, in Weir, C and Milanovic, M, *Continuity and Innovation, Revising the Cambridge Proficiency Test in English Examination 1913-2002*, Studies in Language Testing vol. 15, Cambridge: UCLES/Cambridge University Press.

Van Moere, A (2006) Validity evidence in a university group oral test, *Language Testing*, 23, 411–440.

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

Conference reports

Cambridge ESOL staff have recently taken part in a number of events, reported on below by Fiona Barker, Angela ffrench, Kate Murray, Guy Nicholson and Jennifer Roberts.

5th Annual EALTA Conference, Athens

The 5th Annual European Association for Language Testing and Assessment conference was held in the Hellenic American Union building in Athens from 8–11 May 2008. Its theme was *Assessment Literacy in Europe and Beyond: Realities and Prospects*. The suitability and high standard of the facilities were commented on by delegates and plenary speakers, including grateful comment on the well-functioning technological aspects of giving presentations and talks. The conference was attended by delegates from a broad range of countries and institutions, including Austria, Bulgaria, Finland, the University of Lancaster, Cambridge Assessment, and primary, secondary and tertiary teachers from Greek schools and universities. There was a pleasant social element to the conference proceedings comprising a welcome reception in the roof garden of the Hellenic American Union, where the Acropolis provided a vivid and memorable backdrop to the event.

Publishers and educational institutions were represented at the conference, with stands including Trinity College, City and Guilds and Cambridge ESOL. There was interest expressed by teachers from a range of Greek educational institutions regarding Cambridge Assessment products, notably Young Learners Examinations (YLE), the Teaching Knowledge Test (TKT), First Certificate in English (FCE), and also an interest in the suite of Business English Certificates (BEC). It was noted that the Key English Test (KET) and the Preliminary English Test (PET) were also generating significant interest in the region, and of course, the First Certificate in English (FCE) was also known as a particularly useful qualification for students entering upper secondary study.

To mention two of the plenary sessions which attracted notice during the conference, in the opening keynote paper, Norman Verhelst, from CITO, the Netherlands, presented and explored themes related to the conference on the methodological considerations of vocabulary testing. Mr Verhelst developed his main premise of the need to develop a European vocabulary item bank which would allow a common data base for language testing in Europe. Another plenary presentation by Angela Hasselgreen, from the University of Bergen, Norway, was on how she is studying the reading and writing of children in school classes in order to provide CEFR-linked benchmarks of reading texts and students' writing. There was also a high standard among the concurrent sessions. For example, Dr Gary Buck, from the University of Michigan, presented some of his work on creating and evaluating a theory of test development using the CEFR, and then evaluating that against real world data. Maria Davou from Lancaster University presented a lively argument entitled *"Can Do": What can they do after all?* during which she outlined her pilot study involving learner interviews which explored how learners at different levels use formulaic language to achieve particular communicative objectives and how their performance compares with the "can-do" statements of the CEFR.

9th International Civil Aviation English Association Forum, Warsaw

In May 2008 Angela ffrench and David Booth gave a plenary session at the 9th International Civil Aviation English Association Forum *Testing for ICAO compliance: Best practice in aviation English proficiency assessment* in Warsaw. The forum, which took place over four days, allowed participants to consider testing from a number of points of view including sessions from regulatory officials, test developers and major test providers. Four organisations
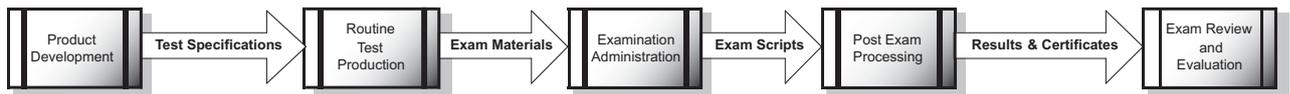
Figure 1: Cambridge ESOL's five phase model of test delivery

which offer tests for Pilots and Air Traffic controllers were also on hand to give conference participants a first hand view of tests that had been developed for this part of the aviation industry.

Angela and David's plenary session focussed on issues of validity, reliability, impact and practicality and how these factors must be considered at all stages of the process of developing and delivering tests. They focussed on a five phase model of test delivery which was developed as part of Cambridge ESOL's quality assurance procedures to conceptualise the core processes involved in developing and delivering high stakes, high quality tests to rigorous quality standards. Those phases are: product development, routine test production, exam administration, post exam processing and exam review and evaluation and are shown in Figure 1.

The plenary focussed on the last three of these phases as the earlier stages had been the topic of an earlier session on test development. David and Angela went through each phase in turn, highlighting the major issues faced at each stage, and talked about ways in which validity and reliability are maintained through the delivery and post examination processing phases. They discussed issues relating to security, examiner and administrator training, inspection criteria, policing irregular conduct, the issuing of results and awarding procedures. For each stage a checklist was provided, to enable participants to focus on the key issues in the demonstration sessions with test providers, which followed the plenary session.

Over 100 delegates from 33 countries attended the conference representing schools and training institutions as well as aviation specialists and industry representatives. The presentations and discussions were of a very high quality, lively and well-informed.

IVACS 2008 Conference, Limerick

Fiona Barker represented Cambridge ESOL at the 4th Inter-Varietal Applied Corpus Studies conference which was held at the University of Limerick, Ireland, on 13–14 June 2008. This event was attended by around 100 delegates and the conference theme was *Applying Corpus Linguistics*. Over fifty papers and three plenaries addressed a range of contexts in which corpora are being applied, many having relevance for fields associated with language assessment such as language education and publishing course books, reference materials or language tests.

Fiona Barker presented on *Using corpora for language assessment: trends and prospects* in which she outlined the development and use of corpora within language testing and assessment from the perspective of a large examination board.

The paper sessions were accompanied by many

opportunities for informal discussions over refreshments between sessions and whilst walking around the impressive parkland campus.

AILA 2008, Essen

The 15th World Congress of Applied Linguistics, organised by AILA (Association Internationale de Linguistique Appliquée), took place in the German city of Essen in the last week of August 2008. Attended by over 1400 speakers and delegates from around the world, the congress was spread over two sites – the Universität Duisburg-Essen and the Congress Center Essen (CCE).

The theme of this conference was multilingualism and much of the week's debate centred upon the importance of embracing linguistic diversity in all its manifestations. Keynote speakers at the conference included Shi-Xu from Zhejiang University, China, who spoke about the importance of recognising Eastern scholarship and research in the field of Critical Discourse Analysis, and Claire Kramsch from the University of California, Berkeley (USA), who discussed the concept of the 'third place' and its use in the teaching of language, literacy and culture.

Numerous individual papers were presented which related to various aspects of language assessment. Among these were a paper on the development of a new curriculum-based diagnostic assessment system for 8th Grade EFL learners in the Chinese school system, presented by Wu Zunmin of Beijing University and Wang Junju of Shandong University, and a presentation by Yoshihiro Omura from Kinki University, Osaka, which outlined his research into the use of accents in the revised TOEIC Listening test and its impact on Japanese test takers.

Each day of the conference concluded with various symposia, one of which centred upon the Common European Framework of Reference (CEFR) and its usefulness as a reference document for, among others, assessment bodies. Among those debating this issue were Waldemar Martyniuk (Jagiellonian University, Kraków), Sauli Takala (University of Jyväskylä, Finland) and John H.A.L de Jong, the Director of Test Development for Pearson.

Cambridge ESOL was also well represented at the conference, with members of the Research and Validation department presenting papers and chairing debates. Karen Ashton presented a paper on Asset Languages entitled *Can you read what I can read?: Case studies of German, Japanese and Urdu learners*; Neil Jones presented on *Studying impact in a new assessment framework* also in relation to Asset Languages; whilst Nick Saville (with Piet Van Avermaet from University of Ghent, Belgium) chaired a symposium on *Current perspectives on language assessment for migration and citizenship*.

34th Annual IAEA Conference, Cambridge, UK

As part of its 150th anniversary celebrations, Cambridge Assessment hosted nearly 500 assessment and education experts from around the world at the 34th annual conference of the International Association for Educational Assessment (IAEA). The event took place from 7–12 September 2008.

IAEA 2008 brought together participants from Ministries of Education, schools, universities and exam boards from more than 55 countries to exchange the latest research, ideas and experiences of each education system. The theme of the conference was *Re-interpreting Assessment: Society, Measurement and Meaning* and prestigious keynote speakers included internationally recognised experts Professor Robert J Mislevy, University of Maryland, and Professor Dylan Wiliam, Institute of Education – University of London.

The conference was opened by IAEA President Yoav Cohen and other speakers included Commissioner for Multilingualism Leonard Orban and Cambridge University Vice-Chancellor Professor Alison Richard. More than 140 parallel paper and poster presentations ran alongside the main conference. The event, which took place at West Road Concert Hall and Robinson College Cambridge, also included a Gala Dinner at King's College. Group Chief Executive of Cambridge Assessment, Simon Lebus, said: "We very much enjoyed welcoming participants to Cambridge by hosting this prestigious event. By sharing knowledge and experience in this way we are able to ensure that assessment continues to enrich lives, broaden horizons and shape futures throughout the world despite ever changing requirements."

The conference proceedings can be found at www.iaea2008.cambridgeassessment.org.uk

Winner of IELTS Masters Award 2008

Each year the IELTS partners sponsor an annual award of £1000 for the Masters level dissertation that makes the most significant contribution to the field of language testing. For the 2008 IELTS Masters Award, submissions were accepted for masters theses completed and approved in 2007. After careful consideration the Committee chose Susan Clarke as the 2008 recipient. Susan's thesis, entitled 'Investigating interlocutor input and candidate response on the IELTS Speaking test: A Systematic Functional Linguistics Approach' was completed at the Department of Linguistics, Macquarie University, Australia. Her supervisor was John Knox. The committee was particularly impressed by the breadth and detail of the thesis and felt that the methodological approach offered a significant contribution to the field. Susan Clarke's abstract appears below. The Committee was also suitably impressed by the shortlisted submission from Kerry Ryan, entitled 'Assessing the OET: The nurse's perspective.' Kerry's thesis was completed at the School of Languages and Linguistics, University of Melbourne, Australia under the supervision of Professor Tim McNamara. Kerry was awarded a 'commended' and will receive a certificate in acknowledgement of his contribution.

Susan Clarke's abstract

During the 1990s, a revision project for the IELTS spoken module was carried out to investigate ways to revise the test to improve validity and reliability. One of the main themes of the research was investigating how *interlocutor effects* can impact on candidate performance and scores candidates receive. Based on findings from the research, a revised IELTS spoken module was introduced in 2001. This study of the revised IELTS speaking test continues to investigate the relationship between interlocutor and candidate discourse and the implications for test reliability and validity.

In the study, audio recordings of two revised IELTS speaking tests from the 'IELTS Specimen Materials Pack' were transcribed and then analysed using a Systemic Functional Linguistics (SFL) based discourse analysis. The study focused on using results from the speech function and interpersonal grammatical analyses to compare *interlocutor input* and *candidate response* in the two tests to highlight consistency and variations in the discourse that could have implications for test reliability and validity. It is considered that the SFL based discourse analysis methodology of this study was able to illuminate features and patterns of the discourse useful for evaluating aspects of test reliability and validity. It is also felt that methods utilised in this study would be useful for employing in larger scale studies

The findings show that there were several types of *interlocutor input* occurring in the two tests, such as *test task framing*, *questioning prompts* and *task setting prompts* (which may include *topic-priming strategies*), as well as various types of *feedback* and interlocutor reactions to *candidate communication strategies*. The findings also showed that variations in the different types of *interlocutor input* could have implications for test validity and reliability, particularly in part 3 of the speaking test. Furthermore, findings have supported the concept that the social aspect of language needs to be considered more within the defining of the *construct* of language ability. However, as a small scale study the findings can not be broadly generalised and thus are mainly evidence of the need for further larger scale studies.

Susan Clarke will be presented with her cheque and certificate at the 31st Annual Language Testing Research Colloquium (LTRC), 17–20 March, 2009 in Denver, Colorado. For more information about the LTRC see www.iltaonline.com