

13. ASSESSING ENGLISH AS A LINGUA FRANCA¹

Catherine Elder and Alan Davies

This chapter proposes two alternative models for assessing English as a Lingua Franca (ELF). Tests based on the first model resemble existing approaches to assessing English as a foreign language offered by such tests as TOEFL, and IELTS. This model assumes that interlocutors use varieties of English based on Standard English. What distinguishes tests of this model from existing international tests of English is that it explicitly allows test accommodations. Such accommodations modify the test delivery system in order to make it accessible and fair for ELF users without changing the construct. Tests based on the second model assume that ELF may be regarded not as a use of Standard English but as a code in its own right. Similarities to varieties of World Englishes such as Singapore English, Indian English are noted. In tests based on the second model, strategic competence takes precedence over linguistic accuracy. Although both models are somewhat problematic in practice, neither, it is argued, entails any radical reconceptualization of language testing beyond what has already been envisaged and/or enacted in the field. Nevertheless, future tests of ELF may have both symbolic and practical importance, giving greater authority and legitimacy to expanding and outer circle English voices on the one hand and giving flesh to definitions of effective intercultural communication on the other. The chapter concludes by cautioning against moving too quickly to assess ELF before it has been properly described.

Assessing English as a Lingua Franca

The term English as a lingua franca (ELF) may be understood in (at least) four ways:

- ELF 1. The use of English in an interaction where at least some of the participants are non-native speakers (NNS) of English
- ELF 2. The use of English in an interaction where all the participants are NNSs and do not share the same first language
- ELF 3. The use of English in an interaction where all the participants are NNSs and all share the same (or similar) first language

- ELF 4. A (new) code used for interaction among NNSs, not standard English but based on standard English (SE)

Post-colonial or World Englishes (Brutt-Griffler, 2002; Kachru, 1992) such as Indian English could be classified as a subset of ELF (4). As we explain in the following, we exclude them from our main discussion in this chapter, but we use some examples of Hong Kong English to highlight the problems of description (and therefore assessment) faced by ELF (4).

Although, ELF (1), (2), and (3) are dependent on British or American English, ELF in sense (4) is not but is influenced by “the lingua-cultural background of its speakers and the rules of use that characterize the domain within which the interventions take place (such as business, science, the media)” (Seidlhofer, 2001, p. 147). Whether these uses represent a system that is different from SE remains to be explored. The ELF (4) assertion is that English is increasingly being used as a vehicular language among interlocutors who do not speak one another’s language. In fact, such speakers vastly outnumber native speaker (NS) users (Crystal, 1995; Graddol, 1999). Although ELF (1) refers to the use of SE by NS when they are interacting with NNS, ELF (2) and (3) refer to NNS uses of SE, the approximations, adjustments (including errors) they make in their uses of the standard code. What is claimed for ELF (4) is that it is not metropolitan English (British, American, Australian, etc). The claim, supported by corpora such as the Vienna-Oxford Corpus of ELF (Seidlhofer, 2001), is that this ELF (4) is a new international variety. Such a claim, that at some future date there may be a descriptive basis for eventual codification of ELF, is contrary to the claim we have referred to above under ELF (2) and (3), that they are speakers’ use of an existing code. ELF (2) and (3) also differ from one another: ELF (2) refers to interaction by speakers of ELF across linguistic boundaries, that is they are communicating internationally (e.g., Japanese to French, Danes to Indonesians), whereas ELF (3) refers to the use of ELF intranationally (Germans to Germans, Indians to Indians). No doubt such intranational usage is likely to be limited to professional domains, such as that occupied by teachers of English. But for the most part, intranational use of ELF must shade into international use because bankers, business people, scientists and even teachers who use English professionally tend to engage both intra- and internationally in it.

The international role of English as a lingua franca was addressed by Clyne (1994) in his study of intercultural communication at work, in which he analyses a corpus of spoken interactions among NNSs from different first language (L1) backgrounds in the Australian work-place. He extends his argument by challenging the common practice in the academic community of using native speakers of English as reviewers of articles submitted to international journals. He argues that anglophone readers find nonanglophone rhetorical structure unacceptable, citing Kaplan (1966), a view Kaplan has himself rejected (Odlin, 2002). Equally unacceptable are proposals for establishing one or other interlanguage as a code in its own right (Davies, 1989). These proposals have not been taken up, presumably because all interlanguages are seen as approximate states, approximate, of course, to the native speaker, and therefore in deficit. But if ELF (4) is to be regarded as

discrete, it cannot be regarded as a deficit model, which, in essence is what ELF in senses 1, 2, and 3 are, conceived as they all are on native speaker norms.

Norms for ELF, Sense (4): A New Code

What distinguishes natural languages from one another as discrete codes is not their formal uses but their intimate and personal interactions. A code that is used for one function or a small range of functions in a single, usually public, domain is said to be an LSP (language for a specific purpose). Most LSPs such as scientific English or legal English, present the language of public and formal domains. This suggests that ELF (4) is in essence an LSP. Just as LSPs such as scientific English and medical English represent the textual and discoursal engagement by those involved in science or medicine, so ELF (4) represents the formal language interactions by NNSs in a range of fields (not just one as with most LSPs), such as commercial, business, academic, governmental. As always, it is difficult to keep separate the public and the private: while our emphasis will be on the public, there must be many occasions when the private intrudes, occasions when two old friends speak across national borders of their shared business interests and then move, while still in ELF (4), to friends and family matters; or when a husband and wife from different L1 backgrounds who work together in, say, the law, engage in a legal conversation and then exchange news about their parents, children, and so on. The boundaries are slippery and so for the purposes of our discussion of assessment issues, it may be wise to restrict ourselves as far as possible to (spoken) formal and public texts.

What will always be difficult will be to restrict the data to NNSs on the grounds that, whereas NSs using English for these public international purposes are now fewer than NNSs, NSs too are often involved in these matters and so there is something artificial about excluding them. No doubt, the reason for the exclusion is to avoid the constant appeal to native speaker judgements, triggering the stigma of error for NNS performance. And so we will—as far as possible—restrict ourselves to wholly NNS texts, recognizing that in doing so we have accepted the rationale of porous norms, that what would be regarded as an error by an educated NS may be normal accepted usage for an educated NNS. So much is clear. But it will not be so easy to determine which NNS usages are themselves errors for NNSs until we are able to describe a non-Standard English for NNS, equivalent to the NS Standard English. In other words, we are just not sure what norms to apply for ELF (4) and without that knowledge, assessment remains speculative. However, we attempt that speculation here. What is distinct about ELF, what distinguishes it from a natural language is that it functions largely, as House (2002) reminds us, as a language for communication and not for identification. Firth (1991) concurs, emphasising that ELF is also interactional.

There are some clues to what a description of ELF (4) might look like. These clues do tend to be confined to one country or speech situation. The question is how far these features can be generalized internationally. Our speculation is in two parts. First, we examine the analyses offered by Seidlhofer (2001, 2004) and by

Jenkins (2002, 2005), both of which, in somewhat different ways, present ELF (4) as a deficit L1 NS model. Second, we exemplify one World English context, using data provided by Joseph (2004) for Hong Kong English: what this example suggests is that it may be that the ELF (norms) that Seidlhofer and her coworkers are concerned with are realized in post-colonial World Englishes. Hong Kong English, after all, like other World Englishes such as Singapore English (Crewe, 1977) can be, or is in reach of being described in its own right; a syllabus could be constructed which would inform both teaching and testing. The goal aimed at in testing would then be not a deficit L1 NS English speaker but the Hong Kong English syllabus. In effect, what would emerge would be an alternative (New) NS model of Hong Kong English, in itself as open to the charge of hegemony as is the present anglophone NS.

Seidlhofer (2004) recognizes that so far there is little evidence on which to base a model: “the bulk of the descriptive work still needs to be done” (2004, p. 222). However, she does offer a few pointers. She emphasises Jenkins’s work on phonology (Jenkins, 2002) “culminating in what she (Jenkins) has termed the *Lingua Franca Core*” (that will “assess which phonological features are—and which are not—essential for intelligible pronunciation when English is spoken in *lingua franca* contexts” (2004, p. 216). But those words “which phonological features” raise the question “of what?”, to which the answer surely has to be of L1 NS English. Again, in her section on pragmatics, Seidlhofer tells us “some fairly clear insights are emerging” (2004, p. 218). In general what these studies of ELF pragmatics appear to show is a greater tolerance among ELF (4) speakers—they are less focused on form, more concerned with reaching for the message. Seidlhofer quotes McKay (2002) with approval, that what matters is intelligibility rather than correctness. Thus, misunderstandings are resolved by communication strategies, interference from L1 interactional norms is rare. Over all, ELF (4) talk seems to be “overtly consensus-oriented, cooperative and mutually supportive, and thus fairly robust” (2004, p. 218).

When Seidlhofer considers lexicogrammar, the details we are given are of common errors which “appear to be generally unproblematic and no obstacle to communicative success” (2004, p. 220), for example, dropping the third person present tense—making no distinction between the relatives *who* and *which*, omitting definite articles, failing to use correct forms in tag questions” (2004, p. 220). Again, then, what counts is intelligibility rather than correctness. Jenkins proposes, as we have seen, a *lingua franca core* for phonology, maintaining that this core contains the essential phonological features which are required for phonological intelligibility. The features she cites appear to be a reduced and simplified version of L1 NS norms and include, for example, the consonantal inventory with a number of exceptions such as dental fricatives; the aspiration of word initial voiceless stops, the presence of all sounds in word-initial clusters, maintenance of the contrast between long and short vowels; and placement of contrastive nuclear stress.

As Seidlhofer acknowledges, work on ELF (4) phonology has gone ahead of other areas. But once again, the direction is not yet towards a separate, discrete description of a stand-alone model which could provide its own EFL (4) syllabus. In all cases, what we are offered is a reduced version of L1 NS English. NNSs are, it is

asserted, more tolerant of these reductions, but in traditional terms, what is being described, and what is proposed, is an L1 phonology with (acceptable/unobtrusive) errors.

Mauranen (2003) discusses a particular variety of ELF (4), an academic NNS-NNS register. She maintains that the academic texts in the corpus collected in Tampere, Finland, show marked pragmatic features such as hedging and self-repair. We consider that the interactions she describes belong primarily to an LSP variety and as such may conceivably be subsets of L1 English. The pragmatic features she instances are, after all, found in regular use in L1 NS speech. If there is a greater preponderance of such features that is interesting but interesting from an LSP not an ELF (4) point of view, unless it could be demonstrated that hedging and self-repair are more prevalent across all domains of ELF use than is the case for SE.

For Joseph (2004), Hong Kong English is a reality in the sense that it has features that mark it as distinct and discrete. It remains true that Hong Kong English, like other World Englishes, and indeed like anglophone standard Englishes, such as American and Australian, remains very close to British English. But the plausible claim in the case of Hong Kong English is that this is not based on a deficit model. That is the claim: We acknowledge that there is an equally compelling argument that states that the extreme acrolect of Hong English is not distinct from educated British English, with the exception of accent and vocabulary range. Joseph (2004) suggests that the following are typical of what he calls Hong Kong English:

- The flattening of count noun versus mass noun distinction, reflected in the use of singular for Standard English plural and in the different distribution of definite and indefinite articles (e.g., ‘group of ... shirt’; ‘for alternative’)
- Highly distinctive distribution of prepositions
- Semantic differences in individual lexical items (e.g., ‘prepare’ meaning ‘have available, bring’) (Joseph, 2004, p. 141)

Here is a transcript, also from Joseph, 2004, of part of an interview with Szeto Wah, a prominent Hong Kong pro-democracy politician. Underlined are those features Joseph considers typical of Hong Kong English. He notes particularly the different use of verb tenses from Standard English:

Q. The Alliance has raised a lot of money from the citizens through its activities all these years. What is the financial picture now? What if all the money are spent? Will the Alliance accept foreign sponsorship?

A. As of April, we still have three million Hong Kong dollars in the bank. We have been trying our best to cut all unnecessary expenses. I think this year we’ll have no problem. And every year, especially during the commemoration activities, we receive a lot of donations from the citizens. However, as Hong Kong is going through an economic down turn recently, we shall

have to see. If we can raise a million and a half this year at the commemoration activities, it will be okay. Last year, we have raised more than two million Hong Kong dollars. Money is a problem, but not the major one. We will adjust to work with what we have. We will never seek foreign sponsorship. All our past resources are based on the money donated to us directly from the citizens.

(Joseph, 2004, p. 141)

Issues of Measurement

We now turn to our main topic, the implications of what has been outlined previously for ELF assessment, by outlining what we regard as key issues that need to be addressed in any discussion of measurement for these will inform our subsequent discussion of the special case of ELF testing.

First and foremost, before embarking on any test development endeavour we must consider the nature of the criterion or the construct (the *what* of assessment). It is generally agreed that all subsequent issues of test design and test use hinge on how the construct is defined. At issue in this chapter is the question (discussed earlier) of whether the ELF construct should be defined in terms of Standard English (ELF, Senses 1, 2 and 3) or whether (as Jenkins and others propose) we need to define our target differently and hence rethink our testing designs and practices (ELF, Sense 4).

Our second concern in developing tests is with the various facets of the test method used to elicit the test takers' performance (*the how* of assessment). Test items or tasks will contain input (whether a short written or spoken prompt or a longer text or texts possibly involving multiple speakers or writers) which, for the sake of authenticity, are sampled from domains of language relevant to the construct, and will require a particular type of response which may involve a candidate selecting from given alternatives (e.g., true-false and multiple-choice), or which may be open-ended (e.g., an essay or speech). We focus here on speaking tests, given the centrality of speaking in ELF communicative contexts; there the test taker's interaction between the test task/item may also be mediated by an interlocutor, whether this be the examiner who elicits the performance, or another test-taker as is the case in a paired or group oral task. The nature of the construct will impact all the above aspects of the elicitation procedure.

Last but not least we must pay attention to the scoring process which is always mediated by a set of criteria or a marking key and, in the case of speaking, a trained rater who uses the marking key/criteria to make decisions about where a test-taker sits on the measurement scale. A further stage in the scoring process known as standard-setting, typically involves members of what Brown and Hudson (2002) call the decision target community, or in other words the test users themselves, who are sometimes consulted in an attempt to determine at what score level the cut-off between acceptable and unacceptable performance should be set. Both scoring and

subsequent decision-making about acceptable and unacceptable performance are made with reference to the test construct.

Because the stakes of testing are often high, there are agreed-upon test qualities which tests must meet for public accountability and fairness, including the requirement that the scores generated from a test or assessment procedure are consistent rather than arbitrary (test reliability), that the inferences we draw from these test scores are defensible and devoid of bias with respect to the specified construct (test validity), and that the test does not cause undue harm, and any unintended negative consequences are minimized (test impact). In the latter regard, it is often advocated that attention be paid to the washback of the test on the curriculum, in the hope that what is tested does not adversely influence what is taught, but rather that it will promote good teaching and learning practices. Validation studies, mentioned in the next section, are directed to investigating the extent to which these test qualities have been achieved within the constraints of practicality which places inevitable limits on the how far tests can succeed in capturing the characteristics of real world language use.

What Would an ELF Test Look Like?

In the following sections we attempt to map out the terrain for ELF assessment using the two main construct definitions (ELF 1, 2, & 3; and ELF 4) proposed at the outset of this chapter and considering the implications of these definitions for our elicitation methods and scoring procedures.

Proposal 1. If we accept, in accordance with our definitions of ELF 1, 2, and 3, that the construct to be measured is Standard English (SE) communication, then all that we need to cater for ELF users is standard tests of SE, such as IELTS or TOEFL or indeed literacy tests designed for NS users such as the OECD PISA tests or the verbal component of the SAT. These may require some accommodations in the delivery and scoring of such tests on the grounds that the test population includes ELF users whose speech or writing may deviate from the codified standard, not necessarily because they are deficient in English, but perhaps because they inhabit communities where English is acquired nonnatively and particular nonnative features have assumed the status of stable varietal differences (Lowenberg, 2002). Strict adherence to native speaker norms of correctness are arguably unreasonable and irrelevant to the target language construct, given that successful communication does not depend on them (although some explicit statement about the relaxed norms will need to be made). What is proposed here is a series of “accommodations” in the testing sense (Abedi, Lord, Hofstetter, & Baker, 2000; Bailey & Butler, 2004.), that is, modifications to the test delivery system in order to make it accessible and fair for ELF users without changing the construct, which remains as the ability to communicate effectively in something approximating SE.

An analogous example may be useful at this point. Accommodations are routinely made in high stakes testing encounters for students with special needs, for example those with poor eyesight, who may be given texts to read in a larger font.

These accommodations are based not on the notion of a deficit in the ability measured by the test (the visually challenged test taker may well be a good reader) but on the notion of inadequacy of the test delivery process—such a person might be more affected by construct-irrelevant aspects of this process (such as the font size of the textual input) than other test takers with normal vision and hence deserves special consideration. What is important for test validity purposes is that the accommodations made do not alter the fundamental nature of the construct (in this example, the ability to read) and hence change the meaning of the scores assigned. This can be empirically tested, by giving the same accommodations to people with normal sight and (hopefully) demonstrating that the score they achieve is the same regardless of which font size they are reading in.

Test accommodations on a test of speaking for ELF users on an SE test could take the following forms:

a) Vet texts used in SE tests for potential bias against ELF users who might have limited opportunities to encounter particularly topics or genre (e.g., nursery rhymes) in particular contexts of use and therefore lack the background knowledge needed to make sense of these texts. Such vetting might result in different text choices or the provision of relevant background information by way of introduction to the passage. Help would be offered only with areas of background knowledge deemed to be of marginal relevance to the construct under test.

b) Gloss or avoid altogether any lexical items or structures which are likely be unfamiliar to NNS users (of ELF 1, 2, or 3). (An example would be the word *prepare* which, as noted earlier, has a different meaning in SE than in Hong Kong English). The idea would be to avoid single lexical or grammatical items, which NNS test takers could not be reasonably expected to have encountered in the local or wider community of ELF users, to interfere with the measurement of their ability to communicate effectively in English.

c) Use interlocutors (either examiners or other candidates) who are expert NNS/ELF users and therefore have experience in ELF contexts and know how to adjust their speech in ways familiar to the test takers. The argument here is that ELF candidates who have limited experience of communicating with native speakers deserve to receive the benefits of this kind of scaffolding. It would of course be important to ensure that raters did not evaluate such accommodations negatively by scoring its recipients more severely as revealed by Ross (1992).

d) Train raters, whether NSs or NNSs, to ensure that only those errors which result in miscommunication are penalized. Presumably nonstandard linguistic and discursal features characteristic of ELF would be overlooked if the overall message was clear. Features already documented as new norms in one or other Outer or Expanding circle context include nonstandard use of phrasal verbs (e.g., “cope up with”), the pluralization of uncountable nouns (e.g., “equipments”), the use of neologisms (“prepone”), topic comment structure which overturns standard rules of word order (“TV I don’t usually watch”) and tense/aspect (“She was having a

headache”) (Higgins, 2003; McKay, 2002). What we are proposing then is the application of the Let it Pass principle (Seidlhofer, 2004) in rating, in keeping with our claim that EFL 1, 2, and 3 involve a relaxation of SE norms.

e) Involve ELF users in standard setting exercises to check whether that cutoffs are low enough to ensure that competent ELF communicators can pass the test, regardless of any superficial differences in accent or formal usage. Again it would be important to ensure that NNS judges were indeed more permissive than their NS counterparts, given that a number of studies have shown the reverse to be true (Fayer & Krasinski, 1987; Santos, 1988).

The purpose of such accommodations would be to boost test scores for those ELF users, who might otherwise have been unfairly judged or whose communicative competence in English might have been underrepresented. The problem with making the abovementioned adjustments for the ELF user is that, on a test where language serves both as the vehicle and the object of assessment, it may be difficult to determine where the test method stops and the test construct begins. In fact it seems very likely that some of the preceding modifications to the test will result in scores from the same test instrument having different meaning for NS and ELF users; Davies, Hamp-Lyons, and Kemp (2003); Elder (1997, 2000), Elder, McNamara, and Congdon (2004) discuss the complex issue of determining bias in language testing. A further problem with this proposal is that giving ELF users the benefits of positive discrimination may in fact have a negative impact on test-takers by reinforcing the notion of a deficit in their competence with respect to the target SE norms.

Proposal 2. If we reach a point where ELF 4 (like Hong Kong or Singapore English) can be demonstrated to constitute a new code, then we need to use that code (rather than SE) as a basis for our test construction. Accommodations made in Proposal 1 to compensate test takers for lack of access to standard English, would no longer be seen as accommodations but instead would become an integral part of the new test construct. Although ELF researchers acknowledge that attempts to describe the new ELF 4 code are at a very early stage, it is worth speculating about what form such a test might take. Test items or tasks used to elicit performance would be sampled directly from domains relevant to ELF communication and therefore would contain lexical, grammatical, phonological, prosodic and discoursal features peculiar to ELF. We have listed some possible features earlier but it is too soon to claim that these features operate cross-nationally in all ELF contexts. However, for the sake of argument, let us assume that the Hong Kong idiosyncracies of tense/aspect described previously were true for all ELF users. A grammar test might therefore include an item such as the following, with (a) rather than (d) chosen as the correct response:

Complete the sentence with the appropriate form of the verb in parenthesis
 Last year we _____ more than a million dollars.

- a. have raised*
- b. are raising
- c. raise
- d. raised

Native speakers of one or other SE might conceivably be disadvantaged on an item of this kind, although, given the limited number grammatical differences between standard English and the new English varieties which have been codified thus far (McKay, 2002), the impact of such differences on overall test performance is likely to be slight.

Grammatical accuracy however is not generally what ELF 4 proponents are concerned with. It is more likely that a test of ELF would take the form of a performance assessment with listening texts or speaking tasks designed to mirror those carried out in relevant communicative contexts (Britt-Griffler, 2005; Canagarajah, 2005). Input for a listening comprehension test might involve a news feature from, say, a “This week in Asia” radio broadcast produced by speakers with a variety of NNS accents. Since one of the valued qualities of an ELF speaker, according to Jenkins (2002, 2005, in press) is the ability to accommodate to a variety of nonnative accents, a candidate might be allowed to replay the texts as often as needed to answer the comprehension questions, but given bonus points for completing the task quickly on the grounds that immediate comprehension would be indicative of greater flexibility or adaptiveness as a listener. (In fact such adaptiveness may have less to do with language ability than with the personality of the test taker, so personality testing may be an avenue worth exploring in ELF contexts.)

A speaking test might involve a simulated interaction between business people in Denmark and Saudi Arabia negotiating an export/import deal (see the following example), between researchers from Germany and Italy discussing the possibilities for an academic exchange between their respective countries, between Korean and Japanese teacher participants at an international teacher education conference in Singapore, or between sports organizers from Greece and China discussing plans for the Olympic Games in Beijing. Such tests would be likely to involve integrated tasks, with participants drawing on paper- and Internet-based sources containing pricing information, conference schedules, events calendars, and so on. If, as has been claimed, the majority of such communicative interactions involve NNS rather than NS of English, it would make sense, for the sake of authenticity, to use NNS interlocutors to elicit the relevant speech samples from test takers or to set up paired or group tasks where it is the test candidates who assume the various communicative roles. Herein lies a conundrum. In current tests involving paired interaction, care is generally taken (in the interests of eliciting a valid sample of performance) to ensure that the performance of one candidate is not adversely affected by a lack of proficiency in the test taker’s partner or partners (Iwashita, 1996). In an ELF test, however, the fact of being paired with a less competent partner might be seen as essential for test validity/authenticity because the use of English by NNS around the globe inevitably involves speakers at various levels of

proficiency even though ability to deal with such variation is crucial to successful ELF interaction. The criteria for assessing performance would be concerned primarily with task fulfilment—whether the relevant negotiation was carried out successfully regardless of the partner's/partners' competence (Prabhu, 1987)—and with the participants' ability to accommodate to the other party and to self-repair or use other strategies to disambiguate in the event of misunderstanding.

An example from Firth (1991) illustrates the point quite well. The following conversation takes place between a Danish exporter/producer of cheese (A) and one of his international buyers, a Saudi Arabian based Indian importer/wholesaler (B), both NNSs of English:

- 1 B... so I told him not to send the cheese after the, the blowing (.) in the customs. We don't want the order after the cheese is blowing
- 2 A: I see, yes.
- 3 B: So I don't know what we can do with the order now. What do you think we should do with this is all blowing mister Hansen?
- 4 A: I'm not er (0.7) blowing er what er, what is this er too big or what?
- 5 B: no the cheese is bad mister Hansen, it is like fermenting in the customs' cool rooms.
- 6 A: ah it's gone off
- 7 B: yes it's gone off
- 8 A: well you know, you don't have to do anything because it's not... (turn continues)

(Modified from Firth, 1991, p. 275)

What we see here is the use of the (non-SE) term blowing to describe the fermenting process which has ruined the cheese in transit to the Saudi Arabian purchaser. The term creates some confusion for the Danish producer, although this confusion is resolved in turns 4 to 7. It is hard to determine who is the more proficient user here; B's speech displays more obvious markers of nonnativeness ("in *the* customs" "after the cheese *is* blowing"), but from an ELF 4 testing perspective what is important is the fact that the communication problem was successfully repaired by the conversational partners concerned, both of whom show strategic resourcefulness. Eventually, in turn 4, A signals lack of understanding and B is able to disambiguate through suppliance of an alternative term 'ferment' and quickly adjusts and responds to A's use of the more colloquial term 'gone off.'

Such strategic competence on a test of ELF 4 would presumably take precedence over linguistic accuracy (which if deemed to be relevant to the success of a particular encounter, would be defined in terms of the new code rather than any NS standard). The top level of an oral rating scale might therefore read as follows:

Speech is clear, fluent, and sustained. The candidate is able to use his/her linguistic resources with confidence to express ideas in a clear and efficient manner including appropriate detail where this is needed or otherwise fulfil the requirements of the task. Any linguistic limitations are overcome by using paraphrase or gesture or any other available strategies for conveying the message. Pronunciation and intonation are such that listeners, including those with limited English proficiency, ultimately succeed in understanding the content or intent of the communication. In the event of any difficulty, the candidate makes appropriate lexical, accentual and structural adjustments to his or her speech to resolve the problem and get his or her meaning across. The candidate is likewise able to ask for clarification in the face of difficulty in understanding others and to adapt quickly to unfamiliar accents and communicative styles which are different from his/her own.

It could be argued that what is being assessed here is a kind of aptitude, not for learning a new language but for coping with different speech varieties, although aptitude has traditionally been characterized as an individual psycholinguistic phenomenon. The preceding description emphasizes the reciprocal nature of communication and also challenges the notion that competence resides entirely within the individual user. The case could therefore be made that group scores are more appropriate than individual ones and indeed that the best judges of the success or otherwise of ELF communication are the users themselves and, accordingly, that peer assessment would be an appropriate means of determining speakers' competence as ELF users. Certainly assessors would need to be NNSs familiar with the ELF code, and the pragmatic demands of the contexts in which it was used. Ideally such raters would be from different L1 backgrounds to neutralize the possible effects of familiarity with candidate speech or communicative style on rating behaviour. Note however that raters, regardless of their status as NS or NNS, as participants in or observers of the communicative exchange, would have to demonstrate that, after training, they were able to apply the rating scale consistently with one another (Taylor, 2005). There would be, in our view, no grounds for relaxing the normal standards of reliability needed to ensure consistency of judgement. Presumably NSs of English could also be used employed as raters provided that they showed the capacity to rate similarly to NNS ELF users.

Standard-setting exercises would probably involve members of the relevant professions (academic administrators, teacher educators or bureaucrats in the sporting industry who would be asked to decide whether or not the task had been completed sufficiently well to meet the standards of the relevant profession. It is not inconceivable however that the standards applied by such professionals might differ considerably from one another, given the subjective nature of such judgements and the difficulty (already noted above) of specifying what constitutes an unacceptable departure from the ELF code. In a testing environment there is no escaping judgement of unacceptability. While Seidlhofer, (2001, p. 144) has made the interesting suggestion that what are currently defined as learning strategies in current learner corpora might well be redefined as communication strategies in an ELF data base (because they are part and parcel of normal communicative competence in ELF, rather than an indication of deficit with respect to SE norms). However, some

communication strategies are no doubt more efficacious than others. If this were not so, there would be no grounds for testing ELF at all and even unintelligible encounters could be classed as legitimate instances of ELF use.

Discussion

At this point it should be noted that many of the ideas put forward thus far are not new and have already been broached by language testers. Proposal 1 is essentially a soft version of Standard English testing, but have we not already embraced this relaxation of standards to some extent? The push for communicative language testing has brought with it a tendency to accept interlanguage approximations in the ways that have been discussed in Proposal 1 with a strong emphasis in assessment criteria on meaning as opposed to form. Moreover the need to invoke *expert user* rather than the *native speaker* as the benchmark for assessing language ability has already been acknowledged, with support from empirical data from writing tests showing that native speakers themselves vary widely in their performance (Hamilton, Lopes, McNamara, & Sheridan, 1993). The effect of interlocutor accommodations on language test performance has also been explored in language testing contexts (e.g. Lazaraton, 1996; Malvern & Richards, 2002; Ross, 1992) even if the implications of this research have not been explicitly spelled out from an ELF perspective. Work on dynamic assessment by Poehner and Lantolf (2003) has relevance here, because it stresses that probing the test takers' potential through a mediated activity may be a better means of assessing an individual's communicative capacity than a more static assessment, where intervention from the interlocutor may be seen as interfering with the measurement of a test taker's true ability. Also relevant to the ELF testing agenda is work by O'Hagan (1999) and Weigle, Boldt, and Valsecchi (2003) and exploring the different rating patterns of Faculty members from different disciplinary backgrounds when rating NNS and NS students' writing. These studies suggest that the softening of norms implicit in Proposal 1 is already occurring in contexts where NNSs are present in substantial numbers (although the reasons for this may be largely pragmatic, rather than stemming from a conviction that NNS deserve special treatment).

As for the ideas offered in Proposal 2, which involves a more radical rethink of the test construct, language testing research is now embracing the notion that nonnative accents may need to be incorporated in language tests. More than one study has been commissioned to explore the effect of these on comprehensibility (Major, Fitzmaurice, Bunta, & Balasubramanian, 2002). The idea of scoring test takers on their ability to accommodate new input (as proposed earlier in relation to tests of listening comprehension) and/or on the amount of support or prompting needed during test performance, has already been entertained in aptitude tests, such as the Lern test (Guthke, Heinrich, & Caruso, 1986) designed to measure international students' capacity to profit from a compulsory course in German for university entrance purposes. It is also widely accepted that proficient users of English, regardless of their L1 background, might be used as raters provided that they meet the necessary criteria (Taylor, 2005) and there are number of studies which have explored the impact of language background on rater behavior (Chalhoub-

Deville & Wigglesworth, 1995; Hamp-Lyons & Davies, 2005; Hill, 1996; Lazaraton, 2005). Tests using paired or group oral tasks have long been practiced and discussed in the field (e.g., Bonk, 2003; Egyud & Glover, 2001; Foot, 1999; Fulcher, 1996; Iwashita, 1996; O'Sullivan, 2002; Saville & Hargreaves, 1999), although the practice of assigning scores for groups rather than individual performance is usually limited to classroom contexts.

Peer rating has likewise been researched (Cheng & Warren, 2005; Patri, 2002,) but again studies tend to focus on classrooms rather than high-stakes testing situations. The difficulty of assessing individual competence when communication is necessarily co-constructed has been explored by a number of researchers (e.g., Brown, 2003; Lazaraton, 1992; Lumley & Brown, 1996; McNamara, 2001; Swain, 2001) although further work with NNS-NNS dyads as opposed to NS-NNS pairs is clearly needed. ESP tests for NNS users containing some of the features described in our second proposal are already in existence (for example, Brown & Lumley (1998) account of a "localized" test of English proficiency for teachers of English in Indonesia), even if such tests have thus far been developed only for intranational rather than international use. Finally, the use of task fulfilment as a rating criterion is well established in the field. Douglas (2001) and Jacoby and McNamara (1999) working in different ESP contexts, have advocated the use of "indigenous" rating criteria and standard-setting procedures which reflect the values of users in particular professional environments. Brown (2004) and Elder (2001), among others, have explored the application of such criteria in professional communication contexts (tourguiding, and language teaching, respectively) and shown how the judgments thus derived sometimes sit oddly with more traditional criteria focussing on the accuracy of the language system. In sum, the current quest by LSP researchers to bring testing practices into line with what really matters for participants in particular communicative contexts (Douglas 2000), seems to be precisely what ELF researchers now seek.

At the moment it is very hard to envisage that tests for ELF users could ever have applications across rather than within particular ELF domains and become a stable undifferentiated generic code, as implied in our previous definition of ELF 4. Indeed, as Bruthiaux (2003) has cogently argued with respect to the now widely used categories of inner, expanding and outer circle Englishes, attempts to establish such uniformity may prove to be sociolinguistically naïve because they ignore the variation that exists within and between national boundaries and may end up reinforcing the reification of the NS and NNS categories and the stigmatization of the latter that the ELF movement attempts to counteract. Moreover, although the various features of ELF use in particular contexts can conceivably be captured in domain-specific ESP tests, there are important practicality considerations to bear in mind. Special purpose testing is, by its very nature, restricted in scope and as such likely to have limited generalizability and less sway with score users, and possibly test takers themselves (see research by Bolton, 2004; Timmis, 2002), than is the case with current tests of SE which have greater prestige and wider currency.

Conclusion

Calls for new approaches to the assessment of ELF (Jenkins, 2005; Lowenberg, 2002) have thus far come from those outside the professional language testing field unfamiliar with the constraints and requirements of language testing that we have alluded to in the discussion of issues of measurement here. These approaches are stronger on politics than applied linguistic realities and appear to be a push by claimants from outer and expanding circles for ownership of English (Higgins, 2003), or perhaps a plea for official recognition of their legitimacy as users of English and acknowledgment of validity of their intra- and cross-nationally negotiated language identities. Although such claims are to be applauded for their attempts at achieving distributive justice, they carry within them a paradox, which we have previously adumbrated in reviewing issues for possible norms for ELF.

One of the arguments for establishing ELF norms and using these rather than SE as a basis for measurement is that the resultant tests would offer more valid representations of target language use domains, and have positive impact on test takers resulting in a reduction in anxiety on the part of ELF users, who would no longer feel pressured to adhere to norms of SE English. Tests such as these would also have positive washback on teaching in that the syllabus would be designed around their likely communicative needs rather than on unattainable native speaker norms, which in any case do not apply in the contexts of concern. However once such ELF norms reach the point of being structurally stable enough for codification purposes and hence operationalizable in the form of language tests, they presumably would have the same power to demoralize, oppress, and disenfranchise nonstandard or nonproficient users of ELF as have current tests of SE. What is currently a proposal for legitimization of nonstandardness and affirmation of NNS identity could risk becoming a new monolithic standard with all the attendant consequences for those lacking the command of the new code.

Even so, we believe that the diversity of speakers and purposes involved in ELF communication make it unlikely that a description of ELF which holds across multiple contexts of use and can function as a code in its own right will eventuate in the near future. Although norms of English acceptability are inevitably shifting to accommodate the variations characteristic of ELF users, it is at best premature (James, 2000) and at worst misguided to attempt across-the-board structural typologizing of these variations. Methodological and conceptual uncertainties still abound in the ELF literature and it may be more useful, until greater clarity is achieved, to conceive of ELF as a series of register varieties, each serving a highly specific communicative purpose. Such a context-specific formulation, we have argued, does not require a testing agenda radically different from what has already been adopted or at least entertained in other LSP contexts.

We should nevertheless acknowledge the potential contribution of this vibrant area of sociolinguistic enquiry, which is likely to result in an intensification of what the communicative language teaching and testing movement has already set in train: a reduced emphasis on the linguistic code, which can offer only partial

explanations for the communicative phenomena we try hard to capture in our language tests, and further refinements in our understanding of the pragmatics of particular intercultural and cross-cultural encounters. Such research may help us achieve greater explicitness in our descriptions of communicative effectiveness (or its obverse) in a variety of LSP contexts. This will give more flesh to our construct definitions and, accordingly, inform our choice of assessment criteria, thereby helping us produce more contextually sensitive descriptions and profiles of NS user behavior for both predictive and diagnostic purposes.

Whether or not the ELF project is, as we have suggested, stronger on politics than applied linguistic realities, it forces us to recognize that, when used in interaction, language is not an abstract construct but is embodied in people and therefore we, as students of language, need to take account of the politics of language, of language as an identity marker. “Our identities,” writes Joseph (2006, in press) “are not something essential and permanent that stand aloof from our relations with others (but) manifested in those relations.”

Note

1. The authors wish to acknowledge the contribution of Liz Hamp-Lyons to our thinking about this chapter.

REFERENCES

- Abedi, J., Lord, C. Hofstetter, C., & Baker, E. (2000). Impact of accommodation strategies on English language learners' test performance. *Educational Measurement: Issues and Practice*, 19(3), 16–26.
- Bailey, A., & Butler, F. (2004). Ethical considerations in the assessment of the language and content knowledge of U.S. school-age English learners. *Language Assessment Quarterly*, 2 & 3, pp. 177–193.
- Bolton, K. (2004). World Englishes. In A. Davies & C. Elder (Eds.), *The handbook of applied linguistics* (pp. 367–396). Malden, MA: Blackwell.
- Bonk, W. (2003). A many-facet Rasch analysis of the second language group oral discussion. *Language Testing*, 20(1), 89–100.
- Brown, A. (2004). LSP testing: The role of linguistic and real world criteria. In r. Khoo (Ed.), *LSP: Problems and prospects* (pp. 202–218). Singapore: SEAMEO Regional Language Center.
- Brown, J. D. & Hudson, T. (2002). *Criterion-referenced language testing*. Cambridge, Cambridge University Press.
- Brown, A. (2003). Interviewer variation and the co-construction of speaking proficiency. *Language Testing*, 20(1), 1–25.
- Brown, A., & Lumley, T. (1998). Linguistic and cultural norms in language testing: a case study. *Melbourne Papers in Language Testing*, 7(1), 80–96.

- Bruthiaux, P. (2003). Squaring the circles: Issues in modelling English worldwide. *International Journal of Applied Linguistics*, 13(2), 159–178.
- Brutt-Griffler, J. 2002. *World English: A study of its development*. Clevedon, UK: Multilingual Matters.
- Brutt-Griffler, J. (2005, September). *Striving to be the native speaker*. Paper presented at British Association of Applied Linguistics (BAAL) Conference, Bristol, UK.
- Canagarajah, A. S. (2005, July). *Changing communicative needs, revised assessment objectives*. Paper presented at Symposium on The Assessment of World Englishes AILA Conference, Madison, WI.
- Chalhoub-Deville, M., & Wigglesworth, G. (2005). Rater judgement and English language speaking proficiency. *World Englishes*, 24(3), 383–391.
- Cheng, W., & Warren, M. (2005). Peer assessment of language proficiency. *Language Testing*, 22(1), 93–112.
- Clyne, M. (1994). *Inter-cultural communication at work: Cultural values in discourse*. Cambridge: Cambridge University Press.
- Crewe, W. (1977). Singapore English as a non-native dialect. In W. Crewe (Ed.), *The English language in Singapore* (pp. 96–119). Singapore: Eastern Universities Press.
- Crystal, D. (1995). *The Cambridge encyclopedia of the English language*. Cambridge: Cambridge University Press.
- Davies, A. (1989). Is international English an interlanguage? *TESOL Quarterly*, 23(3), 447–467.
- Davies, A., Hamp-Lyons, L., & Kemp, C. (2003). Whose norms: International proficiency tests in English. *World Englishes*, 22(4), 571–584.
- Douglas, D. (2000). *Assessing languages for specific purposes*. Cambridge: Cambridge University Press.
- Douglas, D. (2001). Languages for specific purposes assessment criteria: Where do they come from? *Language Testing*, 18(2), 171–185.
- Egyud, G., & Glover, P. (2001). Oral testing in pairs—a secondary school perspective. *ELT Journal*, 55(1), 70–76.
- Elder, C. (1997). What does test bias have to do with fairness? *Language Testing*, 14(3), 261–277.
- Elder, C. (2000). Is it fair to assess native and nonnative speakers in common on school “foreign” language examinations? In A. Kunnan (Ed.), *Fairness in language testing: Proceedings of the 19th Language Testing Research Colloquium* (pp. 82–104). Cambridge: Cambridge University Press.
- Elder C. (2001) Assessing the language proficiency of teachers: Are there any border controls? *Language Testing*, 18(1), 149–170.
- Elder, C., McNamara, T., & Congdon, P. (2004). Rasch techniques for detecting bias in performance assessments: An example comparing the performance of native and non-native speakers on a test of academic English. *Journal of Applied Measurement*, 4(2), 181–197.
- Fayer, J. M., & Krasinski, E. (1987). Native and nonnative judgements of intelligibility and irritation. *Language Learning*, 37, 313–326.
- Firth, A. (1991). Lingua franca negotiations: Towards an interactional approach. *World Englishes*, 9(3), 269–280.

- Foot, M. (1999). Relaxing in pairs. *ELT Journal* 53(1), 36–41.
- Fulcher, G. (1996). Testing tasks: Issues in task design and the group oral. *Language Testing* 13, 23–51.
- Graddol, D. (1999). The decline of the native speaker. *AILA Review*, 13, 57–68.
- Guthke, J., Heinrich, A., & Caruso, M. (1986). The diagnostic program of “syntactical rule and vocabulary acquisition”—A contribution to the psychodiagnosis of foreign language learning ability. In F. Klix & H. Hagendorf (Eds.), *Human memory and cognitive capabilities: Mechanisms and Performances*, (pp. 903–911). Amsterdam: Elsevier.
- Hamilton, J., Lopes, M., McNamara, T., & Sheridan, E. (1993). Rating scales and native speaker performance on a communicatively oriented EAP test. *Language Testing*, 10(3), 337–353.
- Hamp-Lyons, L., & Davies, A. (2005, July). Bias revisited. Paper presented at Symposium on The Assessment of World Englishes, AILA Conference, Madison, WI.
- Heritage, J. (1984). *Garfinkel and ethnomethodology*. Cambridge, MA: Polity Press.
- Higgins, C. (2003). Ownership of English in the Outer Circle: An alternative to the NS/NNS dichotomy. *TESOL Quarterly*, 34(3), 615–644.
- Hill, K. (1996). Who should be the judge? The use of non-native speakers as raters on a test of English as an international language. *Melbourne Papers in Language Testing*, 3(2), 29–31.
- Iwashita, N. (1996). Validity of the paired interview format in oral performance assessment. *Melbourne Papers in Language Testing*, 5(2), 51–67.
- Jacoby, S., & McNamara, T. (1999). From language proficiency to interactional competence. *The Modern Language Journal*, 70(4), 366–372.
- James, A. R. (2000). English as a European lingua franca: current realities and existing dichotomies. In J. Cenoz & U. Jessner (Eds.), *English in Europe: The acquisition of a third language*, pp. 22–38. Clevedon: Multilingual Matters.
- Jenkins, J. (2002). A sociolinguistically based, empirically researched pronunciation syllabus for English as an International Language. *Applied Linguistics*, 23(1), 83–103.
- Jenkins, J. (2005, September). The spread of English as an international language: A testing time for testers. Paper delivered at British Association of Applied Linguistics Conference, Bristol, UK.
- Jenkins, J. (In press). *English as a lingua franca: Attitudes and identity*. Oxford: Oxford University Press.
- Joseph, J. E. (2004). *Language and identity: National, ethnic, religious*. Basingstoke, UK: Palgrave Macmillan.
- Joseph, J. (2006, in press). *Language and politics*. Edinburgh: Edinburgh University Press.
- Kachru, B. (1992). *The other tongue* (2nd ed). Urbana: University of Illinois Press.
- Kaplan, R.B. (1966). Cultural thought patterns in intercultural education. *Language Learning*, 16, 1–20.
- Lazaraton, A. (1992). The structural organization of a language interview: a conversational analytic perspective. *System*, 20, 373–86.

- Lazaraton, A. (1996). Interlocutor support in oral proficiency interviews: The case of CASE. *Language Testing*, 13, 151–72.
- Lazaraton, A. (2005, September). Non-native speakers as language assessors: Recent research and implications for assessment practice. Paper presented at British Association of Applied Linguistics Conference (BAAL), Bristol, UK.
- Lowenberg (2002). Assessing English proficiency in the expanding circle. *World Englishes*, 21(3) 431–435.
- Lumley, T., & Brown, A. (1996). Specific purpose language performance tests: Task and interaction. *Australian Review of Applied Linguistics*, 13, 105–136.
- Major, R. C., Fitzmaurice, S. F., Bunta, F., & Balasubramanian, C. (2002). The effects of nonnative accents on listening comprehension: Implications for ESL assessment. *TESOL Quarterly*, 36(2), 173–190.
- Malvern, D., & Richards, B. (2002). Accommodation in language proficiency interviews. *Language Testing*, 19(1), 85–204.
- Mauranen, A. (2003). The corpus of English as a lingua franca in academic settings. *TESOL Quarterly*, 37(3), 513–527.
- McKay, S. (2002). *Teaching English as an international language*. Oxford: Oxford University Press.
- McNamara, T. (1997). Interaction in second language performance assessment: Whose performance? *Applied Linguistics*, 8, 446–466.
- McNamara, T. (2001). Language assessment as social practice: challenges for research. *Language Testing*, 18(4), 333–349.
- Odlin, T. (2002). Language transfer and cross-linguistic studies: relativism, universalism and the native language. In R. B. Kaplan, (Ed.), *The Oxford handbook of applied linguistics* (pp. 253–261). Oxford: Oxford University Press.
- O'Hagan, S. (1999). Assessment criteria for non-native and native speaker essays: Do uniform standards work? *Melbourne Papers in Language Testing*, 8(2), 20–52.
- O'Sullivan, B. (2002). Learner acquaintanceship and oral proficiency test pair-task performance. *Language Testing*, 19(3), 277–295.
- Patri, M. (2002). The influence of peer feedback on self- and peer-assessment of oral skills. *Language Testing*, 19(2), 109–132.
- Poehner, M., & Lantolf, J. (2003). Dynamic assessment of L2 development: Bringing the past into the future. *CALPER Working Papers, No. 1* (pp. 1–26). State College, PA: Center for Advanced Language Proficiency Education and Research (CALPER).
- Prabhu, N. S. (1987). *Second language pedagogy*. Oxford: Oxford University Press.
- Ross, S. (1992). Accommodative questions in oral proficiency interviews. *Language Testing*, 9, 173–86.
- Santos, T. A. (1988). Professors' reactions to the academic writing of non-native speaking students. *TESOL Quarterly*, 22(1), 69–70.
- Saville, N., & Hargreaves, P. (1999). Assessing speaking in the revised FCE. *ELT Journal*, 53(1), 42–51.
- Seidlhofer, B. (2001). Closing a conceptual gap: The case for a description of English as a lingua franca. *International Journal of Applied Linguistics*, 11, 133–158.

- Seidlhofer, B. (2004). Research perspectives on teaching English as a lingua franca. *Annual Review of Applied Linguistics*, 24, 209–239.
- Swain, M. (2001). Examining dialogue: Another approach to content specification and to validating inferences drawn from test scores. *Language Testing*, 18, 275–302.
- Taylor L. (2005, May). Linguistic diversity: Language varieties and their implications for testing and assessment. Paper presented at the Association of Language Testers in Europe (ALTE) Conference, Berlin.
- Timmis, I. (2002). Native-speaker norms and international English: A classroom view. *ELT Journal*, 56(3), 241–249.
- Weigle, S., Boldt, H., & Valsecchi, M. (2003). Effects of task and rater background on the evaluation of ESL student writing: A pilot study. *TESOL Quarterly*, 2(1), 345–354.