

Rethinking English Language Certification

New Approaches to the Assessment of English as an Academic Lingua Franca

David Newbold

7 The Shape of Certification to Come

Abstract This final chapter offers a reflection on possible future directions for English language certification. The major problem to solve (or to attempt to solve) seems to be not so much *what* to assess (fifteen years of ELF research have offered lots of insights into this) as *how* to do it. After discussing a series of problems related to rater rubrics, and the notion of error, we consider the format that future ELF aware certification may take, concurring with Harding and MacNamara that an add-on ELF component currently seems the most practical way of incorporating an ELF element into mainstream certification. We conclude that the development of ELF certification is likely to be slow and painstaking, it may combine local and global elements, but in the long term it is inevitable, since the demand for valid and reliable certification of competences in the use of the world's lingua franca is destined to grow, perhaps for many more years.

7.1 The Need for New Approaches

In the new world order envisaged in *The Shape of things to come* Wells does not make it clear whether citizens need to certify their level of Basic English to access the jobs market (which seems to be controlled by their “educational guardians”), or any other position in society which will require them to use the lingua franca. Perhaps it is no longer necessary to do so; Basic English appears to be an easily acquired lingua franca in the new “body of mankind” which has become “one single organism” (444). In Wells’s brave new world of well-behaved citizens the acquisition of the world language has in fact proved most difficult for the native speakers of English, who require special training “to restrict themselves to the forms and words” needed for successful lingua franca communication. This is an interesting reflection on the role of the native speaker in lingua franca interaction, and it raises questions not only about *what* ELF ‘certification’, if it is ever to exist, should attempt to certify, but *who* should be taking the test. The native speaker vs non-native speaker is just one of a number of dichotomies that the test developer, or examining board, will need to address in the preparation of any test of ELF.

After a lifetime in language testing, in a “State of the Art” interview, James Dean Brown (Salmani Nodoushan 2015, 139) argues that there are (at least) fourteen approaches to testing English language proficiency (“whatever that may be”), six of which are “top down” and eight of which are “bottom up”. Only one of these (the first top down approach) is rooted in a native speaker

model approach. The others (top down) he labels as “truth-in-advertising”, “multiple world Englishes”, “English as a lingua franca”, “global standard English”, and “functional approaches”, while the bottom up approaches include “the effective communicator”, “scope of proficiency”, “scale of range”, “intelligibility”, “resourcefulness”, “symbolic competence”, “intercultural communication skills”, and “performative ability”.

The list is useful, not because it is exhaustive, but because it is long. Apart from the “top down” approaches which might loosely correspond to the agendas of language planners and curriculum designers, the list of “bottom up” approaches suggests a wide range of user-focused competences, most of which could be of interest to an ELF test designer. For example, “effective communicator” suggests developing tasks which have a measurable outcome in terms of successful communication, “intelligibility” suggests a focus on perception rather than (native speaker like) production, and “resourcefulness” could include a raft of strategies (such as paraphrasing, self-repair, and requests for clarification) which have been described in the ELF literature and which tend to facilitate successful outcomes in ELF interaction.

In an early (2006) publication Elder and Davies outlined a number of tasks which might feature in a test of ELF, such as avoiding native speaker-centric lexis, listening to non-native speakers, and participating in a role play with a speaker from a different lingua-cultural background. The first of these seems conceptually problematic, since it requires raters to look for an absence of something, and evaluate it positively, while they are asked to overlook non-standard features which do appear but which do not impair communication. The second (listening to non-native speakers) was the focus of the project described in chapter 6; the third, task-based interaction, was already being used in the form of paired assessment, such as the speaking tasks in the Cambridge exams. In any case, the authors themselves conclude their proposal by warning “against moving too quickly to assess ELF before it has been properly described”.

A more structured approach has been put forward and experimented by Harding (2015), who took an information gap activity, carried out by two participants from different lingua-cultural backgrounds, one of whom was the “information provider”, the other the “information receiver”. Ten raters were invited to observe ELF features relating to accommodation, negotiation of meaning, and discourse maintenance. Although they agreed broadly on which of the two participants performed better, Harding concludes that it was not clear how the holistic rating scale they were using was actually being interpreted.

This kind of information gap task has been familiar since the communicative language testing revolution announced by Morrow (1979), and it brings with it a series of rater-related problems which, as Harding acknowledges, will need to be addressed if such a thing as an ELF test is to be developed. If the main focus of ELF assessment is to be spoken interaction – the co-

construction of meaning between two or more participants who have different mother tongues – then the major challenge for examining boards will be to develop reliable rating scales to evaluate this interaction. Of course, every area of language activity can be undertaken in an ELF context, and an ELF assessment could thus be extended to include listening, reading, writing, and spoken (monologic) production, all of which could be relevant to an assessment for academic purposes. But it is not the *what* to test which is the primary problem for the ELF-aware test developer; this should be directly linked to the target language use domain envisaged, which (for English in academic contexts) emerges clearly in needs analyses such as the one we described in chapter 5. Rather, the problem is *how* to assess the one-off, unique, never-to-be-repeated performance moment of any ELF interaction through an assessment tool (such as a holistic grid) which is nonetheless fixed, stable, and (ideally) potentially reliable.

7.2 Re-Thinking Rating

Paran and Sercu (2010) analyse four aspects of language education which they consider to be “untestable”, yet worthy of testing: literature and literary competence, learner autonomy, CLIL and inter-cultural competence. To these could be added ELF, but there is a difference. Paran and Sercu (2010) take a process, learning-based approach to strands which have come to occupy important positions in school curricula, and for which evidence of acquisition and/or progress would be useful. ELF, as ELF researchers are at pains to point out, is use of English beyond a learning context (see chapter 5). The strategies that ELF users bring to bear in interaction may of course be fostered in language classrooms, by ‘ELF aware’ teachers, but they may also develop in users independently of any formal learning process. Indeed the familiar (and perhaps cosy) environment of the classroom is at odds with the unpredictable nature of ELF interaction, and any test of ELF interaction would need to guarantee a degree of unpredictability in the task it attempts to assess. This is just one aspect of the “rating problem”, and it concerns the identity of the participants, as well as the nature of the task. We turn now to consider briefly some of the areas in which an examining board engaging with ELF interaction would need to rethink existing communicative tests.

7.2.1 The Identity of Participants

By definition, participants in any ELF interaction do not share the same native language. If the paired assessment model is to be used, this is likely to cause logistic problems for examining boards, especially if a traditional

format is used, i.e. with the test takers physically in the same room together; one of the participants would have to be brought in from a different lingua-cultural community to the local one. In relatively stable monolingual communities which are still the norm in Europe, this would be difficult.

Of course, existing communicative-type paired assessments could similarly be criticized when they somewhat unnaturally invite candidates who have the same mother tongue to converse in English, which may cause unexpected comprehension problems for the native-speaker examiner.¹ But interactive tasks in a traditional communicative test are primarily designed to elicit appropriate language, and not to sample a range of pragmatic strategies which enable ELF communication to take place.

Alternative formats could include setting up a video interaction using the Internet – but this would require negotiating criteria for matching test takers, and bring into play a number of variables related to the use of technology – or to revert to one-to-one interaction, between examiner (or facilitator) and candidate, in which the examiner is herself part of the meaning-construction process. This takes us to the next aspect of the problem, the need for empathetic raters.

7.2.2 The Empathy of Raters

More than a decade ago, House (2003, 573) suggested that

the yardstick for measuring ELF speakers' performance should [...] be an 'expert in ELF use', a stable multilingual speaker under comparable socio-cultural and historical conditions of use, and with comparable goals for interaction.

The monoglot native speaker, it is implied, would be at a disadvantage for assessing ELF interaction. This would probably be a consensus view for most ELF researchers and 'ELF-aware' teachers today, although Canagarajah (2007, 927) points out that there is "nothing stable about the multilingual speaker". This is not to assert that a trained native speaker rater would be unable to make judgements about the effectiveness of strategies used by test takers, but by referring to "comparable goals for interaction" House seems to be alluding to the collaborative nature of meaning making;

¹ The author was once told the following anecdote by an examiner who had attempted to make a paired assessment in Naples. The two candidates chatted away comfortably in 'Neapolitan' English, fluently, respecting time limits and turn taking, clearly understanding each other, and thereby achieving a degree of communicative success, but the examiner understood little or nothing of what was being said, and consequently found it difficult to rate the candidates' performance.

whoever is doing the rating also needs to be part of this process, whether she is interacting directly (in an interview) with the test taker, or simply observing performance. In short, raters need to be empathetic participants and/or listeners. This is at odds, of course, with a traditional view of an examiner as detached, impartial, and objective.

Other queries also arise about the identity of raters. What if they were to share the same mother tongue as one of the test takers? Would that compromise fairness and impinge on test validity? Examining boards would need to draw up a recruitment and training policy for raters, define the competences required, develop scoring rubrics, and implement a validation process to ensure a degree of inter rater-reliability. The starting point could be the trialing of a holistic grid, such as the one suggested by Harding (2015).

7.2.3 The Need for Evidence

Harding tentatively suggests a check list of strategies for a holistic rubric organized under the principle competence areas of “accommodation”, “negotiation”, and “maintaining smooth interaction”. The first of these includes making oneself intelligible and adjusting to the interlocutor’s speech or style. “Negotiation” lists four well documented ELF strategies, clarification, self-repair, repetition, and paraphrasing; the final area of discourse management includes turn-taking and politeness.

This is a good start, but other strategies could be added. In many ELF interactions, progress is anything but smooth; communicative success, if it is achieved, is achieved against the odds (Newbold 2015a, 214), and it may involve such diverse ploys as explicit or implicit requests for help, the use of body language, or specific references to shared cultural resources. In short, there can be a messiness to the negotiation of meaning which should not be mistaken for lack of competence(s), but an attempt to harness all possible resources.

On the other hand, interaction may indeed be ‘smooth’; so seamless, in fact, that there is nothing to observe in the way of self repair, repetition, paraphrasing, or any other criterion which may be taken from a taxonomy of pragmatic strategies for ELF communication. What happens when test takers converse with no apparent need to resort to accommodation or repair strategies? How would communicative success be measured in these cases, with little or no evidence of ELF strategies being employed? This is an eventuality which test developers would need to anticipate. In Harding’s information gap activity, participants were presumably chosen because of their very different lingua-cultural backgrounds - one a native speaker of Thai, the other of Spanish. The lingua-cultural gap may close when both or all participants come from the same geopolitical area, such as the European

Union which has been the main focus of this book, and this may make communication easier. In an international test of ELF which included spoken interaction, how would an examining board match test takers?

This begs another question about traditional levels of language competence. Like the lingua-cultural gap, a mismatch of levels of fluency (however we might define this) is likely to cause more strain for participants, and as a result more opportunity for resorting to ELF strategies for both participants; could 'mismatch' be a criterion for pairing test takers? In a test of ELF (if it is ever to exist) should test takers be required to supply information about their presumed level on a well-known scale (such as the Common European Framework) when they enrol for the exam?

7.2.4 The Problem of Levels

In a criterion referenced, task-based, communicative test success is ideally measured in terms of outcomes. To take a simple example from real life: if an information receiver R is able to get to the railway station on the basis of directions provided by information provider P, then the interaction can be considered as having a successful outcome. From this perspective, a 'purely' communicative test can have only two possible outcomes: success or failure. Indicating a degree of success - or even more grotesquely, a degree of failure - would be difficult and irrelevant.

In a hypothetical rating rubric for ELF interaction, even if we are to focus on evidence of ELF strategies which facilitate a successful outcome, rather than the outcome itself, there will be a problem of identifying levels. Luoma (2004, 80) suggests that the norm (to guarantee a degree of inter-rater reliability and therefore consistent results) is from four to six levels of performance, but she is referring to both holistic and analytic grids in traditional tests based on a standard model of the language. When it comes to the 'untestable' areas of language ability, the would-be ELF tester might find Sercu's (2010, 29) discussion of three possible levels for measuring intercultural competence (basic, intermediate and full) relevant, although not transferable in any acritical way, to the ELF context.

In short, the problems of rating ELF interaction seem insurmountable. Wherever we focus our attention on rubrics or on levels of performance, on raters or on the test takers themselves, we find questions but no obvious answers. However, so far we have been considering a hypothetical stand-alone test of ELF; a test which *only* measures a yet to be defined ELF construct. The prospective changes when we think in terms of ELF assessment as an add-on element to a more traditional (Framework related) test. This is the conclusion reached by Harding and McNamara (2017):

It seems more likely that ELF is at least in the short term not going to

replace more static proficiency constructs, but rather would function as an add-on in contexts of language assessment where ELF competences are expected to come into play (which may be all situations).

We shall return to this idea of the ‘add on’ in the section on test formats below. First, however, we need to consider another rater-related dichotomy, about which examining boards attempting to assess ELF would need to issue guidelines, and which, for many teachers preparing students for tests would be crucial: the notion of ‘error’ in international communication.

7.3 Rethinking Errors

The notion of error in language teaching and testing is traditionally, and often unquestioningly, equated to a deviance from native-speaker norms. References to native speakers may be built into rating scales, and there are numerous references in the CEFR to native speakers. Notoriously, concepts such as not “unintentionally amusing or irritating” or “keeping up with native speakers”² are built into the scales for spoken interaction, suggesting that native speaker-like proficiency, and indeed, native speaker-like behaviour, should be the wider target language domain as a testing objective. However, it should be remembered that the CEFR was developed not with a single language (English) in mind, but as a functional description which could be used for all European languages, and it was never intended to describe levels of competence for a lingua franca.

The case of English is doubly exceptional: not only because of its use as a lingua franca, but also because of the emerging paradigm of world Englishes, which embraces variability in all aspects of language use (phonology, syntax, lexis, discourse management, etc.). English does not have one ‘standard’ version, but many native and second language speaker norms, and a growing awareness of this variability, and the choices to be made about which English to teach – and consequently test – have become a major subject for discussion in training courses and publications for the ELT (English Language Teaching) profession. (Newbold 2017a).

For would-be language certifiers, one possible approach to error would be to discard *any* deviation in production from *any* native speaker norm, at least if these deviations were not considered to undermine comprehension; but this would be problematic, not only because of the subjective judgements involved (on the part of the rater, who may not always be sure

² CEFR Descriptors for Level B2 include:

conversation: Can sustain relationships with native speakers without unintentionally amusing or irritating them.

informal discussion: Can keep up with an animated discussion between native speakers.

that understanding has taken place), but also because an initial lack of comprehension (and awareness of such) will often be the trigger for those ELF repair strategies which raters would be looking for, and which are a necessary part of the co-construction of meaning. In any case, it would be advisable for examining boards to establish a public policy on errors, including a definition of error, and the part played by errors (if any) in the assessment process. We shall briefly consider how these might vary from one aspect of language use to another; these considerations could be addressed in a policy document on errors which could be incorporated into test specifications.

7.3.1 Phonology

Outlining a ‘lingua franca’ approach to testing pronunciation, Sewell (2017, 238), writing from Hong Kong, suggests that the challenge “lies in navigating the local/global polarity”. This observation seems particularly pertinent in the light of the research by Basso (chapter 5) who found that, for the majority of European students in an international campus in Venice, the most difficult accents to understand were North American (i.e., native speakers of English) and South East Asian (speakers whose mother tongues were Chinese and Japanese). Although this research did not have pronunciation as its main focus, we might speculate that the comprehension problems are linked to two concepts which Sewell refers to: “intelligibility” and “functional load”. The first of these is taken to mean the quantity of understandable speech; the second, the extent to which specific phonemes are used contrastively (an indication of which can be given by the number of minimal pairs a phoneme contrast is required to keep apart). In the case of the north American speech, unfamiliarity with accent, coupled with speed of delivery and lexical load, could have made understanding difficult, whereas in the case of the Japanese and Chinese speakers problems of perception may have been more exquisitely phonological.

Another interesting factor Sewell refers to (243) is the possibility that “written language and worldwide literacy operate as centripetal forces on pronunciation”. This also seems relevant in the context of English as an academic language. The notable mismatch between spelling and pronunciation, as well as the rhythms of stress timed language, which are features of native speaker English, are often eroded in lingua franca interaction. Stress timing is not part of Jenkins’ “core phonology”, and it is easy to see why: careful syllable-timed speech makes perception less, not more, difficult, and it may be adopted as an accommodation strategy.

7.3.2 Syntax and Morphology

Language testers, and especially examining boards delivering high stakes tests, have come to be seen as guardians of standards, and it would perhaps not be unfair to assert that this role has been promoted by a testing culture which has developed around the notion of errors, and especially grammar errors. Generations of test takers have been tricked into selecting erroneous forms in an array of objective test types, from multiple choice to cloze, from true/false to sentence rewriting. One reason is that such tests (or parts of tests) are easy to create and easier to score. But they belong to the written domain. Unsolicited grammar errors in spoken production, and in spoken interaction, may be captured in analytic scoring grids, but (as we saw in chapter 2) grammatical accuracy is likely to be seen as just one of several assessment criteria, and probably not the most important one.

Grammar errors do not usually compromise intelligibility, but they are harshly viewed by the academic community. In the 2006 study by Mollin, in a survey of 435 European academics, 95% responded that omission of the third person “s” (in the example sentence: “Do you know where she live?”) was unacceptable, making it the most despised error of all. Yet, taking the long term view, it is arguably a fossil structure, the only morphological inflection left of a once highly inflected verb system, and doomed to disappear.

In a lingua franca context the focus changes again, since grammar may be manipulated to enhance meaning. A sentence such as

ex 1

I will go to Rome if you will come with me.

mirrors structures in many other languages, while emphasis may be achieved by left dislocation - Mauranen (2010) provides a number of examples of this from the ELFA corpus - such as

ex 2

This problem, I'll come back to it in a minute.

Reduplication, which has only a limited use in standard English, but is a feature of some world varieties, as well as other languages (including Italian), might also be used for emphasis, instead of an intensifier:

ex 3

It's a small small problem.

These are just a few instances of deviation from a standard which could be used to inform a new approach to errors in a test of ELF.

7.3.3 Lexis

We referred to lexical creativity in chapter 5 as a major focus of ELF research and an effective strategy for creating meaning. In fact, much lexical creativity takes place at the interface of grammar and lexis, through the manipulation of morphology. Like many ELF strategies, it cuts across the divide between native and non-native speakers. For example, the word *involvable* was recently used by an Italian post doc student in conversation with the author, to refer to a motivating classroom activity, in which everyone could take part:

ex 4

It's a very involvable activity.

A Google search³ asks the information seeker if they didn't in fact mean *insolvable*, and when the offer is turned down, returns a count of just 2,230 hits for *involvable*, some of which are clearly in a non-English context. The meaning that was inferred was both "motivating" and "not difficult to participate in"; which was confirmed by the person who had coined it. It seemed to the author (and still seems) not so much an error as an economic and elegant term for a useful concept.

This kind of creativity shows considerable language awareness. It demonstrates knowledge of lexis (*involve*) and knowledge of word formation processes (affixation). There is nothing in the example to indicate that it is a non-standard form used by a non-native speaker, rather than a term invented by a native speaker to plug a gap. In a context of ELF assessment, it would be an observable strategy promoting communication.

More problematic, from an assessment point of view, is to sanction lexical choices which seem to hinder communication, as in the case of "unilateral idiomaticity" (Seidlhofer 2011, 134). Communication breaks down when a word or words (whether used idiomatically or not) are not familiar to the interlocutor, but it is at this moment that ELF accommodation strategies can kick in, and the channel of communication be re-opened. An ELF assessment grid, rather than simply noting errors and breakdowns, should be observing if and how these are transformed into opportunities for co-operative meaning making.

3 Search made on 2017-07-03.

7.4 Rethinking Test Formats

So far in this chapter we have been discussing spoken interaction, which lies at the heart of ELF usage, and which probably poses most challenges for any hypothetical “test of ELF”. But a test of ELF, or a more realistic ELF aware test, may include other skills, and may embrace many formats.

To start with, ELF may be manifested in different ways, and allow for more or less variability, and consequently require a more or less rigid test format. Basic English, with which we began this chapter, is an example of a controlled natural language (Kuhn 2014), with a prescribed word list and specific rules for meaning creation (through the combination of words in the list). Similarly, there are areas of professional use of English today in international contexts, such as so-called “Seaspeak”, for maritime communication, and “Airspeak”, for air traffic controllers and pilots, the main aim of which is the avoidance of ambiguity. In these contexts language needs to be carefully regulated and assessed, not least because human lives daily depend on the successful communication in English between non-native, and also native, speakers.

This is not the kind of lingua franca use we have in mind for certifying competences in academic English. Rather, beyond the challenge posed by the assessment of spoken interaction, future certifications may not look very different from existing certifications described in some detail in this book, and extend, as they always have done, to reading, writing, and spoken (monologic) production. They could, however, be made ELF aware in the choice of texts for reading and listening components, and in their assessments of written and spoken production.

Reading components, for example, could include texts by non-native writers. These could be literary, academic, formal or informal, depending on the underlying construct for reading skills; they could be carefully sourced or specially written, published or unpublished, from a “world English” variety, or from the “expanding circle”, to use Kachru’s well known (1985) model. In a one topic, multi-text approach which has been adopted in the new international version of the Trinity College *Integrated Skills in English* suite (chapter 6), one text could be by a non-native writer. Similarly, tests of listening could incorporate non-native voices, such as the extract from a lecture in the updated co-certification (chapter 6), but also genuine short ELF interactions which might be relevant to the overall aims of the test. These latter might not be very different from the “extracts from life on campus” which are a feature of the TOEFL test, with the difference that both participants would be non-native users of English.

In the productive skills the problem of native speaker norm returns, and with it, the problem of rating. In writing, especially formal writing of an academic nature, it is harder to justify deviations from native speaker norms. But the advent of computerized testing of writing may alleviate

these, since test takers could switch on spell and grammar checks to reduce low-level formal inaccuracies. After all, in most non-testing contexts of writing, writers would normally be able to make use of tools (such as dictionaries, grammars, and style guides) to help them; it is thus what they can do *with* such tools, rather than without them, which should be of greater interest for assessment purposes and provide most information for the test user. This would allow an empathetic ELF user/rater to shift her focus to higher level aspects of discourse management, such as structural cohesion and clarity of argument; an objective which might also be within the range of some future (non-native) machine marking system.

The assessment of spoken production seems to us to be particularly important in the context of ELF. More than ever, English, or rather, ELF, plays a role in the professional lives of non-native speakers, and universities can provide a training ground for future professionals who may have to give reports in meetings or address audiences, by offering opportunities to hone their presentation skills. With the reform of the university system in Italy, and the introduction of the *laurea magistrale*, student presentations have become a staple feature of many courses, and may be used as part of a continuous assessment process. Certification provides an excellent opportunity for an ELF-type presentation, of a topic chosen by the candidate, and addressed to a putative non-native speaker audience. The skills which might feature on a check list for raters could include, to give just a few examples, voice control (speed, volume, use of pauses), repair strategies, and discourse management features such as signposting. Newbold (2015a, 219) suggests that these could be usefully assigned to a higher order of categorization for rating purposes: *control* (of voice, lexis, etc), *range* (of repair strategies, etc.) and *alignment* (or ways in which the speaker connects to the audience).

Of course, all of these skills would be part of the stock-in-trade of a competent native speaker, but none of them belong exclusively to the native speaker domain; they cut across the language divide, and there would thus be no point, indeed, no meaning, in including “native speaker like” behaviour on the assessment check list. Rather, a hypothetical future test, or certification, of “speaking to an international audience” could be aimed at both native and non-native speakers; and the native speakers (as Wells predicted) might find it more difficult to score highly on such a test than their battle-hardened, ELF-using, non-native counterparts.

Such a test could be a free standing ‘certification’ of spoken production in its own right, of interest to prospective employers in an international jobs market. We have already referred to a test of spoken interaction as a possible ‘add-on’ component to an otherwise conventional certification. At this point it seems that a modular approach to certifying ELF competences, whichever skill(s) we are interested in, is likely to be the most practical, for at least three reasons. Firstly, it would keep ELF and

non ELF approaches to rating separate, allowing a generic component to be linked to a framework such as the CEFR. Secondly, it recognizes that some skills might relate to a 'specialized' ELF construct (academic writing, interacting with patients in a healthcare context, etc) and could also be offered as 'add-ons' or stand alone tests. Thirdly, a modular approach would allow local versions of a test, on the global/local interface, so that, for example, a European test of ELF for academic purposes might include both local and global elements; it might offer 'local' contents but look for global ELF strategies in the test taker.

Versioning certifications obviously has a cost for examining boards, but allows them to reach more candidates. This approach has long been adopted by IELTS (chapter 2), which offers an "Academic" and a "General Training" version of the exam, in which the listening and speaking parts are the same for all test takers, while reading and writing are different. A recent switch to a more modular approach has been made by Trinity College in the ISE exam, (chapter 6), so that the reading and writing exam, which is done on a different day from the listening and speaking, is now certified separately, making it possible for a test taker to have a certificate for just one part of the exam, and consequently, in the case of failure of one part of the exam, to re-sit only that part, with a subsequent reduction in the fee.

The greater flexibility offered by a modular approach would also allow test users to make informed choices about which elements would supply the information they were interested in, by adapting those modules most relevant to a local context, and in this way, mirroring the fluid nature of ELF itself. The modular approach, one could maintain, is more 'ELF aware' than a 'one size fits all' certification. The test format, of course, is not the test construct, but it could grow naturally out of it.

7.5 Conclusion: Evolution, not Revolution

In this book we have tried to show that, although the certification industry has grown enormously over the last two decades, it has still to address the underlying cause of that expansion: the unprecedented growth of a genuinely global lingua franca, and the need for reliable independent measurements of what ELF users can do with it. We have noted the aspirations and also the shortcomings of existing tests, and we have presented a small local project of an 'ELF aware' certification, only to return, in this chapter, to the fundamental problem of rating, to which we have offered tentative approaches but no real solutions.

If the primary focus of the book had been "assessing ELF" we might have managed to write most of it without referring to any certifications. We would have discussed a range of more alternative approaches, such as

peer assessment, self assessment through reflective feedback, or continuous observation-based assessment such as the project described in Tsagari and Kouvdou (forthcoming). Assessment which involves reflection on the part of all participants is likely to be richly formative as well as informative, and to fulfil an essential role in any ELF-aware language programme.

But certifications are here to stay, and they are important. They have a function in today's globally mobile society because they provide independent assessments which (as we have seen) prioritize fairness, reliability and security. They drive a large sector of the English language teaching and publications market, as well as providing a high stakes gate-keeping function for immigration services, potential employers, and higher education. In short, they have a controlling function which is more apparent than ever before (and which has more than a faint analogy with the controlled global society portrayed in *The Shape of Things to Come*).

This is why examining boards need to reconsider constructs, to invest in ELF assessment research, to be able to stay in touch with emerging new language needs. In the long term, to do so would make commercial sense, and assert an ethical role which not-for-profit organisations typically subscribe to. If they do not, then other locally-based organizations may emerge to do so. Indeed, a strong case could be made for locally developed tests which combine specific professional, vocational or academic content with specific international settings, such as a university access test for European University students.⁴

So far, the major examining boards seem to have shown little interest in engaging with the phenomenon of ELF, beyond the co-certification project described in these pages, although it is to be presumed that they are aware of the issues involved. Whatever the future of English language certification, formal ELF assessment is likely to come about slowly, piecemeal, perhaps through more small-scale projects, and assisted by developments in technology. When communicative language teaching was being theorized, in the late nineteen seventies, Keith Morrow (1979, 156) concluded his seminal article "Communicative Language Testing: Revolution or Evolution?" by speculating that "there is some blood to be spilt yet". Four decades later, there is not yet much evidence of blood having been spilt in the testing profession (as far as the author is aware), but rather an ongoing consensus which has evolved out of different assessment traditions and, more recently, the CEFR. The time is now ripe to move a bit further along the communicative route. As if to underline the urgency, an email alert has just arrived on the author's desktop which reads "Once You Go Global,

4 For the form such a test might take, see Newbold 2015b.

There Is No Coming Back”.⁵ To engage with ELF in a language certification also means going global, and to follow English along its evolutionary path as a hybrid, many-faceted tool of communication, and from which there is indeed no going back.

5 On closer inspection, it turns out to be an invitation to a webinar organized by IATEFL, the International Association of Teachers of English as Foreign Language.

Appendix 1

Feedback sheet for readers

Please write brief answers or circle the appropriate responses.

- 1 What is your mother tongue?
- 2 Have you lectured or given a lesson in English before? YES / NO
- 3 Did the texts seem to be 'authentic' (i.e. similar to a real university lecture)? YES / NO
- 4 If not, can you briefly say why not?
.....
- 5 Did you find them difficult to read? YES / NO
- 6 If yes, can you briefly say why?
.....
- 7 Do you think your reading of the texts sounded (reasonably) natural? YES / NO
- 8 If not, can you briefly say why not?
.....
- 9 Are you aware of having made any 'errors' typical of non-native speakers? YES / NO
- 10 If so, which?
.....
- 11 Do you think non-native speakers would find it easier to understand these texts if they were read by a native speaker? YES / NO
- 12 Do you think non-native speakers will find your readings as easy to understand as a native speaker would? YES / NO
- 13 Do you think you would have used simpler language if you had given the lecture? YES / NO

Thank you for providing this feedback!

Appendix 2

Post-exam feedback sheet for test-takers

Please answer these questions about the recorded listening task. This won't take long: Circle the answers which seem most true for you or write short answers where appropriate.

1 Did you find the content of the listening text difficult? YES / NO
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 THE RIGHT SPEED?

4 Did the speaker's accent interfere with your understanding? YES / NO

5 Did the speaker sound like a native speaker of English? YES / NO / DON'T KNOW

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 was
EASIER / MORE DIFFICULT / NEITHER EASIER NOR MORE DIFFICULT to understand.
Can you say why?

.....
.....

8 If you have any other comment about this listening task, please write here:

.....
.....

Please return this form to (.....)

Thank you! Your feedback will help us to develop the co-certification

