DE GRUYTER MOUTON

David Newbold*

"Co-certification": a close encounter with ELF for an international examining board
"Co-certificazione": un incontro ravvicinato con inglese lingua franca per un ente certificatore internazionale

Abstract: Over the last decade ELF has become a reality in European universities, but this is not reflected in the major international language tests designed for access to higher education and for university students. In this paper I describe an experiment in "co-certification," a test of English set at level C1 of the Common European Framework of Reference (CEFR), jointly developed by Trinity College London and the University of Ca' Foscari Venice, in which an international version of the test was adapted locally to include locally relevant components and enhance validity and task authenticity. One feature of the co-certification was the introduction of an independent listening task which took the form of understanding an extract from a lecture given in English by a non-native speaker. Possible problems we had anticipated ranged from the kind of discourse features they should include, and whether or not there was a fairness issue related to accent recognition. In fact, the overall results showed no significant difference from the previous version of the exam, while the listening part had a higher pass rate than the overall oral test which involved interaction with a native speaker. As far as we know, the co-certification is the first example of an internationally recognized board engaging systematically with characteristic features of ELF communication, notably non-native phonology and intonation patterns, in a test of listening. The findings so far suggest that, for the test takers at least, this can be unproblematic and uncontroversial.

Keywords: testing, listening, higher education, examining boards, certification

Abstract: Nell'arco dell'ultimo decennio inglese lingua franca è diventata una realtà per il mondo accademico europeo, ma questo fatto non si rispecchia nelle più importanti certificazioni internazionali (quali IELTS e TOEFL) che sono

*Corresponding author: David Newbold, Dipartimento di Studi Linguistici e Culturali Comparati, Università di Venezia Ca' Foscari, Venice, Italy, E-mail: newbold@unive.it

utilizzate in Europa per dare accesso all'istruzione universitaria. In questo articolo descriverò un progetto di 'co-certificazione', un test di inglese livello C1 del Quadro europeo, sviluppato in un contesto locale in collaborazione tra Trinity College London e l'Università Ca' Foscari di Venezia per incorporare elementi 'locali', allo scopo di aumentare validità e autenticità (*task authenticity*) del test. Una delle caratteristiche peculiari del test è la presenza di una prova di ascolto indipendente come parte dell'esame orale, che consiste nell'estratto da una conferenza tenuta (in inglese) da un *non native speaker*. Le problematiche che sono state prese in considerazione nella costruzione del test comprendevano, tra le altre, aspetti del discorso e la possibilità di *test bias* collegati al (non) riconoscimento dell'accento. In realtà, i risultati sono stati simili a quelli di precedenti versioni (non ELF) dell'esame, con un tasso di successo più alto per il test di ascolto paragonato con il resto dell'esame orale che includeva l'interazione con il native speaker. La co-certificazione ci risulta essere tutt'ora l'unico progetto di un ente certificatore internazionale che si impegna in modo sistematico con un aspetto di inglese lingua franca (fonologia di parlanti non nativi) in un test d'ascolto. I risultati finora sembrano indicare che questo impegno, almeno per i candidati, non crei particolari difficoltà.

Parole chiave: testing, l'ascolto, università, enti certificatori, certificazioni

1 Testing ELF: the growing challenge

More than a decade ago, Davis and Elder warned against "moving too quickly to assess ELF before it has been properly described" (Davies and Elder 2006: 282). Other words of caution from the world of language testing focused on reliability and validity issues (Taylor 2006) and possible test bias (Elder and Harding 2008). Most of the early attempts to grapple with theoretical issues in testing ELF, and to open up new areas for test development, had come from what Jenkins and Leung call "scholars working with a critical perspective" (Jenkins and Leung 2014: 1612) in the field of World Englishes, such as Lowenberg (1993, 2002) and Canagarajah (2006). Jenkins herself used the *ELT Journal* to invite the big examining boards to engage with ELF since "it is changes in teaching which keep pace with changes in testing, and not vice versa" (Jenkins 2006: 49); an invitation which, until the project described here, has gone unheeded.

But as the phenomenal growth in the use of English as a lingua franca has become apparent, and new perspectives on ELF have developed, such as its relation to world Englishes and the implications of ELF for language teaching, so

too has the interest in engaging with assessment in the world of mainstream testing, in spite of real or imagined difficulties. Fulcher claims that "language testing research has always been very sensitive to changes in the focus of applied linguistics, particularly with regard to new uses of language" (Fulcher 2015: 58) and goes on to refer to the case of English as a Lingua Franca, citing Brown (2014) for "a masterly summary of the issues" (Fulcher 2015: 58).

To engage with assessing ELF is to grapple with a paradox. Traditionally, testers rely on standards against which they attempt to measure performance; but even though it may be "a myth that ELF eschews any kind of standards" (Jenkins and Leung 2014: 1607), it would be hard to deny that ELF delights in fluid norms, and that every interaction invents its own norms, through a process of accommodation, invention, and a range of pragmatic strategies. Unsurprisingly, the words *challenge* or *challenging* recur frequently in the titles of the scant but gradually growing literature on ELF assessment (McNamara 2012; Newbold 2014; Harding and McNamara Forthcoming). But to date there is no such thing as a "test of ELF"; the challenges are seen as future scenarios for a testing revolution which is waiting to happen.

Tests need constructs, which are abstractions of the underlying skill or skills they are attempting to measure. Since constructs will vary according to test purpose, it would be difficult to theorize a catch-all "ELF construct," except at an extremely generic level; the common element being the absence of native speaker norms. McNamara (2012) puts it thus:

> The articulation of the construct of English as a lingua franca communication is a complex task but an urgent one if assessment is to play its part in ELF education and in policies in which language competence features. Current conceptualizations of proficiency in terms of gradual approximation to the competence of the native speaker will need to be drastically revised. The consequences are likely to be as revolutionary as the advent of communicative language teaching some forty years ago. (McNamara 2012: 202)

McNamara has in mind the dynamics of ELF interaction, and the skills participants bring to bear when they co-construct meaning. Identifying these skills is only part of the challenge for would-be testers of ELF. Other, related, problems are likely to include distinguishing levels of performance, and training raters. For House, one of the first commentators to consider the problem of rater reliability, "the yardstick for measuring ELF speakers' performance should [...] be an 'expert in ELF use', [...] with comparable goals of interaction" (House 2003: 573).

Spoken interaction is at the heart of ELF usage, as it is for the communication process in any language. The Common European Framework of Reference (CEFR) exemplifies it with nine illustrative scales which cover a wider range of

communicative purposes and contexts than it uses for any of the other macro-skills (spoken production, listening, reading, and writing). In spoken interaction, the interlocutor is key; successful communication is determined by both parties, by the recipient as well as by the initiator, and this poses considerable problems for the tester. Thus attempts merely to "disentangle" the interlocutor's contribution from the performance of an individual in a paired assessment (such as those which have been used in the Cambridge ESOL suite since the 1990s), and to treat it as "a variable like any other" (McNamara and Roever 2006: 48) is to overlook the fact that meaning is co-constructed in interaction. In ELF interaction a lack of shared cultural and linguistic co-ordinates of participants is likely to make this process of co-construction less effortless than it would be for native speakers interacting with other native speakers. Pragmatic and extra-linguistic strategies are brought into the process, and these too would need to be accounted for in any "ELF" assessment.

This might be one reason why, as we noted above, there is as yet no "test of ELF" on the market, and why most testing organizations still seem reluctant to engage with it. However, given growing teacher awareness of ELF (Vettorel 2015), and the need to create tests locally to reflect local needs for ELF use, the development of tests which are "ELF aware," and which would differ from traditional tests in their approach to errors and in the kinds of texts they used, seems to offer a realistic and useful future direction for English language testing.

ELF is not just spoken interaction. Written production, spoken production, and the receptive skills of reading and listening, can all be addressed within an ELF framework, and are likely to prove to be less problematic from the test development viewpoint. This paper reports on a test with an "ELF element" which focused on a specific skill (listening to non-native speakers) and which was developed to fit the local needs of an Italian university (Ca' Foscari Venice), but which was incorporated into a well-known international certification (Integrated Skills in English, offered by Trinity College London). In a way, it is about letting ELF into a mainstream test through the back door; but far from being perceived as an intruder, for most test takers the new component had a familiar feel to it.

2 The "co-certification": a brief history

In 2005 the University of Venice (Ca' Foscari) and Trinity College London, one of the oldest examining boards for English as a foreign language, joined forces to produce a "co-certification." The idea behind it was simple. At a time in which

the Italian ministry of education was promoting and supporting the use of CEFR-related certification in schools and universities, Ca' Foscari was interested in adapting an international exam to reflect more closely the needs and profiles of its students. The "Integrated Skills in English" (ISE) suite which Trinity had just developed, and which was based on the recently published CEFR (2000), offered a broad skills-based exam which, it was felt, would have a good washback effect on language courses in the university and could be used to substitute in-house exams and, at the same time, provide students with internationally recognized certification.

A particularly attractive part of the ISE, for test takers, was the Portfolio, a series of writing tasks which test takers did in their own time in non-exam conditions, and which they discussed in the oral exam, and the presentations on topics chosen by candidates, with which the oral exam began. However, the "creative writing" task in the written exam and the portfolio of the international version was considered to be more appropriate for younger learners than for university students, as were some of the prescribed topics for the conversation phase in the oral exam.

The "co-certification" which was eventually agreed on thus saw a number of modifications from the international version. The main change was to replace the "creative writing" tasks in the Portfolio and in the controlled written exam with "critical writing," for which a Framework-like descriptor was provided. The co-certification was offered at two levels (B2 and C1), equivalent to ISE 2 and ISE 3 in the main suite; the descriptor at C1 level, which attracted the most candidates, since it could be used to substitute the high-stakes third year exam for students majoring in English, eventually settled down, after a series of revisions, to:

> Critical writing C1
> Can write a critical appraisal of a work of art, such as a novel, a film, or a collection of poetry, or present a critical overview of a cultural phenomenon, such as an institution or a lifestyle, or of an economic, historical or linguistic issue, isolating and developing the main thrust of the argument with some assurance, identifying supporting themes or typical features, and evaluating the work appropriately against the background to which it belongs.

The idea of a local version of an international exam brought with it undoubted advantages, but also raised problems. The critical writing construct made it possible for the Ca' Foscari item writers to refer to Europe-specific or Italy-specific themes with which all candidates would be familiar but which could not be used internationally. Like EFL publishers, examining boards have policies about which topics are taboo in the materials they use, for fear of causing

offence, and they cast their net wide. Sex, religion and politics are the first thematic areas to be sacrificed. But politically correct policies make for predictable and anodyne topics for discussion. Within an Italian (or European) university context, topics relating to (say) religious or political themes, usually avoided in an international test, may well be appropriate, challenging and interesting for test takers.

In the co-certification this was a potential source of tension. From the outset, the collaboration between an international examining board and a local institution had required a clear stipulation of roles: the "local" input (such as the questions for critical writing) was to be provided by Ca' Foscari; the assessment was to be carried out exclusively by Trinity College. But in one of the first sessions, most of the items provided by Ca' Foscari were rejected by Trinity College on the basis (for some of the items at least) that the subject was inappropriate. One of these, for example, invited candidates to reflect on how, in the space of one generation, Italy had changed from being a country of emigration to one with a large community of immigrants. A crisis meeting re-established the principle that, just as the local institution relinquished any claim to the assessment process, only they could decide what was appropriate in terms of content for the "local" part of the test. (For a full report of the collaboration, see Newbold 2012a.)

This episode seems worth recounting within a context of English as a Lingua Franca, since it illustrates the abandonment of monolithic standards (in this case, of what *not* to test) on the part of the examining board, to embrace local norms. But the close encounter with ELF with which this paper is concerned came later. In 2014, Trinity College decided on a major overhaul of the Integrated Skills Suite, with a view to integrating reading and writing skills more closely and to introduce an independent listening test within the oral part of the exam. This led to a radical re-designing of the structure of the exam. In the reading to writing section, among other changes, test takers would be required to synthesize material from four different text types, one of which was to include graphic material; in the oral exam, listening was assessed both in the conversation phase (as "interactive listening") and again, (as "independent listening") through a recorded excerpt from a lecture (or similar monological spoken text). While listening to the excerpt, the test taker would be invited to take notes. The recording was to be played twice; after the initial listening, the test taker should identify the theme, and after a second listening, report the main points to the oral examiner.

If the co-certification was to continue it would have to adopt the same structure as the new exam, while continuing to offer more locally appropriate content. This gave the university the opportunity to review the needs of

potential candidates for the new co-certification, and to reflect on how these had changed over the last decade. It was a decade which had seen the emphasis move from English as a subject of study to English as an essential element in a process of internationalization which ranged across all subjects, and which had fuelled an upsurge in student and teacher mobility. Like other universities across Europe, in a process documented in Doiz et al. (2013), and, for Italy, Costa and Coleman (2012), and sanctioned in a report made to the European commission in 2013,[1] English had rapidly emerged as a medium of instruction (EMI) at both undergraduate and master's levels, making major demands on human resources, and opening up a new target language use domain for language testing.

3 The independent listening task: construct and specifications

For the new co-certification we drew on a 2010 needs analysis which had been carried out at the university and which indicated the skills in English that students across the four faculties (economics, sciences, humanities, and languages) now needed as full-time students. These included reading research articles (70 %), using the Internet for research (53 %), attending lectures and seminars (21 %), and interacting with foreign students (18 %).[2] Five years down the line, with the introduction of new courses delivered entirely in English, these percentages would certainly have grown. But even for students not attending EMI programmes, and not involved in mobility – the stay-at-homers – the chances of English never intruding into their student lives were becoming remote.

In Trinity College's proposed new international version of Integrated Skills in English, the reading to writing part built on a theoretical socio-cognitive framework which had been developed by Weir (2005), Shaw and Weir (2007), and Khalifa and Weir (2009). A strong feature of the exam was to be the synthesizing of information from multiple sources (including a visual source, as noted), intended to tap into information-getting processes used by students when using the Internet for research purposes, which in the light of the needs

---

1 In June 2013, the "EU high level group on the modernisation of higher education," recommended that knowledge of English should become a pre-requisite in European higher education. See Recommendation 12 at http://europa.eu/rapid/press-release_IP-13-554_en.htm?locale = en (accessed 1 February 2017).

2 For a report, see Newbold (2012b).

analysis quoted above appeared to be a timely and appropriate update for the co-certification, too. The final, free-standing "extended writing" task, which could target a variety of genres in the international version, we decided should remain as "critical writing" for the co-certification, and retain the same construct.

It was in the speaking and listening part that we saw an interesting opportunity for change for the new co-certification. The independent listening task, a recorded text activated by the oral examiner as the final part of the oral exam, was a new departure for Trinity assessments, which had previously relied on the oral examiner to assess listening during face-to-face interaction. The "expert listener" construct hypothesized by Trinity, and based on commissioned research, draws in part on the work of Field (2012, 2013) on cognitive validity, and brings the Trinity exam more in line with other well-known international certification such as International English Language Testing System (IELTS) and Test of English as a Foreign Language (TOEFL). The monological text types include extracts from lectures or complex discussions, podcasts, and documentaries; test-taker tasks include listening for gist, taking notes, and reporting the main points to the examiner.

This type of listening test was intended to have real-world relevance, especially in an academic context. This would certainly hold true at Ca' Foscari, where one-off lectures or seminars given by visiting academics had become an almost everyday occurrence, as part of the policy for internationalization pursued by the university. If the lecturer was not an Italian speaker, then the language of the lecture would almost certainly be English.

But which English? Or rather, whose English? Most foreign academics visiting Ca' Foscari would not be native speakers of English. Yet the test specifications for the new certification refer only to native speaker accents:

> **Accents**
> May include varieties that can be processed using southern British and general American as a point of reference[3]

In this specification too, Trinity College seemed to be aligning itself with the native speaker parameters of IELTS, TOEFL and Cambridge exams. IELTS notably claims to offer ten varieties of English in its listening tests, including Welsh English, Southern US English, and South African English,[4] and invites

---

3 ISE specifications document p 47, retrieved 24 January 2017 from http://www.trinitycollege. com.
4 http://ielts-academic.com/2015/10/31/ielts-listening-english-accents/ (accessed 24 January 2017).

would-be test takers to familiarize themselves with the varieties so that they "get comfortable with each accent." In actual fact, it seems that many, perhaps most of these accents are imitations produced by a small group of actors hired by the IELTS organization.[5] The target language use domain envisaged by the creators of IELTS and TOEFL (both used primarily as university admission tests) is the native English-speaking university campus, and the model of interaction on which the tests are based non-native speaker (NNS) to native speaker (NS). This format does not take into account the fact that many lecturers in English-speaking countries are themselves non-native speakers, nor does it address the fact that international certification is increasingly requested to access EMI programmes in Europe and elsewhere. In Italy (and probably elsewhere in Europe), in any given course taught through the medium of English, and in any "one-off" lecture given by a visiting foreign academic, the most likely context is one of NNS–NNS interaction. This is the narrow sense in which we defined ELF for the purposes of the project; while Seidlhofer's wider definition of ELF as "any use of English among speakers of different first language for whom English is the communicative medium of choice, and often the only option" (Seidlhofer 2011: 7) would hold for the rest of the oral exam (which included a presentation and discussion of a topic chosen by the candidate, a collaborative task, and conversation on a topic from a set list) in which the candidate interacted with a native speaker/examiner. In this way, in the same 20-minute exam, candidates would be exposed to both native and non-native accents.

We thus proposed to modify the specifications of the independent listening task to ensure a closer fit with the target language use domain by changing two specifications of the new international certification, *topic* and *accent*, and leaving intact all the others (*speech rate, syntactic complexity, processing*, and *task outcomes*). The topic/content specifications in the international version include debates, documentaries, podcasts, which we cut, as well as extracts from a lecture, which we wanted to keep as our focus. The revised specifications for the independent listening for the co-certification thus became:

> **Topic** Information generally of a discursive nature. Could be expositional, summative, or procedural. The context would always be academic, such as an extract from a lecture or a seminar.
> **Accent** Fluent non-native speaker of English.

---

5 See, for example, the blog on quora.com: https://www.quora.com/Who-are-the-narrators-for-IELTS (accessed 24 January 2017).

4 Test development

Alderson and Bachman claim that "the assessment of listening abilities is one of the least understood, least developed and yet one of the most important areas of language testing and assessment" (Alderson and Bachman 2001: x). This is partly due to the complexity of the process of listening, and the inherent difficulty of measuring a receptive skill which by definition does not involve any observable production. But when compared to tests of reading, listening tests are harder to develop, since more decisions have to be made. With a reading test, choices need to be made about the text, such as text type, length, and degree of authenticity. The same choices need to be made for a test of listening, but since the text is mediated through performance, a range of other factors, such as speed of delivery and accent, also have to be taken into consideration. To allow for the fact that the test taker has no control over a pre-recorded listening, and so cannot ask for repetition of information, some tests may have a repeat listening built in. This is not the case with IELTS and TOEFL, in which there are multiple texts, for each of which only one listening is allowed. The new ISE exam, in contrast, offers a repeat listening of a single text of about 440 words, which takes about two and a half minutes to read.

Both approaches (discussed in Buck 2001: 170; and Ardeshir and Taylor 2008) can be justified: a single listening replicates real-life listening, in which one can usually only hear a text (such as a lecture) once, and rapid processing is needed; but a second listening can be said to compensate for all the extra-linguistic, visual or pragmatic information (gestures, PowerPoint slides) which might be available to the real-world listener attending a lecture and which are lost in a pre-recorded monologue. In addition, it might be argued that many lecturers now put their lessons on line, and in any case most students in Europe now have the means (smart phones) to record a lecture and to listen to it again in their own time. For the new Trinity College certification, the repeat listening of a single, lengthy (more than two minute) text lent itself to the dual nature of the listening task: to first report the gist (or main point) of the text, and then (after the second listening) to identify the supporting arguments.

This, then, was the structure into which the university "co-certified" version of the exam would have to fit. The texts were to be provided by the university; they would be recorded using the voices of non-native speakers who were "expert users" of English (and not actors imitating foreign accents). A team of four item writers, two researchers and two language teachers was formed with the task of producing 40 texts which would be sufficient for one

or possibly two administrations of the new certification, which would continue to be offered just once a year.

An initial choice which had to be made concerned text authenticity. In the communicative tradition which has influenced English language testing for more than three decades, "authenticity" is seen as a desirable goal for texts and tasks. In practice, however, no test can be said to be truly communicative (Buck 2001: 92), and no test task completely "authentic"; in their definition of authenticity, Bachman and Palmer (1996: 23) refer to the "degree of correspondence" between a test task and the features of a target language use task. Ideally, we would have liked to record visiting academics delivering lectures in English and then extract samples to use in the test. In practice, beyond the obvious problems of obtaining permission, quality of recordings, and a possible problem of test security (since students might end up listening to a lecture they had already attended), the two-minute samples were unlikely to offer the sort of content required for testing purposes. This needed to be, at the same time, information-dense, yet accessible, while not appealing to subject-specific knowledge which some students might have, and which would lead to construct-irrelevant variance.

The approach favoured was to look for reading texts on topics which were not in mainstream academic areas which might replicate information students were familiar with, but peripheral topics which might nonetheless crop up in an academic context, and then to rewrite them as lecture extracts, building in some redundancy and signposting typical of a good lecturer. Item writers were asked to produce sample texts, which were then discussed at a meeting with one of Trinity College's lead item writers. This provided useful feedback, and fuelled a discussion about how an "ELF text" (in our narrow definition, see above) might be different from a native speaker text at discourse level. At this point, it seems appropriate to clarify that all our item writers were native speakers of English, not through choice, but because no expert non-native item writers could be found. So texts, inevitably, were written in standard English. However, the ELF context which was our target language use domain, and the non-native speaker we had in mind, implied avoiding, as far as possible, idiomatic language, especially unilateral idiomaticity (Seidlhofer 2011: 134), and unnecessary specific cultural references to the United Kingdom.

Of the 40 texts which we wrote, ten were rejected by Trinity College on the grounds that they were probably below level (i.e. more suited to a B2 than a C1 test of listening). The remaining texts were scrutinized, and returned with a number of suggested changes. Some of these concerned style, and seemed appropriate. Others called for changes on the basis of topic suitability, reminiscent of the "local institution – global examination" tension reported in Section 2.

For example, in one text *Christmas* had been unnecessarily changed to *Birthday*. In another text, *church attendance* had been changed to the rather meaningless *religious attendance*. Both seem to have been dictated by an instinctive reaction of a global examining board anxious to avoid sensitive topics for a world market. Similarly, texts on migration and the division of Cyprus had been branded as "sensitive," whereas they were both likely to be of interest to co-certification candidates (especially, perhaps, those students following the Master's level course in Comparative International Relations). But, as in the earlier confrontation about topic suitability, these issues were clarified, and the roles of both partners in the co-certification re-asserted: Trinity did well to weed out texts which were below level, Ca' Foscari had the right to present topics which were relevant to the target language use domain, and which would not cause offense to test takers.

5 Recording the texts and reader feedback

Our initial idea was to ask colleagues from Ca' Foscari with different language backgrounds to record the texts. We began by getting a French L1 and a German L1 lecturer to record three texts; in spite of (or perhaps because of) a number of false starts, self-corrections, misplaced stresses and mispronunciations, we felt that the texts they recorded would be accessible to students and ring fairly true as non-native lecturers; which of course is what the speakers were.

However, after discussion with Trinity College, it was decided, for reasons of reliability and quality control, that the recordings would need to be made in the same studios in London along with the other (English native speaker) recordings for the international version. So Trinity, together with the recording studios, identified four competent non-native speakers of English to do the recordings. There were two men and two women; their first languages were French, Italian,[6] Spanish, and Catalan.

All four had noticeable accents. But since all four had been living in the United Kingdom for some years, they had acquired a range of features (especially vowels) which approximated to native British accents. One of the speakers occasionally used a glottal stop reminiscent of Estuary English[7] in words such as

---

6 The choice of an Italian speaker was justified since most of the EMI courses at Ca' Foscari are delivered in English by Italians, to Italians (as well as to international students).
7 For a brief description of the main features of Estuary English, see John Wells, "What is Estuary English?" retrieved 24 January 2017 at http://www.phon.ucl.ac.uk/home/estuary/estuary.pdf.

about [ə'baʊʔ] or *but* [bʌʔ]. Nonetheless, they also retained obvious features from their first languages, such as (for the Spanish speaker) the velar fricative [x] for [h], and (for the French speaker) nasalization of some vowels.

The Venice research group prepared guidelines which were sent to the speakers, and which they were invited to apply, as follows:

> You have been asked to read the text because you are a competent user of English whose mother tongue is not English. The listening texts which you produce will, we hope, be accessible to students not only because of the content, but also because they are familiar with the accents and speech habits of Europeans using English.
> In particular we would like to ask you
>
> (a) to read the texts in as natural way as possible, in your "best" English, without unnaturally exaggerating either your native speaker accent, or any English accent;
> (b) to imagine that you are speaking to an audience of about 100 students, most of whom will be Italian, a few of whom will be from other countries, none of whom will be native speakers of English;
> (c) if you wish to make any very small changes to the text (adding words like *so* or *and*) to do so;
> (d) if you make any small "errors" (e.g. of pronunciation or grammar) and self correct, please leave the correction (i.e. don't re-record the text);
> (e) if you are aware of any small "errors" (e.g. of pronunciation or grammar) only at the end of the recording, please leave them (i.e. don't re-record the text).

There were a number of noticeable errors in the thirty texts which were recorded, particularly in the frequently inappropriate division of text into manageable chunks, with readers making pauses in the middle of nominal groups, or between verb and object, or getting the wrong tonic stress, as in:

(1)   "One of the <u>group</u>'s keys to success" (instead of <u>success</u>)

(2)   "<u>Some</u> two and a half thousand years ago" (instead of <u>two and a half thousand</u>)

Errors like these for Jenkins (2000) relate to "core phonology" and might compromise intelligibility. They typically appeared in long phrases, and, to the alert listener, would probably give the impression that the speakers were reading texts which they hadn't prepared very well, thereby dispelling any illusion of "authenticity" that they were extracts from lectures. However, many lecturers in EMI programmes, anxious about their use of English, do read their lectures, or partially read them, constructing them from notes, and may as a result produce similar errors.

Word stress errors (which Jenkins consigns to non-core, i.e. unlikely to compromise comprehension) were less frequent, and usually occurred in low-frequency words (cons<u>e</u>quently, del<u>i</u>cacy, inf<u>a</u>mous, subs<u>e</u>quently, comm<u>u</u>nal, ref<u>u</u>ge).

There were also occasional false starts and self-corrections:

(3) "a group of woman... women"

(4) "and the attempt to evangelize ends here ... ends there."

(5) "the most controversial area is what to, is to what extent..."

(6) "which is now being a reality, which is now becoming a reality"

In a few cases, readers commented on, or apologized for having stumbled:

(7) "and that the future of art and, sorry, and that the future of art restoration..."

These glitches, unlike misplaced tonic stress, had the ring of genuine performance errors, which may crop up in any lesson or lecture. There were also a number of "core" phonological errors involving vowel length (such as /hɒl/for *whole* and /ˈɛrəz/for *eras*), and, on the interface between morphology and phonology, the lack of a plural marker in:

(8) "the nature of student–teacher relationship"

(9) "they are repository of history"

Finally, a few barely noticeable non-standard grammar forms appeared, in spite of the fact that readers were reading from a text with the correct forms:

(10) "*on* the front line" (instead of "*in* the front line")

(11) "it *is largely consisted* of" (instead of "it largely *consists* of")

(12) "back *in 1940s*" (instead of "in *the* 1940s")

On balance, the recordings made in London did not sound too dissimilar to the ones we had made with colleagues in Venice, although we would have preferred

to have at least one reader with a mother tongue background from northern Europe, such as a German speaker; and we would also have preferred the readers to be living and working in Europe, using ELF on a day-to-day basis, rather than to be living in the United Kingdom, and interacting with native speakers on a day-to-day basis. Nonetheless, we felt we had a corpus of texts which would be accessible to our test takers, because the performance errors, as well as the accents, would be familiar, probably more familiar than the regional native speaker accents imitated in the more polished performances of actors in the well-known international examinations. We intended to verify this hypothesis through focused post-test feedback.

First, however, we were also interested in feedback from the four readers. All four completed a form intended to elicit their impressions about the difficulty, and also the authenticity, of the texts and their readings of them. In most of their answers there was a lack of agreement. Two thought the texts seemed "similar to a real university lecture"; two did not. Two thought test takers would find them easier to understand if read by a native speaker; two did not. When asked if they thought non-native speakers would understand their readings as well as native speakers, three agreed. Three out of four agreed that the texts had been difficult to read, and gave reasons, such as sentence length and time constraints. Only two thought they had made errors typical of non-native speakers.

The only question which produced a complete consensus was "Do you think your reading of the texts sounded natural?", to which all four answered "yes". "Natural," of course, might mean different things to different people, but in our guidelines we had asked readers to be "as natural as possible," without manipulating their accent, or trying to sound more like native speakers, and they presumably had this in mind in their answers. Of course, "natural" usually has positive connotations, although for a native speaker communicating with a non-native speaker, being "natural" might imply using a range of non-transparent language and cultural references which will make communication more difficult. In an ELF context, in contrast, being "natural" in one's use of language seems like a good starting point for successful interaction, since it is on the basis of participants' real capacities, and their limitations, that the fluid norms which we referred to earlier kick in.

6 Test results and feedback

The results which we present in this section come from the first two administrations of the new co-certification (ISE 3) in 2016 and 2017. The exam comes in two parts, "reading and writing" and "speaking and listening." The reading and

writing part is allocated a fixed date, concurrently with the international ver-
sion, with which it shares most of the exam material. The date of the speaking
and listening part is chosen by the test centre (i.e. in this case, the university),
usually a month or so after the written part.

To date, 69 candidates have attempted the exam, of whom 51 (74 %) passed
the complete exam, statistically comparable with the pass rate for the first
version of the co-certification over an 11-year period.

However, unlike the earlier version, the two parts of the new exam (reading
to writing, listening and speaking) are reported separately, and each part is
certified. It is thus now possible to pass just one part of the exam, and fail the
other. The results in the listening and speaking part of the co-certification were
considerably better than the reading to writing, yielding a 93 % pass rate.
Looking more closely at these results, which report "speaking" and "listening"
separately, we find only two fails for the independent listening task,[8] making
this the easiest part of the whole exam.

Why should this be so? An obvious answer is that this was the shortest part
of the test, and the task probably more straightforward than the other, inter-
active, tasks which preceded it. In the collaborative task of interactive listening
and speaking, for example, the candidate has to adopt a persona and respond to
a cue from the examiner, perhaps by giving advice or making a suggestion
appropriate to a particular situation. Interacting with strangers in a role play
requires quite different skills from retrieving information in a recorded text; the
former might be a common activity for students of EFL in a language school, the
latter is part of the everyday reality of survival at university.

As they came out of the oral exam, all candidates were given a feedback
sheet to fill in (see the appendix) which they were asked to do while the exam
(and the examiner) were fresh in their minds. These concerned primarily their
perceptions of the difficulty of understanding of both the recorded non-native
speaker, and the live examiner, a native speaker of (British) English.

All 69 candidates completed the short (eight questions) form. Although five
of them found the contents difficult, 68 thought that the speaker spoke clearly,
and 61 "at about the right speed". Nonetheless, 11 students reported that the
speaker's accent interfered with their understanding. On closer investigation, it
turned out that five of these were referring to the native speaker of French, four
to the Catalan speaker and one each to the other two speakers. As mentioned,
although all the speakers had marked accents, it was probably the French

---

8 "Listening," recorded as a separate skill in the certification, refers only to the independent
listening task; the interactive listening which is part of spoken interaction with the live
examiner is assessed as part of the speaking skill.

speaker who had had most difficulty producing unambiguous vowel phonemes. Most students (43) found the accents familiar, which is perhaps not surprising, given the speakers' countries of origin, but at the same time 25 found that the speakers sounded "like a native speaker".

Perhaps the most interesting feedback came from question 7:

> In comparison with the accent of the examiner the speaker of the recorded listening text was:
> (a) easier (b) more difficult (c) neither easier nor more difficult
> to understand. Can you say why?

Most candidates (55%) chose (c), suggesting that the accent had not affected their understanding in any way. Thirteen students (19%) thought that the non-native speaker was easier to understand; 18 (26%) thought that the native speaker examiner they were interacting with was easier. They were asked to give reasons. Of those who found the non-native speaker easier, one said that the accent was "less strong"; another, that they were not used to talking to native speakers. More articulated responses included:

> she [the NNS] was less fluent than the examiner so she spoke more slowly and it was easier to keep up.

> I think it's psychological: if I know someone is a non native speaker I feel closer to him.

Two students who found the NNS accent difficult referred to an "unfamiliar accent" or a "foreign accent". Another wrote "I think the speaker was French and had a very strong accent", while other comments were "some words, especially place names, were not immediately understandable, due to the speaker's accent" and "his pronunciation was not very clear". One candidate wrote:

> I couldn't stay focused on what the speaker was saying because I was being distracted by the accent.

Only one candidate referred implicitly to the advantages of extra-linguistic features which the live examiner was able to use to facilitate comprehension when she said: "the speaker spoke faster, and we couldn't see the gestures and expressions".

Most candidates who reported that the NNS was neither easier nor more difficult to understand then the NS did not add comments, but those who did seem to confirm the everyday reality of ELF for young Europeans:

> I have many foreign friends around Europe, so I'm used to different accents

> I listen to both native and non-native accents regularly.

It may seem strange that for the vast majority of students the NNS recordings, with all the hesitations, errors, and self-corrections they contained, were at least as easy, if not easier, to understand than the live native speaker with whom they were interacting. However, these findings seem to reflect data which emerged from an earlier (2010) study in an EMI programme for international students in Venice (Basso 2012: 25), in which participants found it easiest to interact in English with fellow Europeans, and in which the most difficult interlocutors to understand were (in descending order of difficulty), Chinese, Japanese, and native speakers of (American) English. The simplest explanation, which we identified at the beginning of this paper, could be that European students now have considerable and frequent exposure to non-native speakers of English (especially Europeans) as part of their everyday academic and social life.

7 Conclusion: the shape of tests to come?

The project we have described is a small-scale attempt to make an international exam more relevant to local needs, in an ELF context. It can be seen as a niche product, financially not viable for Trinity College, organizationally demanding for the university, and with an uncertain future. It probably raises more questions than it provides answers, for example about text contents and the degree to which they should be manipulated to simulate authenticity, such as building in redundancy, or exophoric references to address a putative audience, or, more crucially, whether or not recordings should be left with the false starts and errors made by readers or tidied up in a more polished reading.

One of the most important questions concerns possible test bias: are some accents more likely than others to impede or facilitate comprehension? In particular, is there a bias in favour of test takers listening to native speakers of their own language using English? In the co-certification project, as noted, one of the speakers was Italian. Of the 23 test takers administered a recording made by the Italian speaker, all of them found that he spoke clearly (Q2), although only fifteen that the accent sounded familiar (Q5). Twenty-two were themselves Italian; the non-Italian was one of two students who did not respond to the question which asked whether he was familiar with the accent. The two who failed, both Italian speakers, were administered texts read by, respectively, the French and the Catalan mother tongue speaker.

The question of bias has been examined by Harding (2012). Unsurprisingly, it seems that familiarity with a particular accent may aid comprehension, although the evidence is by no means conclusive. However, as Harding notes,

the problem of bias is resolved if familiarity with a particular accent, or range of accents, is built into the test construct. In the case of the co-certification, the original "fluent non-native speaker of English" requirement was in practice restricted to Europeans when we discussed with Trinity the most appropriate accents to use. Naturally, in the interests of transparency and fairness, it is important that test takers know in advance what kind of accents they might be exposed to. For this reason, we held introductory meetings to explain to candidates that they would be listening to competent European non-native speakers in the listening task.

The fact that all the readers were mother tongue speakers of Romance languages may be seen as a shortcoming, as we have noted; as was the fact that the input texts were written by native speakers and not by non-native lecturers. But all tests, including high stakes tests, have to operate under a series of constraints and compromises, dictated by conflicting demands of validity arguments and available resources, and the co-certification is no exception to this.

Ultimately, of course, the co-certification is not a "test of ELF," nor was it ever meant to be. It is, if anything, a close encounter with the reality of ELF in European universities, driven by a quest to make an existing test more relevant to a language use domain. If a full-blown test of ELF is ever to be attempted, then the challenge for test developers will be to assess spoken interaction, and hence to create assessment grids which can capture the "one-off" nature of any ELF interaction, and measure the extent of the communicative success achieved by the participants. Such a test may be a long time in the making: test construct, test content, rater training, and the abandonment of monolithic standards are all likely to hold up this process.

But what the co-certification project does show is that the major examining boards can have a part in that process. It was through an attention to local needs, rather than to a default international standard, that Trinity College began the process of collaboration which we have described here. But the findings, which suggest that the use of non-native speakers in a listening test can be unproblematic for listeners, as well as enhancing test validity, have a significance which extends beyond Italy and Europe, to wherever English medium instruction, or everyday exposure to international English has become the norm. It seems to us that the time is now ripe for all the examining boards to engage with this phenomenon. Indeed, it may not be long before examining boards begin to include non-native speakers, not only in small-scale projects such as the one recounted in these pages, but in their mainstream tests of listening. The biggest challenge, however, will come in experimenting with new criteria for assessing production. To do so will mean to acknowledge the reality of international communication which relies on English as its lingua franca.

Appendix

Test Taker Feedback Sheet

Please answer these questions about the *recorded listening task* in the ISE 3 exam.
You should boxed(circle) the answers you select.

1 Did you find the <u>content</u> of the recorded listening   YES / NO
  text difficult?
  If so, can you say why?
  ………………………………………………………………………………
2 Do you think the speaker spoke clearly?   YES / NO
3 I think the speaker spoke
  TOO QUICKLY /TOO SLOWLY /AT ABOUT THE RIGHT
  SPEED
4 Did the speaker's accent interfere with your   YES / NO
  understanding?
5 Did the speaker sound like a native speaker of   YES / NO /DON'T KNOW
  English?
6 Are you familiar with the speaker's accent?   YES / NO
7 In comparison with the accent of the examiner the speaker of the recorded
  listening text was:
  EASIER /MORE DIFFICULT /NEITHER EASIER NOR MORE DIFFICULT to
  understand
  Can you say why?
  ………………………………………………………………………………………………
  If you have any other comment about this listening task, please write here:
  ………………………………………………………………………………………………
  ………………………………………………………………………………………………
  ………………………………………………………………………………………………

References

Alderson, Charles & Lyle Bachman. 2001. *Preface to Gary Buck, assessing listening*. Cambridge: Cambridge University Press.

Ardeshir, Geranpayeh & Linda Taylor. 2008. Examining listening: Developments and issues in assessing second language listening, *Cambridge ESOL research notes*, 32. 2–4.

Bachman, Lyle & Adrian Palmer. 1996. *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.

Basso, Nadine. 2012. Dealing with the unexpected: The use of ELF in an international academic context. In Geraldine Ludbrook & David Newbold (eds.), *English Lingua Franca: Contexts, strategies and international relations*, 21–40. Venice: Cafoscarina.

Brown, James Dean. 2014. The future of world Englishes in language testing. *Language Assessment Quarterly* 11(1). 5–26.

Buck, Gary. 2001. *Assessing listening*. Cambridge: Cambridge University Press.

Canagarajah, Suresh. 2006. Changing communicative needs, revised assessment objectives: Testing English as an international language. *Language Assessment Quarterly* 3(3). 229–242.

Costa, Francesca & James Coleman. 2012. A survey of English-medium instruction in Italian higher education. *International Journal of Bilingual Education and Bilingualism* 16(1). 3–19.

Davies, Alan & Catherine Elder. 2006. Assessing English as a Lingua Franca. *Annual Review of Applied Linguistics* 26. 282–304.

Doiz, Aintzane, David Lasagabaster & Juan Manuel Sierra (eds.). 2013. *English-medium instruction at universities*. Bristol: Multilingual Matters.

Elder, Catherine & Luke Harding. 2008. Language testing and English as an international language: Constraints and contributions. *Australian Review of Applied Linguistics* 3. 34.1–34.11

Field, John. 2012. The cognitive validity of the lecture based question in the IELTS listening paper. In Linda Taylor (ed.), *IELTS collected papers 2: Research in listening and reading assessment*, 391–453. Cambridge: Cambridge University Press.

Field, John. 2013. Cognitive validity. In Ardeshir Geranpayeh & Linda Taylor (eds.), *Examining listening: Research and Practice in assessing second language listening*. Cambridge: Cambridge University Press.

Fulcher, Glenn. 2015. *Re-examining language testing*. London: Routledge.

Harding, Luke. 2012. Accent, listening assessment and the potential for a shared-L1 advantage: A DIF perspective. *Language Testing* 29(2). 163–180.

Harding, Luke & Tim McNamara. Forthcoming. Language assessment: The challenge of ELF. In Jennifer Jenkins, Martin Dewey & Will Baker (eds.), *Routledge Handbook of English as a Lingua Franca*. London: Routledge.

House, Juliane. 2003. English as a Lingua Franca: A threat to multilingualism? *Journal of Sociolinguistics* 7(4). 556–578.

Jenkins, Jennifer. 2000. *The phonology of English as an international language*. Oxford: Oxford University Press.

Jenkins, Jennifer. 2006. The spread of English as an international language: A testing time for testers. *English Language Teaching Journal* 60(1). 42–50.

Jenkins, Jennifer & Constant Leung. 2014. English as a lingua franca. In Anthony John Kunnan (ed.), *The Companion to Language Assessment,* vol IV, 1607-1616. Chichester: John Wiley & Sons.

Khalifa, Hanan & Cyril Weir. 2009. *Examining reading: Research and practice in assessing second language learning*. Cambridge: Cambridge University Press.

Lowenberg, Peter. 1993. Issues of validity in tests of English. *World Englishes* 12(1). 95–106.

Lowenberg, Peter. 2002. Assessing language proficiency in the expanding circle. *World Englishes* 21(3). 431–435.

McNamara, Tim. 2012. English as a Lingua Franca: The challenge for language testing. *Journal of English as a Lingua Franca* 1(1). 199–202.

McNamara, Tim & Carsten Roever. 2006. *Language Testing: The social dimension*. Oxford: Blackwell.

Newbold, David. 2012a. Local institution, global examination: Working together for a 'Co-certification' In Dina Tsagari & Ildiko Csepes *Collaboration in language testing and assessment*, 127–142. Frankfurt: Peter Lang.

Newbold, David. 2012b. The role of English as a lingua franca in a university entrance test. In Geraldine Ludbrook & David Newbold (eds.), *English Lingua Franca: Contexts, strategies and international relations*, 103–110. Venice: Cafoscarina.

Newbold, David. 2014. Assessing ELF in European universities: The challenges ahead. In Paola Vettorel (ed.), *New frontiers in teaching and learning English*, 205–226. Newcastle: Cambridge Scholars.

Seidlhofer, Barbara. 2011. *Understanding English as a Lingua Franca*. Oxford: Oxford University Press.

Shaw, Stuart & Cyril Weir. 2007. *Examining writing: Research and practice in assessing second language writing*. Cambridge: Cambridge University Press.

Taylor, Linda. 2006. The changing landscape of English: Implications for language assessment. *English Language Teaching Journal* 60(1). 51–60.

Vettorel, Paola (ed). 2015. *New Frontiers in teaching and learning English*. Newcastle: Cambridge Scholars.

Weir, Cyril. 2005. *Language testing and validation*. London: Palgrave Macmillan.

Bionote

David Newbold

David Newbold is a researcher in English language and linguistics at Ca' Foscari University of Venice. He has a background in language teaching and materials development and a longstanding interest in testing. In recent years his research has focused on aspects of English as a lingua franca, ranging from attitudes towards English, the use of English by non-native writers, and, especially, the need to develop frameworks for the assessment of communicative interaction in ELF.