6 Co-Certification Revisited

Abstract In 2015 Trinity College overhauled its *Integrated Skills in English* suite, to bring it more in line with other academic certification, notably by introducing a reading to writing task based on multiple input texts, different text types, and an independent listening task. This had repercussions on the co-certification (chapter 4); if it were to continue, the University would have to adopt the same structure. The revision was seen as an opportunity to update the co-certification by introducing an “ELF element” – listening to a non-native lecturer – as the independent listening task. In this chapter we report the results after two administrations of the certification, and note that, for most candidates, the “ELF task” seemed realistic and unproblematic.

6.1 Envisaging an ELF Element for the New Co-Certification

One of the new features of the revised Integrated Skills certification was to be a free-standing pre-recorded listening task (chapter 4). In the previous version, listening had been tested only as an interactive component of the oral exam, in conversation, and in collaborative tasks, reflecting the performance-based approach of Trinity College exams. The new format marked a change in direction, aligning the Trinity exam more closely with academic certifications, and their target language domains, by acknowledging the importance of listening to monologue, and related academic skills such as note-taking and summarizing (whether orally or in writing). The rationale for this ‘expert listener’ construct hypothesized by Trinity drew in part on the work of Field (2012, 2013) into cognitive validity, and it can be seen as complementing the socio-cognitive framework (Khalifa and Weir 2009) which lies behind the new reading to writing part of the certification.

This free standing listening attracted our attention as a part of the new exam which could be easily adapted in an ‘ELF-aware’ co-certified version, and which could reflect students’ needs as ELF users in a European context. As we noted in the previous chapter, a 2010 survey had shown that more than twenty per cent of all students looking back over their experience as full-time students in the period 2007-2010 had been expected to participate in seminars, or to listen to lectures, in English, as part of their course. A decade later this percentage would surely be much higher. But it would also be true that most visiting lecturers giving these seminars or talks, to non-native speakers of English, would themselves be non-native speakers. This fact is not however captured in the specifications of the new international version of the certification, where we read:

Accents
May include varieties that can be processed using southern British and General American as a point of reference[1]

In this, too, the ISE exam follows substantially the same line as IELTS and TOEFL by offering a variety of accents, all of which, however, are native speaker accents. Yet some of these accents are less likely to be encountered on a regular basis by European students than (say) French, or German accents, in a context of English lingua franca.

We thus proposed to adapt the specifications for the listening task at C1 level, while keeping the structure and level of difficulty the same as in the international version. Two of the specifications, *topic* and *accent,* needed to be changed, in the interests of task authenticity, while all the others – *speech rate*, *syntactic complexity*, *processing* and *task outcomes* – could be left intact. Our revised specifications for the free standing listening (co-certification) 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.

We also made slight changes to the rest of the exam, (adding "education" and "higher education" to the list of possible topics of conversation in the oral, and continuing to provide the input for the final, free-standing writing task in the reading-to-writing paper).

As far as we were aware, this was the first time that non-native speaker accents were to be exclusively used in a high stakes listening test. It also offered potential research questions, such as:

– Is understanding a non-native speaker more problematic than understanding a native speaker?
– If so, why? If not, why not?

We could imagine that familiarity with a particular accent might make it more accessible to the listener, just as we could imagine that entrenched attitudes towards some accents might make them less accessible. In any case, although we did not expect to get any definitive answers to such questions, we hoped that a judiciously administered post-exam survey could elicit some interesting insights.

1 ISE specifications document, 47. URL http://www.trinitycollege.com (2017-01-24).

6.2 Test Development

A series of meetings with the Trinity College research and development team brought the project into clearer focus. Although we would have preferred to use extracts from real lectures, it would have been problematic (and extremely time consuming) to find authentic texts which had the right level of information density for the two-and-a-half minute intensive listening tasks we had in mind. In this respect, the co-certification would be no different from other certifications: we would use specially written texts following guidelines which would be drawn up by the team at Ca' Foscari, and mediated by Trinity College. Texts would be supplied by Ca' Foscari, but recorded in London in the recording studios regularly used by Trinity College, by expert non-native speakers identified by Trinity College or by the recording studios. Here too, we would have preferred to use colleagues from the University, with different mother tongues, whom we knew to be competent users of English; but we accepted that for organisational reasons, and comparability of our version with the international version, the uniform recording conditions offered by the studios were a positive feature.

The agreement, then, was to supply Trinity College with enough texts for two administrations of the certification (2016 and 2017), which would cover the two year renewable contract which had been a feature of the partnership since 2004. Firstly, however, a training session was arranged for the university team of four item writers who would produce the texts, with input from a senior item writer from Trinity. Each member of the team was invited to supply, in advance of the meeting, sample texts at levels B2 and C1. A rationale for writing was drawn up, focusing on how the texts in the co-certified version might differ from the international version, such as in the choice of topics, and how they could be made similar to real extracts from lectures, for example by (limited) use of signposting, redundancy, and hedging, as well as by focusing on a specific mode of delivery, 'procedural', 'expository' or 'summative'.

The meeting produced a consensus of opinion on some points, such as the need to limit the use of non-transparent idiomatic language, and long noun phrases more suited to written tasks, and the possibility that, given the large number of cognates with Italian words in academic texts, the B2 level texts could be more lexically dense than their counterparts in the international version. One useful activity was for each writer to read their own text aloud and note where they stumbled, and why, and to reflect on the nature of hesitations, stumblings, and self-repair in the actual delivery of a lecture. The main discussion focused, perhaps not surprisingly, on bridging the gap between a written text, and the immediacy of live oral performance.

In the end, because of the time involved in preparing and editing the texts, we agreed to limit the ELF input, at least initially, to the higher C1 level certification, and to provide Trinity College with forty texts by the

end of the summer (2015), allowing us time to edit our work, and Trinity time to process and record the texts for the spring 2016 session. Drawing topics from the humanities and the social sciences, we aimed to produce texts which would be accessible to European students, especially Italians, or international students in Europe, especially in Italy. So topics ranged from young peoples' voting habits in Europe, to ancient Greek science, to young writers in Wales (although we avoided texts which focused on traditional aspects of British culture). For each text we provided a sample gist question, which test takers had to answer after a first listening, and then the four or five main points which we expected them to be able to report after a second listening (during which they were allowed to take notes). This sequence, of course, followed the standard procedure for the international version.

Of the forty texts we wrote, ten were rejected by Trinity on the grounds that they were more suited to a B2 level test than C1. This was useful feedback: what these ten texts seemed to have in common was a more conversational style, and more self-reference, than the others, rather than an obviously simpler lexis or structure. Trinity also suggested some style and content changes to the other texts; however, some of the suggestions, especially those concerning content, seemed to be dictated by the 'default' position of the international examining board, and the need to avoid topics which referred (even superficially) to religion or politics. Thus we were invited in one text to change *Christmas* to *Birthday*, in another *church attendance* to the rather meaningless *religious attendance*, and to avoid altogether the topics of migration and the division of Cyprus (branded as "sensitive") which were the subject of two other texts. In actual fact, such topics would be unlikely to cause offence to university students, and indeed, at Ca' Foscari, could be of particular interest to students of International Relations, a heavily subscribed master's level course which regularly provided candidates for the co-certification.

The proposals made by Trinity were reminiscent of the crisis reported in 4.5, and so, as before, we had to remind our partners of the content rationale for the local version, before proceeding to the recording of the thirty mutually agreed texts for the C1 co-certification.

6.3 The Recordings

The recording studios engaged four component non-native speakers of English to read the texts. All of them had been living for some time in the UK, all of them had noticeable non-native accents, but (to the native speaker author of this book) these were in no way difficult to understand. All of them used vowels which approximated to native English vowels, especially in their use of diphthongs, while one of them had acquired a glottal stop reminiscent of Estuary English in words like *about* [ə'baʊʔ] and

but [bʌʔ], and made use of syllabic consonants, e.g. at the end of the word *written*. Nonetheless, they also all retained characteristic features of their mother tongue, such as the velar fricative /x/ (for the Spanish speaker), and nasalization of some vowels (for the French speaker).

The mother tongues were Italian, Spanish, French and Catalan. Ideally, we would have preferred a sample of accents from across Europe, including, for example, a native speaker of German (which has more native speakers than any other European language). Instead, we had only speakers of Romance languages from southern Europe. Furthermore, one of them was Italian: an accent with which, we presumed, most of our test takers would be familiar. However, despite the problem of potential bias (Harding 2012), there was a strong validity argument for including an Italian accent, precisely because this would be the most common non-native accent to which our students would be exposed, in English taught programmes for which most lecturers would be local faculty using English, or in international events held at the University. Two of the readers were men (Italian and Catalan); two were women (Castilian Spanish and French). We prepared a guide for them, which explained the background to the project, and then went on to give instructions about how to read, 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.
>
> As far as we are aware, this is the first time that a major examining board has used non-native speakers (NNS) in a test of English, and so we are keen to collect as much data as possible about the processes involved in NNS-NNS interaction, especially in the context of a test.
>
> 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 mother tongue 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).

The instructions were intended to encourage as far as possible a uniform approach to reading, as well as to create the impression of live performance. In actual fact, on listening to the recordings, we found numerous hesitations, self-corrections, and errors, in phonology, word stress, and organization of tone units. This latter was the most common error of all, with all readers making inappropriate pauses, in the middle of nominal groups or between verb and object; an error type which, perhaps more than others, indicated that the speaker was in fact reading (and was probably not very familiar with the text).

Partly because of this incorrect chunking, nuclear stress was sometimes compromised, as in:

ex 1

“One of the group’s keys to success” (instead of success).

ex 2

“Some two and a half thousand years ago” (instead of two and a half thousand).

For Jenkins (2000) this is an error of “core phonology” which risks compromising intelligibility. There were also word stress errors (for Jenkins, “non-core”, and so potentially unproblematic for the listener), for example in compound nouns, where the stress moved to the second element, as in *love story*, *travel writer*. Most word stress errors occurred with low frequency words (*consequently*, *delicacy*, *infamous*, *refuge*), while others involved selecting the wrong form of words with two pronunciations (*process* and *record*, both nouns, were articulated as if they were verbs).

Phonological errors were infrequent, and included /’kɒmræd/ for *comrade* /hɒl/ for *whole*, and /’ɔːtʃɪd/ for *orchard*, and included several mispronunciations of proper nouns, such as the names of places and people, for which some speakers used a default mother tongue pronunciation (for *Pythagoras*, *France*, *Vatican*).

There were also noticeable errors in the interface between phonology and morphology, such as the omission – or addition – of plural “s”, as in:

ex 3 “is interesting to university student” (instead of “students”)

ex 4 “banks and local governments” (instead of “government”)

ex 5 “they are out of sights and also out of mind” (instead of “sight”)

The words appear in their correct form in the text being read, of course. Example three we might suppose to be phonologically induced, in which the reader reduces the final consonant cluster in “students”; in examples four and five, however, the additional “s” may have been induced by knowledge of grammar (selecting “government” and “sight” as count nouns, the former perhaps prompted by the plural marker in “banks”, the first part of the co-ordinated noun phrase).

There are a lots of hesitations and false starts, such as:

ex 6 “history of art and ah, ah, architecture”.

ex 7 “despite the presen, despite the presence”.

which occasionally lead to apologies:

ex 8 “the future of art, sorry, the future of art restoration”.

There are also misreadings with self corrections:

ex 9 “and the attempt to evangelize ends here.... ends there”.

ex 10 “which is now being a reality, which is now becoming a reality”.

ex 11 “the most controversial area is what to, is to what extent...”.

Some of these slips have the feel of performance errors which might be made by any speaker (whether native or non-native) in a lecture. But perhaps the most interesting errors were those grammar errors which passed unnoticed by the readers, as if they had subconsciously adjusted the text to fit an internalized grammar, and which are hardly noticeable even to the most attentive listener reading simultaneously from the script:

ex 12 "*on* the front line” (instead of “*in* the front line”).

ex 13 “it *is largely consisted* of” (instead of “it largely *consists* of”).

ex 14 “back *in 1940’s*’” (instead of “in *the* 1940s”).

ex 15 “working in *the* job for which they are qualified” (instead of “*a* job”).

In the end, we felt we had a corpus of texts which, although featuring numerous hesitations, slips, and stress errors – not one of the thirty texts was completely free of these – they would nonetheless be accessible to our students, and in some cases, the performance errors would be familiar to

students from their own experience of listening to non-native lecturers.

We were also interested in feedback from the readers themselves, especially their own estimates of how ‘authentic’ the texts felt, and how they rated their readings of them. Each reader completed a feedback form (appendix 1), in which two stated that they had some experience in lecturing in English themselves. The feedback revealed considerable disagreement in their opinions. Two (including one of the former lecturers) felt that the texts seemed to be “authentic”; two felt that they were not. Three found them difficult to read, because of time constrictions and/or the lengthy sentences; two said they were aware that they had made errors “typical of non-native speakers”, which they identified as vowels, the failure to articulate the inter-dental fricative, and intonation. In fact, none of the speakers seemed (to the author of this book) to have problems with the inter-dental phonemes (which for Jenkins 2000 are “non core”). Two believed that the texts would have been easier to understand if read by native speakers; two did not. However, when asked if they thought that non-native speakers would understand them as easily as native speakers would, three were in agreement.

There was only one question which produced a unanimous response. All four readers answered “yes” to the question “Do you think your reading of the texts sounded natural?” The word “natural” had been offered in the questionnaire with no explanation, but clearly was understood to mean something different from “like a native speaker”. Clearly, too, the four readers were unanimous in their confidence that there can be a ‘naturalness’ to lingua franca communication, which transcends the ‘naturalness’ of native speakerism, and which is ‘naturally’ fluid and variable, making conscious or non-conscious use of nonstandard features, which do not necessarily compromise intelligibility but may actually promote it. We shall return to this idea when we consider the feedback from the test takers in the following section, and their comparison of native and non-native speaker intelligibility.

6.4 Test Administration and Test Taker Feedback

The data which we present in this section comes from the first two administrations of the new co-certification (ISE 3) in the spring of 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 version, 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.

In 2016 there were 29 candidates for the co-certification at ISE 3 (C1) level, a lower number than usual, perhaps because it was the first admin-

istration after a gap year; in 2017 the number grew to forty, closer to the average number of candidates for the first decade of the project, from 2005 to 2014. We thus have data for 69 candidates over a two year period.

In the new certification, the two parts (reading and writing, and speaking and listening) are certified separately, making it possible for candidates to fail one part of the exam, but to receive a certificate for the other part. A candidate passing both parts will thus receive two partial certificates, and an overarching certificate for the four skills when both parts of the exam are passed. In all, 64 candidates passed the speaking and listening part; 57 passed the reading and writing. Of the five who failed the speaking and listening, only two went below the minimum score for the free-standing listening, which therefore appears to have been the easiest section of the whole exam. This is confirmed by the number of candidates (fourteen) earning a distinction for listening (compared to nine distinctions for speaking).

Why should this be so? Firstly, we need to clarify that this was the shortest part of the test, carrying the least weight. The tasks (identifying the topic, and then, after the second listening, listing the main points) were probably more straightforward than the interactive speaking and listening tasks in the same exam, in which the candidate had to assume a persona in response to a cue from the live examiner, and take the initiative, by making suggestions, giving advice, and generally being imaginative. This sort of 'empathetic' listening is quite different from the focus on content required in the independent listening task. Trying to understand the content of lectures (in English or not) is part of the day-to-day reality of being a university student; engaging with strangers in role plays is not.

The feedback from students shed further light on the results. All 69 students completed a short, one page form with eight questions (appendix 2), all of which concerned the independent listening task. Sixty three students said that they had not found the content difficult, while sixty eight thought that the speaker spoke clearly. This almost unanimous response was in spite of the numerous errors, hesitations, and false starts which we noted above. It would seem, then, that performance imperfections do not necessarily impede communication in lingua franca, if the content is accessible. Clarity was presumably enhanced by an appropriate speed of delivery: sixty one students believed the speaker spoke "at about the right speed".

There was more variation of responses when it came to making judgements about the speakers' accents. Eleven students thought the accent had interfered with their understanding; fifty-eight did not. Of the eleven supposed comprehension problems, five were caused by the native speaker of French, and four by the Catalan; the Italian and Spanish speakers, in contrast, each caused problems in only one case. Given the marked accents of all four speakers, these results seem to bear out findings that communication can be successful in the face of noticeable or strong accents (Levis 2005, Derwing and Munro 2015).

The case of the French speaker merits a reflection. In spite of the fact that there were fewer performance lapses in her recordings, her accent was the most problematic, which could have been due to the nasal vowels which we have already mentioned. One student commented that the accent was “too thick”, another that it was “very strong” while a third reflected:

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

This is an interesting comment, because it suggests that it was not so much the intelligibility of the phonology as the listener’s own attitude, or low tolerance level to a marked accent, which errected a barrier to understanding. Nonetheless, the authors of these comments both passed this part of the test.

Forty three students said they were familiar with the accent in the recording they listened to, and a similar number (44), unsurprisingly, recorded that they did not think the speaker sounded like a native speaker of English. Perhaps the most interesting feedback of all came in the answers to question 7, which compared the accent of the non-native speaker in the recording with the accent of the live native speaker examining conducting the exam with the student. The question read

> 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.

A large majority (78%) thought that the recorded text was neither easier nor more difficult (55%) or even easier (19%) to understand than the native speaker who was with them in the examination room. Only 18 students (26%) found the non native speaker more difficult to understand than the native speaker. Given that the native speaker examiner (British, male), spoke clearly and used an accent unmarked by regional inflections, this is perhaps surprising. After all, the examiner had ways of making himself understood – such as repetition and the use of non-verbal language – which the recorded voice did not have. Again, students’ comments are illuminating; predictably, those who found the recording more difficult referred to an “unfamiliar accent”, or a “foreign accent”, but also to the fact that “we couldn’t see the gestures and expressions”. For those who found the recordings easier, reasons given included being “accustomed to recordings, not used to talking with native speakers”, “I’m more familiar with stranger (*sic*) accents”, and the self reflective: “I think it’s psychological: if I know someone is a non-native speaker I feel closer to him”.

The majority found no difference in difficulty between native and non-native speaker; some students felt the need to explain why:

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

> "I listen to both native and non-native speakers regularly."

This is a timely reminder, not only of the increased mobility of university students across Europe, but also of the fluid nature of ELF communication, which, in its widest sense includes interaction with native as well as non-native speakers, and one of the defining characteristics of which is the ability to cope with variability.

Only a few students added any additional comments on the listening task (question 8), mostly to comment on the accessibility of the accent, or to approve of the perceived rationale behind the test:

> "I consider the British accent more difficult to understand but I can imagine that the aim of this task is not to make the exam more difficult but to test our understanding of the foreign accent."

> "I found the speaker's accent really understandable. His hesitations did not influence the clarity of the speech."

> "The non-native speaker's level of English was good enough to be understood easily. As most of English speakers nowadays aren't natives I think it's a good test."

However, at least one student questioned the validity of using non-native speakers:

> "I believe it's nice to hear a non-native speaker speaking, but probably not for an English exam."

6.5 Test Results: Unproblematic and Uncontroversial?

The test results (first reported in Newbold 2017b) suggest that for most students the listening part was unproblematic, and even those students who flagged up difficulties related to the accents for the most part demonstrated sufficient understanding of the texts to pass the exam. The potential issue of fairness which Harding (2012) raises – namely that a candidate might be at unfair advantage if he or she shares the same first language as the speaker, does not seem to arise. For Harding, reporting research carried

out in Australia, and which made use of Chinese, Japanese, and Australian English accents, the evidence of unfair L1 advantage is not conclusive. But, he suggests, the problem is avoided if the accent is written into the test construct, or diluted, if the listening test uses a range of accents.

In the co-certification, the Italian accent was part of the construct of the "fluent non-native speaker"; in a meeting held for candidates before the exam, in which the structure of the new exam was explained, students were told that they would hear a European accent. Most, but not all, students recognized the Italian accent when they heard it, just as most correctly identified the French accent,[2] although one student wrote "I think the speaker was – or pretended to be – a Spanish woman, so since Spanish is quite similar to Italian, and we have a similar accent, it was really easy to understand her".

Compared with the results for the generic, international version of the independent listening task, the co-certification results are particularly interesting: the pass rate of 97% is matched by 84% for test takers of comparable age (i.e., university students) in the rest of Italy and 72% for candidates worldwide. This comparison, obviously, should be treated with caution, given the small number of candidates for the co-certification.

The new co-certification is not a 'test of ELF', nor was it meant to be, but it is certainly an 'ELF-aware' test in its attention to local needs for the writing part, and, especially, the non-native speaker recordings in the listening. But it is also a small scale project, relying on limited resources, and with an uncertain future; the need to pre-test all items, to align them with the test production procedure for the main international suite, is problematic for a certification which has a small catchment area. But whatever the future of the co-certification, this second, latest version has shown that a language test which looks beyond native speaker models is not only feasible and valid, it can also be uncontroversial for the test taker and the recognizing institution, and potentially generate good washback for the development of future teaching programmes.

More problematic is the development of a 'full-blown' test of ELF – if indeed, such a thing is possible or even desirable. The receptive skills are one thing, the productive skills quite another. If intelligibility, rather than nearness to a native speaker model, is to become the yardstick by which success is measured, then new modes of measuring will be required to assess speaking, and possibly writing.

In the next and final chapter, we shall look at possible future directions for language assessment in general, and high-stakes certification in particular, in the light of the growing need to assess competence in using English as a lingua franca.

2 This emerged in informal feedback after the exam; students were not required to guess the accent when completing the feedback form.