King's Research Portal

[Link to publication record in King's Research Portal](#)

From mythical 'standard' to standard reality:

The need for alternatives to standardized English language tests.

Jennifer Jenkins and Constant Leung

Bio

Jennifer Jenkins is Professor of Global Englishes and founding director of the Centre for Global Englishes at the University of Southampton. She has been researching the phenomenon of English as a Lingua Franca (ELF) since the late 1980s, and has published widely on the subject, including three monographs (*The phonology of English as an International Language*, OUP, 2000, *English as a Lingua Franca: Attitude and identity*, OUP 2007, and *English as a Lingua Franca in the international university*, Routledge 2014). Her current research interests are language in higher education, and issues of empowerment and disempowerment relating to ELF. She is founding editor of the book series *Developments in English as a Lingua Franca* (De Gruyter Mouton), and co-editor of *The Routledge Handbook of English as a Lingua Franca* (2018).

Constant Leung is Professor of Educational Linguistics in the School of Education, Communication and Society, King's College London. His research interests include additional/second language curriculum and assessment, language policy and teacher professional development. His current research and development project is concerned with the assessment of English as an Additional Language in school education. He is Joint-Editor of *Language Assessment Quarterly*, Editor of Research Issues of *TESOL Quarterly,* and serves as a member of the Editorial Boards of *Australian Review of Applied Linguistics, Language and Education,* and *Modern Language Journal*. His work on the English as an Additional Language Assessment Framework for schools (Bell Foundation, 2017) has just won the 2018 British Council Award for Local Innovation.

Both Jennifer Jenkins and Constant Leung are Fellows of the Academy of Social Sciences (UK).

1. Introduction

This position paper builds on three of our earlier publications on the same subject (Jenkins & Leung 2014, Leung, Lewkowicz & Jenkins 2016, and Jenkins & Leung 2017), as well as a number of conference papers we've given both jointly and individually. However, what we have not done up to this point is to propose alternatives to the large-scale standardised English tests administered by the major international examination boards, of which we have been so critical, despite the fact that we have been discussing other possibilities among ourselves for several years. The opportunity to publish a position paper on English language assessment therefore provided an ideal opportunity to present our alternatives, and this we do in the final part of the paper.

Our focus is on standardization in respect of one particular kind of language assessment: English language testing for university entry[1]. We chose this because of its overriding gatekeeping function in preventing many candidates from achieving university entry, thus blighting their career prospects and potentially damaging their entire lives on the basis of test scores that are premised on a set of standardized expectations and norms whose claims to validity and relevance have been questioned. More specifically, our general reference points in this discussion are large scale internationally-marketed products such as IELTS (International English Language Testing System) and TOEFL (Test of English as a Foreign Language), the

---

[1] In this discussion 'assessment' is used as a broad covering term referring to both the idea and practice of ascertaining language proficiency, and 'test' as an instrument of assessment.

tests currently used by more universities globally than any other to determine which students they will and will not accept. Having said this, we believe at least some of our points may have relevance to English language testing for other purposes.

Our paper begins with a discussion of the conceptual background. Here, we explore how international English language tests for non-native speakers have, up to now, always been benchmarked to an idealized native English model, itself based on native intuition and/or native English corpora, with a focus on some generic version of 'correctness' and abstracted notions of academic English. In this section, we also consider how critical language assessment scholars have more recently started to move away from such idealisations and even the fixation on native English itself. On the other hand, to date, we argue, the majority have not moved away from an acceptance of generic tests and (some kind of) standardization *per se*.

We go on to discuss the changing world, one of growing mobility, migration, and superdiversity (Vertovec 2007), in which English is used frequently in lingua franca communication (mainly, and often entirely, among non-native English speakers), with diverse, hybrid uses of English and multilingualism increasingly in evidence. We explore the (socio)linguistic implications of these developments, which include, crucially, the ways in which they require comparable changes in language testing for any context where English serves as a lingua franca rather than only as a means of communicating with native English speakers, and we argue that global higher education provides a particularly strong case in this respect. In the section following on from this, we consider the social justice dimension as it relates to higher education in ways that, with a few exceptions from critical language assessment scholars such as Shohamy (e.g. 2006, 2011, 2017), have tended to be explored within a set of narrow top-down considerations of 'avoiding bias' and 'level playing fields'.

Finally, we present our own alternatives to the language assessment status quo. These alternatives are grounded in the theoretical position we have been enumerating for some years, i.e. that

> … the use of ELF [English as a lingua franca] involves speakers from diverse linguacultural backgrounds [who] use ELF to communicate with one another, to get things done, and to socialize. Therefore the language assessment issues raised by ELF *transcend questions of proficiency conceptualized in terms of a stable variety*; they are concerned with what   counts as effective and successful communication outcomes through the use of English that can include emergent and innovative forms of language and pragmatic meaning (Jenkins & Leung 2014:1610; italics added)

To this, we would add that the most recent conceptualisation of ELF emphasises the multilingual nature of the phenomenon: that for all but monolingual NESs (native English speakers), ELF users are oriented not only to English, but also to the other languages in their multilingual repertoires; therefore, that although English is available to all present, it is not necessarily chosen as the only language appropriate to a particular interaction (spoken or written). Rather, translanguaging is a key feature of ELF communication (see e.g. Jenkins 2015 on the notion of 'English as a multilingua franca', García & Li 2014 on translanguaging). This essential multilingualism of ELF, we maintain, also needs to be incorporated into assessment frameworks.

In presenting our alternatives, we thus move the debate to a new level by arguing that standardized/generic testing of English for lingua franca communication needs to be replaced

with contextualised, socially realistic, and socially fair means of assessing candidates' English language abilities. The time has come, we argue, to abandon testing candidates in tests claimed to be 'international' against any kind of stable variety of English, or even against English only, for future communication in lingua franca contexts.

2. The conceptual background

The large scale internationally-marketed standardized English language tests have tended to be built on a stable portrayal of the English language. That is not to say that the international testing organizations have not taken account of some aspects of language variation such as regional accents. IELTS, for instance, has gone to considerable lengths to incorporate different English accents from different parts of the English-speaking world into its listening tests (http://ielts-academic.com/2015/10/31/ielts-listening-english-accents/). The stability at issue here is concerned with the idealized and typified ways in which English is represented as a medium of communication in real-life contexts, with particular reference to the use of English in academic settings within English-medium institutions. As Harding & McNamara (2018) point out, the international English language assessment industry appears to be quite insulated and slow in response to changes and developments in contemporary language practices. There are a number of possible commercial and operational reasons for this apparent indifference to change. For example, conceptualizing English as a stable and enduring phenomenon provides for a long(er) shelf-life for language tests as products and obviates the need for regular revision and re-development, which is expensive. For the present purpose our attention is on the aspects of the conceptual hinterland of language testing that can be linked to this stasis observed by Harding & McNamara and others.

In the past 30 years or so claims of validity or meaningfulness of a test have been linked to the notion of construct, the fundamental tenets of which have been strongly influenced by the articulation of Messick (1989; also see McNamara 2001, Kane 2006,; for a longer view see Newton & Shaw 2014). It would be fair to say that construct is both a conceptual frame and a principle for operationalization. One of the sources of authority on these matters is the *Standards for educational and psychological testing* (AERA, APA,NCME, 2014:23), which states that 'the construct or constructs that the test is intended to assess should be described'. From the point of view of psychometric measurement, Wilson (2005:28) argues that a 'construct is always an ideal; we use it because it suits our theoretical approach'. In the field of second/additional language testing, construct has been characterized as THE NATURE OF THE KNOWLEDGE AND ABILITY WE WANT TO MEASURE, BY DEFINING IT ABSTRACTLY (adapted from Bachman & Palmer 1996:89). This view is further elaborated as follows:

> … we can consider a construct to be the specific definition of an ability that provides the basis of a given assessment or assessment task and for interpreting scores derived from this task. The construct definition for a particular assessment situation becomes the basis for the kinds of interpretations we can make from assessment performance (Bachman & Palmer 2010:43; see also Green 2014 Part 3).

On this view, it can be implicitly assumed that the construct, once defined and operationalized (at varying levels), is a quality that resides in the individual test-taker, and the assessment task performance is caused by the putative construct within the individual test-taker. So there is a causal relationship between construct and test performance: the construct, as something residing in the individual, CAUSES the performance.
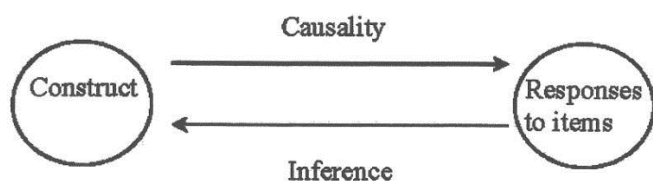
FIG. 1.4 A picture of the construct modeling idea of the relationship between degree of construct possessed and item responses.

(Wilson 2005:13)

When applied to the field of L2 assessment the chain of reasoning embedded in these statements can be presented as follows:

A construct is an idealized and abstracted statement of the ability to be assessed

→the focal ability, as defined by the construct, resides in the individual test-taker

→the assessment task is the operationalized representation of the construct that taps into the focal ability

→the assessment task is conceptually located in a Target Language Use (TLU) context which is specified, e.g. the use of language in a particular occupational or academic setting (see Bachman & Palmer 2010 Chapter 15 on TLU)

→the task and the ability, once specified, are assumed to be context-independent

→the focal ability, as defined by the construct, is actualized in assessment task performance by the test-taker

→the test task performance of an individual test-taker can be used as evidence of their ability in similar tasks in future.

This complex chain of reasoning can be summarized as follows:
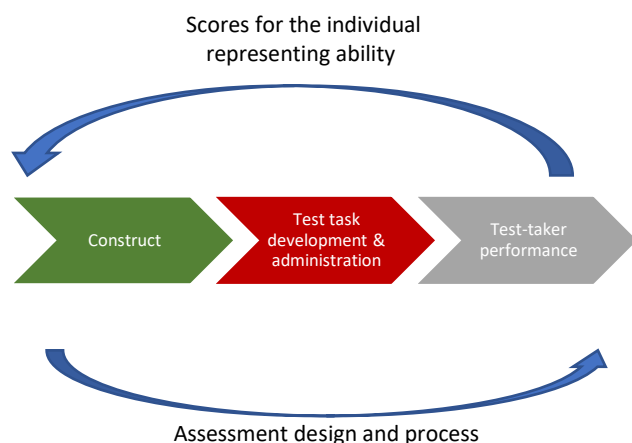
Some elements of the idealization and abstraction process can be seen in two accounts of test (re)development. Something of this propensity to 'see' commonality in language use across different disciplinary domains is found, firstly, in the revision of English Language Testing Service in the late 1980s (ELTS, which became IELTS). The revision was prompted by the perceived need to simplify the ELTS which, *inter alia*, had six separate academic areas for the assessment of discipline-based study skills (e.g. reading comprehension in Life Sciences, Social Studies, Physical Sciences etc.). Weir & O'Sullivan (2017:194–195) report that the revision team found 'there was much in common in the study skills across the disciplines'. And in the process of reducing the six to three academic areas (Physical Sciences and Technology, Life and Medical Sciences, and Arts and Social Sciences) it was found that 'tasks across the three remaining academic subject areas were so similar that the specifications were virtually identical and only differed in respect of the reading texts they employed'.

The revision and re-development of the TOEFL assessment of academic writing provides a second example. Cumming and his colleagues, at the outset of the process, adopted the following conceptual frame for academic writing:

… written texts are produced both in and for specific social contexts, involving communities of people who established particular expectations for genres and standards for writing within these communities, both locally and universally… (2000:4).

The empirically grounded view expressed in this statement accords with the findings of many researchers in the field of academic writing and academic literacy more generally (e.g. Swales 1990; Hyland & Hamp-Lyons 2002; Hyland 2004, , Lea & Street 2006, Lillis & Scott 2007, , Wingate 2016). The transition to test development, however, signals idealization and abstraction: '… our conceptualization of academic writing can be … presented in terms of … task stimuli, rhetorical functions, topic characteristics and evaluative criteria' (Cumming et al. 2000:5). Terms such as 'topic characteristics' and 'rhetorical functions' reflect the move away from 'communities of people' with their writing practices and expectations. This once-removed position supports the next step: to operationalize the test domain in terms of requiring '… students to produce and sustain in writing coherent, appropriate, and purposeful texts in response to assigned tasks' (op.cit.:7). The following is an example of the 30-minute writing topics (300-350 words):

'Some people think it's better to live with a roommate. Other people prefer to live alone. Which do you prefer? Use specific reasons and examples to support your answer.' (TOEFL Independent Writing   https://www.englishclub.com/esl-exams/ets-toefl-practice-writing.htm)

Since 'big' tests such as TOEFL and IELTS are often used for university admission purposes, it is not trivial to note that most university writing tasks are considerably longer than 300 words, that tasks and topics are routinely based on disciplinary content, and that most university discipline-related writing tasks are carried out over a much longer stretch of time (except for written examination papers). This observation is particularly related to content aspects of validity. So it is quite clear that the imperatives of idealization and abstraction have acted on the initial conceptual frame to enable the designing of test tasks that are, in the end, far removed from the practices of real communities of writers. Furthermore, a construct, by virtue of the conceptual parameters of its creation, is imbued with pre-determined and pre-specified individual test-takers' ability/ies and circumstances of language use (ahead of any real-life TLU). All of this takes us straight to the next issue.

The process involved in defining a construct, specifying the TLU, abstracting from complex and often divergent real-life tasks to form idealized assessment tasks, and using the rating/score of assessment task performance to represent an individual test-taker's ability, involves a form of reification in two senses. Firstly, a construct is an idea and an artefact because, in the particular case of large scale standardized English language tests, it is literally created by language assessment professionals. The logic of idealization and abstraction facilitates the production of test tasks that yield performance scores. The scores then appear as a quantitative measure of an individual test-taker's ability. So what starts out as an idea, through a series of transformations involving task setting to task performance, emerges as a score of language ability that carries the appearance of precision and thing-like certainty. The real-life impact of this reified outcome of an intellectual and professionalized process on university and job applicants is well-recognized. Secondly, the theoretical and individual-focussed

measurement orientation requires that real-life situated language use is stripped of its situated variability and contingencies (due to idealization and abstraction). Furthermore, it is specified in advance, and ability is defined in terms of individual performance. However, there is now a substantial body of research literature on academic writing and academic literacy more generally demonstrating: that academic writing varies across disciplines and institutions, that genres and other writing conventions are not immutable (see our references above on this point); that academic writing tasks and associated literacy activities (e.g. reading and discussion on reading) are shared activities; and that students do not necessarily learn on their own (e.g. Gee 2004, James 2006). Seen in this light, the theoretical and measurement concerns have helped reify a much more complex reality. Such reification can have significant negative impact on validity claims.

Another important issue notable for its absence in the English language assessment literature is language development and change. The current L2 assessment literature is informed, broadly speaking, by the ideas associated with the advent of the 'communicative turn' in Anglophone applied linguistics in the late 1970s and early 1980s. The seminal article in the inaugural issue of *Applied Linguistics* by Canale & Swain in 1980 entitled 'Theoretical bases of communicative approaches to second language teaching and testing' can be seen as a landmark publication representing a moment of paradigmatic shift from a grammar-based orientation to a more socially-sensitive view of language. The English language assessment literature, informed by the Anglophone scholarship in applied linguistics, routinely addresses issues related to social conventions of use (sociolinguistics) and cultural meanings in language use (pragmatics) (for an elaboration, see Leung 2011, 2013). However, such discussions have tended to be conducted from the perspective of an assumed standard language variety (usually American or British) and well-to-do mainstream culture (for a wider discussion, see Gray 2010). Social conventions of

use and culture-sensitive meanings in the English language are considerably more diverse than what is currently modelled in the established language tests. Furthermore, there has been a muted response in the English language assessment literature to the widespread use of ELF in different parts of the world for business, educational, industrial and scientific purposes, and in supra-national public organisations (e.g. World Bank) and governmental institutions (e.g. ASEAN). This is the issue to which we turn next.

3. Changing English-speaking world, changing English-speaking university

It is self-evident that the English language has spread over recent decades well beyond its mother-tongue regions and the post-colonial countries of Kachru's (1985) 'outer circle', to most of the rest of the world, or Kachru's 'expanding circle'. Whereas scholarship in the field of World Englishes from the 1970s has long established (if not without a fight) the linguistic rights of the populations of post-colonial states such as India, Nigeria, Singapore and the like to use their own Englishes, and for these to be accepted as legitimate ways of speaking English, the same cannot be said for the rest of the world. Instead, the global ELT industry led from the US and UK continues to thrive while ignoring the sociolinguistic reality around it. Thus, it continues in the main to present native NESs as ideal teachers, non-native English uses as by definition 'errors', and the purpose of ELT to enable non-native English speakers (NNESs) to communicate with NESs. Hence, alongside their idealized versions of formal native English, ELT materials also present native English idiomatic language which, ironically, they often refer to as 'real' or 'real-life' English.

Even ELT materials whose purpose is to prepare students for international higher education, a site of ELF communication *par excellence*, tend to assume that the English they will need, whatever and wherever they study, will be oriented to communicating with NESs and to native English varieties, and will to a great extent be generic. As an example, one current EAP course states the following on its website:

> *Oxford Grammar for EAP* is a grammar reference and practice book whch provides students with the functional grammar they need to succeed in their academic studies, *whatever their chosen subject* (elt.oup.com, accessed 20 November 2017; our italics).

Likewise, the website for the new *Oxford Academic Vocabulary Practice Lower Intermediate* and *Upper Intermediate*, states '[v]ocabulary practice activities help you learn *the key words* you need to use when studying *any academic subject in English at university lev*el' (elt.oup.com, accessed 20 November 2017; our italics). Although the series as a whole makes use of a corpus drawn from four broad academic areas (physical sciences, life sciences, social sciences, and humanities), this is as far as it goes in terms of addressing local disciplinary and linguistic nuance. Meanwhile, the testing of English according to monolithic standard native English 'norms', a 'phantom', as we call them elsewhere (Leung and Jenkins 2018), expands apace, with IELTS reporting increased numbers of candidates and centres year on year.

The status quo in ELT, EAP, and English language assessment is thus deeply unrepresentative of the modern world for which it claims to be teaching and testing English. Apart from minor concessions to the existence of other kinds of English than native made, for example, on the teacher training syllabuses of bodies such as Cambridge Assessment English – who nevertheless tend to mistakenly conflate ELF with World Englishes – there is little or no

engagement with ELF at the practical level (see Jenkins 2015 for the distinction between ELF and World Englishes). Instead, ELT continues to be informed by mainstream SLA research, whose unspoken domain assumption over many decades has been, and remains, to enable learners to 'achieve' nativelike language, with English being the language most frequently used for exemplification.

In this respect, Selinker's (1972) interlanguage theory/continuum is still held in high regard along with the notion that where the L1 differs from L2 English, it will impede the acquisition of English. As Murray observes, both the teaching and testing of English thus continue to '[perpetuate] Selinker's (1972) concept of interlanguage as comprising stages of development, or approximative systems, that increasingly reflect … native-speaker competence' ignoring 'the reality of a world increasingly characterised by multicultural, multilingual interactions' (2018:57-58). To cite but one of many possible examples from current SLA research, a talk was given in London in November 2017 with the title 'What Interlanguage analysis reveals about L2 referent tracking', whose abstract begins as follows:

> Recognizing the learnability problem that the English article system presents for second language (L2) learners, the SLA field has taken a particular interest in documenting its acquisition, especially among learners whose first languages (L1) lack articles (Ekiert 2017).

Over the decades, it is fair to say that SLA research has been extensively developed and refined, and many new theories proposed (and in some cases, abandoned). However, the key premises underlying interlanguage theory still remain firmly in place. These include the notion that the L1 speaker is the only desirable target for L2 learners (including learners of English); that any

differences between the L1 and L2 need to be explored in order for teachers to be able to

eliminate L1 transfer effects and other 'problems' arising from these differences; and that if

learners continue using 'non-target' forms after formal learning has ended, their language has

'fossilized' in these respects. Yet, as far as English is concerned, as McNamara points out:

> …the growing awareness of the nature of English as a lingua franca communication
>
> overturns all the givens of the communicative movement as it has developed over the
>
> last 30 or 40 years. The distinction between native and non-native speaker competence,
>
> which lies at the heart of the movement, can no longer be sustained; we need a radical
>
> reconceptualization of the construct of successful communication that does not
>
> depend on this distinction (2014:21).

Not surprisingly, then, from the perspective of ELF, influence on L2 users' English from their

L1 is considered normal and natural, not something that interferes with the acquisition and use

of English. Thus, for ELF, rather than talking of L1 'transfer' or worse still, 'interference', the

focus is on 'similects' and 'second order contact' (Mauranen 2012). In other words, it is

inevitable and unremarkable that people's L1s have some degree of influence ranging from

slight to heavy on their L2 English. However, as Mauranen also points out, L2 English users

don't habitually speak English with each other. Instead,

> … ELF takes shape in speaker interaction; interactants come together with their own
>
> hybrid variants [ie similects], that resemble those of people who share their
>
> background (that is, who speak their similect) but are different from those used by the
>
> people with whom they speak … Therefore ELF might be termed 'second-order
>
> language contact': a contact between hybrids (2012:29).

In other words, the development of an individual's ELF use depends heavily on the SPECIFIC second order contact in which he or she is involved, ranging from more established ELF communities of practice, e.g. groups of doctoral students from a range of L1s who meet regularly for seminars, to transient ELF encounters where interlocutors have never met before.

Underpinning all ELF research at least to some, and often to a great, extent is the phenomenon of accommodation, or the ability to make both productive and receptive adjustments to speech and writing, primarily to promote mutual intelligibility for interlocutors. This was first documented by Jenkins (2000) in respect of ELF pronunciation, and has since been taken up by all key ELF researchers. Accommodation can be pre-emptive, i.e. the speaker (or writer) uses an alternative in place of an item s/he considers potentially unintelligible for his or her interlocutor(s). Alternatively, it can occur immediately after a problematic item has been uttered. For instance, a lecturer giving a talk to an audience of which the majority were NNES staff and international students used the phrase 'we've got bigger fish to fry', but immediately paraphrased it as 'so we have more important problems' (our data). On the other hand, the problem may only be identified, and accommodation attempted, if an interlocutor indicates non-understanding. Failing this, the outcome is likely to be non-understanding on the part of the receiver(s): an outcome that has been shown to occur more often in ELF communication when the speaker is an NES. This is something that has so far not been addressed in any existing standardised English language assessment where, for example, the use of (potentially unintelligible) native English idiomatic language tends to be rewarded rather than penalised. As mentioned in our Introduction, more recently ELF research has also recognised the key role of translanguaging in effective ELF communication. Both accommodation and translanguaging thus figure in our alternatives to standardised English language assessment.

In addition to the accommodation data, there is now a good deal of published data showing myriad kinds of adaptations and innovations in ELF usages in terms of both linguistic form and meaning at lexical and discourse levels. The following two spoken examples are drawn from Mauranen (2012:102):

'… nothing is *guarantable*, the quest for theoretical certainty …'

''... it's only eight per cent in Slovakia they er they are *in front of* us in regards social and economic reforms'

The standard native English form for 'guarantable' is 'guaranteed'. In the second example above there is a semantic shift in the use of 'in front of'. The more conventional (i.e. native English) phrase would be 'ahead of'. As Mauranen observes, judging from both the co-text surrounding these utterances and the extended context, the non-standard forms/usage were not repaired by either speaker or listener, but "passed unnoticed" (ibid.) and did not cause any communication problems". The same was true of numerous other examples provided by Mauranen (ibid.).

The following extract of talk among international students in a university setting is an example of the complex negotiation of meaning that involves sharing multilingual resources and expansion of semantic possibilities. The data is drawn from Batziakas (2016:138–139). The students involved were in in meeting of a student society in a London university. The discussion was concerned with finding a suitable student to represent their college in an inter-

collegiate event. The participants were: Arvin - Mauritian Creole speaker, Breno - Portuguese, Eshal - Urdu, José - Spanish, Linlin - Mandarin Chinese (all pseudonyms).

Transcription key:

=	Latching
?	Question
(.)	Brief pause
(time in seconds)	Longer pause
BOLD text	Focal expression (for analysis)
↑	Speaker expressed enthusiasm
<u>Underlining</u>	Speaker emphasis

1	**Linlin**	and you know why? (.)
2		because sometimes in all these induction days and freshers fayres
3		i see people who approach you and talk to you
4		and they're they're bad with what they're doing=
5	**Eshal**	=yeah you're right=
6	**Linlin**	=so i wouldn't like someone who can't
7		you know
8		who can't do this or do that
9		(1.2)
10	**Arvin**	[yeah]
11	**Marat**	[true]=
12	**Linlin**	=so I wouldn't like someone who is who is (.)
13		ah in china we say ah haha **diaosi**
14		(.)
15	**Breno**	hm?
16	**Jose**	what?
17	**Linlin**	oh i mean you know **diaosi** (1.4)
18		ah ah in english i think perhaps
19		if there is this word=
20	**Arvin**	=so what's this word? what do you mean?
21	**Linlin**	<u>diaosi</u> (.) someone who is average and normal (1.3)
22		who can't do anything can't manage anything (.)
23		like good for nothing (.)
24	**Arvin**	is he someone loser then?=
25	**Linlin**	=a loser? **diaosi** is not a loser it's not a loser definitely not

26	Jose	is he in spanish we say haha perdedor?
27		like someone who can't manage things and people
28		(.)
29	Linlin	i don't know this word this language i mean
30		but nuh it's not what you said (0.2)
31		you know it's just **diaosi**
32		**diaosi** and nothing else
33		(.)
34	Arvin	ok↑ i get you↑
35		no **diaosi** will be selected
36		and do we all agree no **diaosi** will ever represent the society?=
37	Breno	=[ok]
38	Jose	[yes]=
39	Linlin	=thanks↑
40		yeah it's better this way no=
41	Arvin	=no diaoshi
42	Linlin	thanks↑

In this stretch of talk Linlin used the term 'diaosi' (pinyin, 屌丝) from Putonghua (Mandarin Chinese) to denote a particular (unsuitable) personal quality for the task at hand. Linlin must have been aware that this Chinese term was not known to the other students present in the meeting. It would seem that she wanted to express her idea succinctly and did not think there was an equivalent term in English. So the introduction of the term (line 13) might have triggered unexpected diversions. Instead, the other students, all from different language backgrounds, engaged with the use of this unfamiliar term and asked for a gloss of this term. After Linlin had rendered the meaning of 'diaosi' (lines 21–23) in English, José offered a possible equivalent in Spanish 'perdedor' (line 26) with a translation into English. This led to further negotiation of the meaning of 'diaosi'. At the end of this exchange all involved

appeared to have understood the meaning of 'diaosi', and furthermore 'diaosi' was incorporated into the joint decision-making, as signalled by Arvin (line 41).

The question we need to ask at this point, comes down to what we actually mean by 'English'. And it should by now be obvious that we see a need to distinguish between those contexts in which English is used by NESs among themselves, and those contexts where English serves as a lingua franca. In the former case, it seems reasonable to expect certain codified/established conventions to be acknowledged and deferred to, although even here, local context and language change over (relatively short periods of) time will override any all-purpose native English norms. In the latter case, ELF, there is no codification or established convention that can be deferred to, and the focus is entirely on effective communication skills in context. In this respect, numerous studies of ELF interactions drawn from ELF corpora such as VOICE (Seidlhofer 2001) and ELFA (Mauranen 2003) have demonstrated at both macro and micro level what kinds of phenomena are involved (see, e.g., Seidlhofer 2011, Cogo & Dewey 2012, Mauranen 2012, Pitzl 2018). Ironically, English language tests such as IELTS expect the kinds of English used by NESs AMONG THEMSELVES to be produced by NNESs when the testers should, instead, be assessing NNESs' readiness to operate and convey their meaning in a prime ELF setting: that of a specific international university programme.

From the above discussion, it will be clear that we believe current English language examinations are testing people for things they don't need, and not testing them for things they do need in this increasingly mobile, superdiverse world. In such a world, NNESs are most likely to find themselves communicating with multilingual English users from other first languages in both established groupings and transient encounters, and they, as well as NESs, need assessing in respect of their readiness to do so, not on their ability to reproduce idealized native English

forms. This includes readiness to engage with the specific literacy practices within the candidate's specific target discipline, meaning that divergent literacy practices across different disciplines also need to be factored into the complex equation (see Wingate 2015). And in all these respects, the assessment status quo is not only inappropriate, but also unfair and unjust. In the next section, we explore the unfairness and injustice inherent in current English university entry tests, insofar as they gatekeep and discriminate on a false prospectus, while also causing NNES candidates to waste time acquiring irrelevant English language forms and skills, and ignoring the transcultural needs of NESs.

4. Ethical issues of justice and fairness

It is now common knowledge that the large scale standardized academic English test scores do not strongly correlate with test-takers' subsequent academic performance. Most validation studies report weak and inconsistent correlations (e.g. Cotton & Conrow 1998, Ingram & Bayliss 2007, Lee & Greene 2007, Cho & Bridgeman 2012), with a small number of exceptions (e.g. Yen & Kuzma 2009, Harrington & Roche 2014, the latter involving an institution specific test). In the final chapter of his 2018 book on evaluating language assessment, Kunnan observes that the book's 'primary purpose … is to address two fundamental questions relevant to language assessment: (1) What's the right thing to do to bring about fair assessments and just institutions and (2) What's the right thing to do to remove manifest unfairness and injustice?' (p. 241). We agree entirely with Kunnan's questions, although our conclusions are somewhat different, especially with reference to ELF and language modelling. In this section we therefore consider our own theoretical position on fairness and justice, then go on in the final section to propose our alternatives to (any kind of) standardized English language assessment for

university entry, which we consider both fairer and more just. But first, we discuss what others have said on the subject. We start by considering some issues of fairness and social justice in education more broadly, then turn specifically to language assessment: first, to mainstream approaches, and second, to critical approaches.

From her investigation of a potential link between linguistic diversity and social injustice, Piller concludes that there is a 'collective failure of imagination when it comes to linguistic diversity: the failure to recognize that linguistic diversity undergirds inequality too frequently and the failure to imagine that we can change our social and linguistic arrangements in ways that make them more equitable and just' (2016:222). Our own 'imagination', to borrow Piller's term, has prompted us to seek a new way of evaluating prospective students' suitability for university study, one that rewards rather than penalizes their linguistic diversity in respect of both their use of English and their multilingualism; and one that ends the equation of EMI (English medium instruction) around the world with English according to the 'standard' English of NESs from just two Anglophone countries. Unless such an attempt succeeds, English will remain 'a key mechanism to entrench global inequalities' (op.cit.:165), with both NESs and those NNESs whose English is more 'nativelike' continuing to be privileged, and those NNESs whose English is less 'nativelike' continuing to be discriminated against.

In this regard, Piller's observation that 'schools have maintained their traditional monolingual institutional habitus in the face of students' (and, increasingly, teachers') multilingualism', and that there is therefore an 'entrenched mismatch between schools with a monolingual habitus serving linguistically diverse societies' (op.cit.:120, 127), holds equally true for tertiary education in EMI universities, particularly, although by no means exclusively, in Anglophone settings. For despite the obvious fact that outside the Anglophone context, the

home language is not English, and thus, the kind of English used locally is, by definition, not native English, it is also a fact that most international universities in the non-Anglophone world subscribe to the ideology of NESs' global 'ownership' of the English language and role as guardians of its acceptable use. In other words, as far as their use of English is concerned, those who determine language policy in non-Anglophone EMI universities could be described as 'complicit' in the negative stereotyping of their own English, as Lippi-Green put it twenty years ago (1997:242), as well as the English of their students and prospective students. Edwards noted still earlier that 'this 'minority-group reaction' is a revealing comment on the power and breadth of social stereotypes in general, and on the way in which these may be assumed by those who are themselves the object of unfavourable evaluation' (1994:99). However, in the case of English, we are talking not of a minority, but of NNESs, who vastly outnumber NESs globally, including on many university programmes even in Anglophone settings. The problem, as Li observes, is that 'the myth of a pure form of a language is so deep-rooted that there are many people who … cannot accept the 'contamination' of their language by others' (2017:6). They then extend the contamination metaphor to their perspective on English, ignoring its diverse global reach, hence their widespread negative stereotyping of their own and fellow L1 speakers' English as 'contaminated'.

The 'NES ownership of English' perspective is thus deeply anachronistic. And as ELF and critical multilingualism research has been demonstrating for the past two decades, in today's mobile, linguistically-diverse world, to be an effective English user in ELF communication settings, where most, and often all, participants are NNESs, it is a distinct advantage to be able to accommodate to speakers from a range of first language backgrounds and to have the ability to translanguage. By contrast, it is a distinct *dis*advantage not to have the accommodation skills to understand and be understood easily in ELF communication, and to be

monolingual, both of which have been found to characterise many NESs in intercultural communication. The educational establishment more broadly, and the language assessment establishment more specifically, however, have yet to catch up and to acknowledge these linguistic truths of 21st Century life. Thus, 'monolingual ideologies still dominate much of practice and policy, not least in assessing learning outcomes. The actual purpose of learning new languages – to become bilingual and multilingual, rather than to replace the learner's L1 to become another monolingual – often gets forgotten' (Li 2017: 8).

The NNES advantage to which we referred in the previous paragraph is not intended to minimise in any way the problem currently facing NNESs. And even if the golden age finally arrives when linguistic diversity and translanguaging in and out of English are accepted by high-stakes institutions, there remains the fact that NNESs will still have to function in a language other than their mother tongue, while Anglophones represent what Van Parijs calls 'free riders' on the cost of the language learning of NNESs (2011:50). But this is also not to suggest that there will never be a cost to NESs. The time will come, we believe, when NESs will need intercultural communication skills such as accommodation and the use of multilingualism to enable them to communicate more effectively in professional (including academic) ELF contexts. So eventually, we believe, the 'free riding' will come to an end. By the same token, it is sometimes argued, in line with Bordieu & Passeron's (1977) observation that academic language is nobody's mother tongue, that NNES academic writers are not disadvantaged in relation to NESs. Hyland (2016), for example, claims that it is a 'myth' that there is any injustice to NNESs in this respect. This ignores the obvious fact that it is easier to acquire academic language if your starting point is another version of that language than if it's a different language altogether. And yet if the time comes when ELF communication is better understood and its legitimacy widely acknowledged, the corresponding shift towards acceptance

of diverse English use will lessen the cost (both practically and metaphorically) for NNESs, as they will no longer be obliged to struggle to mimic native English. Meanwhile, it will also lead to the lessening of another kind of linguistic injustice mentioned by Van Parijs: that the privileges given to English mean that equal respect isn't shown to the other languages of the population, which, in the context of our present discussion, means the other languages of NNES university students (as well as NNES staff). Once NESs realise that they need other languages and translanguaging skills for their academic and professional lives, this type of injustice is likely to diminish too.

At this point we turn to some of the relevant ideas and arguments from the field of language assessment that connect with the broad educational ethics related issues discussed above. In many ways ethical issues such as justice and fairness have received a good deal of attention in the language assessment literature. Justice and fairness are closely linked to validity, particularly since Messick's (1989) discussion on validity as a unified concept that embraces, *inter alia*, social consequences of assessment (see, for example, McNamara 2001, 2005, Shohamy 2001). Perhaps the following professional benchmarks presented in the 2014 edition of the *Standards for educational and psychological testing* (AERA, APA, NCME, 2014) can serve a useful reference point:

On validity:

Standard 1.0

'Clear articulation of each intended test score interpretation for a specified use should be set forth, and **appropriate validity evidence in support of each intended interpretation should be provided**.' (p23; our emphasis)

Standard 1.1

'The test developer should set forth clearly how test scores are intended to be interpreted and consequently used. The population for which a test is intended should be delimited clearly, and **the construct or constructs that the test is intended to assess should be described clearly**.' (p23; our emphasis)

On fairness:

Standard 3.0

'All steps in the test process, including test design, validation, development, administration, and scoring procedures, should be designed in such a manner as **to minimize construct-irrelevant variance and to promote valid score interpretation for the intended uses for all examinees in the intended population**.' (p63; our emphasis)

On the face of it we might say that since we have these Standards in place, we have the necessary intellectual accoutrement to address any deficiency in practice; in other words, all we need is more and/or better informed practice. Unfortunately, matters are a little more intractable than they seem on the surface. There are two intertwined aspects to the Standards: application/administration and conceptual/theoretical framing. In terms of application/administration, valid and fair assessment can be achieved through clear articulation of, and adherence to, procedures, e.g. standardized control of administration, interpretation of performance and scoring processes. To the extent that conforming to common processes can help reduce or avoid (unintended) biases and disadvantaging some test-takers, this aspect of the Standards seems reasonably justifiable in the name of universalism. This resonates with Taylor's (1994) notion of equality of entitlement, whereby society will provide the same to all irrespective of their diverse needs.

Conceptual/theoretical framing is less straightforward. On the one hand, validity and fairness can be established by showing evidence that the pre-defined construct/s and other related validity parameters (e.g. content) have been observed. If validation is framed in this way, then a certain circular reasoning is involved: Construct X is valid because of Y (we define it thus); if Y is thus defined, then X is valid. The principles outlined in Section 2 above regarding test measurement approximates this reasoning. We can describe this approach as tight and closed framing. On the other hand, if examination of validity and fairness is framed more loosely and admits alternative models and formulations, then clear specification and application of any adopted construct would only be a secondary issue. A primary concern would be to establish what counts as an appropriate construct/s, a matter of considering and evaluating the suitability and appropriateness of alternatives and the necessity for divergent conceptualizations and practices. As Rawls (2001) recognises, differences in society cannot be avoided, so the task is to find a fair way to co-operate to achieve justice. This would accord with Taylor's (1994) notion of equality of treatment, whereby society recognizes the diverse needs of different groups/individuals and responds accordingly. Kane (2010) provides a relevant legal analogue in relation to fairness in the US context. In this case fairness is understood in terms of two kinds of due process. Procedural due process requires that everyone is treated the same way generally in terms of entitlements and protection of rights within the US Constitution. We take this to be an analogue to application/administration in terms of language testing. Substantive due process requires that the treatment to be applied is reasonable and appropriate in general and in the context of application. In relation to language assessment, CONTEXT is the operative term here.

As we can see from the Standards above, contextual differences related to populations and use of assessment outcomes are well recognised. The requirement of specificity of context

and use embedded within the Standard statements is strongly suggestive of plurality and multiplicity. The case for loose framing is inherent in the Standards, but they have not been openly articulated. We would argue that the call for greater validity and the course of justice and fairness would be better served if we move away from the monolithic universalism premised on a particular variety of English that drives much of current standardized academic English testing. Figure 1 below provides a Toulminian schematic summary of our discussion thus far:

Principle:

Multiple assessment constructs for English language & discipline-related academic language & literac practices

Claims:

No need for monolithic model of academic English

Recognize situated academic language & literacies practices

Warrants:

Warrant

Backing:

Data on student experiences in different kinds of provisions

Long-term and consistent research findings showing diverse language & literacy practices in different EMI contexts

Long-term and consistent research findings showing universalist constructs do not provide for adequate assessment
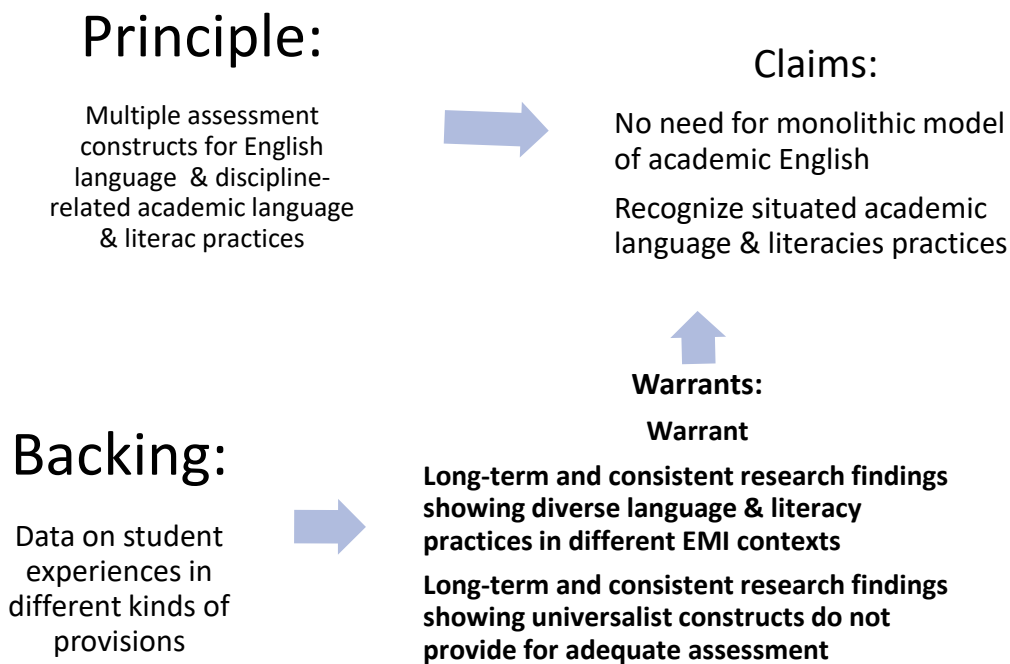
Figure 1   Diverse constructs for language and literacy practices

The principal line of argument is that we need to promote multiple assessment constructs (and designs) to reflect the diverse language and literacy practices that exist in university. This diversity has been clearly and unambiguously demonstrated by long-term research data in academic language and literacy studies and actual accounts of student experiences (e.g. McNamara et al, in press). In addition, as discussed above, validation studies investigating the predictive power of the large-scale standardised academic English tests for

test-takers' subsequent performance in university have yielded only low to moderate correlations. The case for further development is clear.

5. Alternative ways of assessing English for university entry

And so to the finale. Having argued forcefully against the use of 'one-size fits all', standardized English language tests to determine prospective students' readiness for university entry regardless of local context, we will now suggest what we believe should replace them. Piller's (2016) point discussed in the previous section is precisely why we believe the current paradigm is non-viable, and that we should call time on the current practice of standardization based on inadequate and inappropriate language models and norms. We live in a world where linguistic diversity is the norm, and yet students trying to gain entry to international/English-Medium-instruction (the two are usually synonymous) universities, are penalized in the current entry tests for their linguistic diversity in respect of both their use of English and their multilingualism.

In his abstract for a recent talk, McNamara (2017) observes: 'it is remarkable that few if any language tests exist specifically directed at measuring competence in English as a Lingua Franca communication', asks '[w]hat values underlie the resistance of our field to the testing of English as a Lingua Franca?', and criticises 'the fundamentally value-driven and political character of language testing'. We take the argument still further than that of McNamara and other ELF-supporting critical language assessment scholars by arguing that the time has come to abandon any conventional notion of a universal standard. The focus instead, we argue, should be on the individual local context and what is standardly (sic) expected in this respect,

which involves considering the ways in which English is actually used in each individual setting. In this final part of our paper, we explain how, in our view, this goal could be accomplished in both the medium- and longer-term.

This is not to say we suggest dispensing with any kind of measuring of proficiency. As Carlsen (in press) rightly observes, weak correlations between test scores and future academic performance should not be interpreted as meaning that language proficiency is unimportant. Our point is that we need to distinguish between standardization and proficiency and, by extension, between different kinds of proficiencies. We agree with the principle that proficiency should be assessed in relation to specific context of use (e.g. the Standards cited in Section 4 above). This means assessing candidates in respect of their purpose in using English in the focal context and their readiness to do so. For example, even within a single UK institution such as the University of Southampton, a NNES student planning to study engineering, where most other students are likely to be international students from a range of first languages, would need to be 'ready' in a different way from one planning to study English literature, where most other students are likely to be NESs. In other words, we need to take into account a range of considerations including national/local language environments, disciplinary specialisms, institutional curricular requirements, pedagogic approaches, and student cohort compositions. And because of the global spread and contingent diversity of English use, proficiency shouldn't be measured in relation to 'standard' native speaker versions; they are irrelevant to the majority of Higher Education (HE) contexts where NNESs study. In any ELF context, of which international HE is a prime example, the key criterion can only be successful communication in situ. This involves not mimicking a particular variety of native English, but reciprocal intelligibility and rapport in relation to curriculum-related activities. Accommodation skills are therefore paramount: the ability to adjust (pre- or post-emptively) what is said or written and

what is received for the benefit of the SPECIFIC academic interlocutor(s) or reader(s) – skills that are needed by both NNESs and (probably even more so) NESs, whose English skills should also be assessed in such respects.

Some years ago, Bachman & Palmer developed the notion of 'test usefulness', which they present in detail in their 2010 book with the following title: *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Their attempt to model TLU (see also section 2 above) on real life activities in specific domains is very welcome. Equally welcome is their appreciation of the potentially life-changing (including potentially threatening and unfair) consequences of high-stakes tests such as IELTS. In contemporary English-medium international universities we now see a great variety of TLUs. The prevalence of ELF and multilingualism in these settings, along with our increasing knowledge that different academic disciplines tend to have their own language and literacy conventions and practices, are both major reasons why a universal standard language template is inadequate and misleading.

We have argued for several years (e.g. Jenkins & Leung 2014, Leung, Lewkowicz & Jenkins 2016, Jenkins & Leung 2017,) that local context should be paramount in test design. In respect of university English language entry testing, this means taking account of a range of considerations, most important of which are the candidates' first/other languages, the locality (country, region, institution, faculty, and discipline) in which they will be studying, and above all, the kind of communication in which they will be engaging which, in the case of international HE, is primarily ELF, or more precisely, 'English-within-multilingualism' (Jenkins 2018). These considerations lead us to the conclusion that we can't talk of the 'non-native speaker of English', but only of the 'local speaker', one who, to return to Mauranen's

similect theory, has a particular English similect deriving from L1 influence, and whose English is influenced by both the particular second order contact in which they engage and the language(s) of the specific local environments in which they use English. In this respect, see both Li 2017 pp.10-11 on the notion of the bilingual idiolect, and how it differs from conventionally-defined languages, and Mauranen 2018 p.113, who argues that "the multilingual speaker makes use of a whole, composite language resource" which is "a unique combination for every speaker", and thus that it could be argued "that the notion of one's 'own' language, common in folk linguistic beliefs and among professionals, is meaningful with regard to every speaker's idiolect"). The kinds of language use resulting from these various factors, let alone their complex interactions, can't be predefined; and if they can't be predefined, they can't be captured by conventional language rules and assessed in any monolithic standardized manner.

So what do we suggest to replace universalism in standardization? Given the strong grip of the prevailing paradigm on professional practices and the huge commercial interests involved, any change would likely be a complex and slow process, even if the development agenda had widespread consensus and support. But it is not beyond our pragmatic imagination that some nearer-term actions may be possible. For instance, large scale standardised testing could be augmented by local discipline-specific assessment tasks at or after admission, with the local stakeholders (teachers and students) being able to jointly decide how the assessment outcomes are used for formative and summative purposes. Another nearer-term possibility would be to re-design the current 'big' tests to create a space, in addition to the general language proficiency items, for discipline-specific local tasks. There are doubtless many other possibilities.

Looking to the longer term, our fundamental conceptual commitment to prioritizing the role of local context and our understanding of the relevance to contemporary international universities of accommodation and translanguaging skills, received an added stimulus recently from something completely outside language assessment, and even outside linguistics: the traffic experiments of the first decade of the 21st Century, initially in Holland, then in other parts of the EU including Denmark, Germany, and the UK. These have been described by one German commentator (Schulz 2006) as 'controlled chaos'. The basic idea, originally pioneered by Hans Monderman, was that traffic lights and other traffic signage should be removed, because 'the greater the number of prescriptions, the more people's sense of personal responsibility dwindles' and vice versa. The experiment has proved largely successful, with road users and pedestrians seeming to do better at self-regulation than was the case when they were heavily over-regulated with prohibitions, restrictions, warning signs and the like, many of which they simply ignored.

While we wouldn't want to stretch this analogy too far, as language and traffic don't have a great deal in common, these traffic experiments provided the immediate impetus for our longer-term alternative to standardized English language university entry testing: that is, LOCALLY-CONTEXTUALISED SELF-ASSESSMENT as the basis for international university English language entry decisions. What we have taken from the traffic experiments is the idea of SELF-REGULATION IN CONTEXT. In those areas where regulation of traffic has ceased, drivers, cyclists, and pedestrians have to pay close attention to local conditions and respond accordingly. Translating this into international HE, self-regulation would mean prospective students paying close attention to what is needed to operate in a specific local university context, and determining their own ability to do so in respect of self-assessment materials provided by that university department/programme. Meanwhile, the universities themselves (individual

departments/programmes) would be responsible for selecting the materials and activities that best represented those with/in which candidates would subsequently engage. The candidates would then have to decide whether or not they considered themselves ready, communication/ language-wise, to enter a particular programme in a particular discipline/department/university/country. What we are advocating, then, is taking university entry English language decisions away from the external TEST MAKERS and putting them in the hands of the TEST TAKERS as well as of those who will subsequently teach them. In effect we are suggesting, extending the democratic principle advocated by Shohamy (2001), that we should be putting the control, design and use of language assessment directly in the hands of the key stakeholders – students and teachers by means of giving teachers the responsibility for selecting the assessment materials, and students the responsibility for deciding whether they should 'pass' the assessment.

Such assessment would have two major advantages. Firstly, candidates would benefit from the process of the assessment itself, by being presented with the kinds of situation and materials they would subsequently meet/use in their studies, and learning more about their prospective field of study via the experience. They'd thus arrive on campus with the kinds of skills and knowledge, i.e. readiness, for the kinds of activities in which they'd engage in their studies. This contrasts dramatically with the current situation, where students are provided with test materials (produced by providers such as IELTS, TOEFL and Trinity) that bear little resemblance to their proposed subject of study, something about which students regularly complain. To give but one example, in a focus group study conducted by one of us (Maringe and Jenkins 2015), a Saudi-Arabian student who'd applied for a PhD in Education was scathing that she'd been given an IELTS reading comprehension test about cows. By contrast, self-assessment would not only mean a more relevant test and that learning about the prospective

field of study would be built into the test, but also that candidates would no longer be tested (and expected to prepare for testing) on knowledge and skills that they won't need.

Secondly, whereas existing standardized tests have often proved unsuccessful at predicting a candidate's readiness/suitability for university study (see our discussion in part 4 above; also Ducasse & Brown 2009 on IELTS, and Brooks & Swain 2014 on TOEFL, among others), if the entry decision was left to the candidates, aware of the high cost to themselves in terms of both time and money, they'd be less likely to award themselves a high score if they thought it was unwarranted and doubted their ability to manage on the programme in question. In addition, cheating would more likely seem pointless to them. Self-assessment would therefore lead also to more honest outcomes, including removing the risk of rejecting candidates who would have gone on to study successfully in their local context of choice – something that is a potentially grossly unfair outcome of current standardized tests. Likewise, there would be less likelihood of universities accepting unsuitable candidates simply for financial gain. There would also be a third advantage, one for the 'test makers' themselves: the opportunity to participate in a new wave of test materials design. Although ultimate decisions as to what to include in the self-assessment materials in any one place and time would have to remain with individual programme leaders and faculty, the testing experts could take on the role of consultants in the process, especially in its early stages, providing guidance and production support that could be adapted to suit each specific local situation (country/university/discipline/programme/student intake). Meanwhile, although the staff involved in individual courses/programmes may throw up their hands in horror at the idea of designing and preparing these self-assessment materials, once they'd done so the first time, subsequently they'd simply need to update their materials in line with future changes in their

programme and course design as well as factors such as their student demographic, their university's language policy and so on.

As to the nature of these self-assessment materials, although we're not materials designers ourselves, and would prefer to leave the key decisions to those who'll be teaching the respective courses in collaboration with test designers, we make the following tentative suggestions:

- Videos of typical seminars occurring early in the course so that candidates can check their understanding of what's being said. Alongside these, there could be tasks for candidates to enable candidates to contribute to the discussion at various points and then compare their contribution with what was actually said.

- Typical reading texts from an early stage in the course, perhaps with comprehension questions and answer keys for candidates to check their ability to understand the kinds of texts they will have to read.

- Typical assignment titles for the specific course, related to the reading texts provided, and sample good answers for candidates to compare with what they've produced, with annotations pointing out the merits.

- Sample student presentations (if relevant to the specific course), so that candidates can see what will be expected of them both in terms of content and language; and perhaps guidelines for candidates to prepare a presentation of their own, which they can compare with the sample provided.

- Tasks that invite candidates to consider and practise their accommodation and translanguaging skills; the concept of symbolic competence as expounded by Kramsch & Whiteside (2008) and Kramsch (2010) would be relevant for this development. If possible, these could also be built into the other materials.

In all these cases, it would be important not to give the impression that the self-assessment materials were replacing one kind of standardized test with another, albeit more local. The point would be that the materials represented TYPICAL (though by no means exhaustive) discipline-/content-related EXAMPLES for that particular context, but not some kind of language model or target. It would also need to be emphasized that apart from certain specific disciplinary language, the target was effective communication, both productively and receptively.

Harding & McNamara argue that '[t]he sociolinguistic reality of English as a lingua franca (ELF) communication represents one of the most significant challenges to language testing and assessment since the advent of the communicative revolution' (2018:570). The question, then, is this: do the English language assessment establishment and academic community have the willingness to take up the challenge of de-centred and dynamic ELF communication in disciplinary contexts, to 'unthink [their] classic distinctions and biases' (Blommaert 2010:1), replace them with the 21st Century ELF reality that surrounds them, and consider how this reality might be operationalised in entirely new kinds of tests? Or will the world continue to move in the direction of ever-increasing mobility and transcultural communication while the testers remain stuck in a 20th Century mono-groove according to which they still see the English language as the possession of a tiny minority, the NESs who, themselves, are often poor users of English in transcultural communication? As we reach the end of this paper, we quote from a participant in a Brazilian focus group study of Alessia Cogo and Sávio Siqueira (unpublished), which fits particularly well with our themes of local context, non-NES ownership of English, and linguistic fairness. The participant is explaining why s/he sees ELF as 'emancipation':

Emancipation … because if I take the language … I'm the speaker, that language

belongs to me. I have my own way of speaking that language in the sense of emancipating myself … it emancipates the students, and somehow the language empowers.

As this quotation demonstrates, authentic and agentive language use is at the heart of our communicative capacity to engage others and, at the same time, a platform for personal development (for a wider discussion see Leung & Scarino 2016).

We conclude by returning to our title and observing that the approach presented in our paper recognises the 'standard reality' of each individual local context and argues against imposing a 'mythical standard' for all. One thing is certain though: any adaptation and change will require the support of the professional expertise in assessment/test design, implementation and administration.

References

American Educational Research Association., American Psychological Association, National Council on Measurement in Education., & Joint Committee on Standards for Educational and Psychological Testing (U.S.). (2014). *Standards for educational and psychological testing*.

Bachman, L. & A. Palmer (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford: Oxford University Press.

Bachman, L. F., & A. Palmer (1996). *Language testing in practice*. Oxford: Oxford University Press.

Batziakas, V. (2016). *Investigating meaning-making in English as a Lingua Franca (ELF)* (Unpublished PhD thesis). King's College London, London.

Blommaert J. (2010). *The sociolinguistics of globalization*. Cambridge: Cambridge University Press.

Bourdieu P. & J.C. Passeron (1977). *Reproduction in education, society and culture*. London: Sage Publications.

Brooks L. & M. Swain (2014). Contextualizing performances: comparing performances during TOEFL iBT and real-life academic speaking activities. *Language Assessment Quarterly* 11.3, 353–373.

Canale, M. & M. Swain (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics* 1.1, 1–47.

Carlsen, C. (In press). The adequacy of the B2-level as university entrance requirement. *Language Assessment Quarterly*.

Cogo A. & M.Dewey (2012). *Analysing English as a Lingua Franca*. London: Continuum.

Cho, Y. & B. Bridgeman (2012). Relationship of TOEFL iBT® scores to academic performance: Some evidence from American universities. *Language Testing* 29.1, 421–442.

Cotton, F. & Conrow (1998). An investigation of the predictive validity of IELTS amongst a group of international students at the University of Tasmania. *IELTS Research Reports* Vol. 1, 72–115. Retrieved from http://www.ielts.org/researchers/research/volume_1.aspx

Cumming, A., R. Kantor, D. Powers, T. Santos & C. Taylor (2000). *TOEFL 2000 writing framework: A working paper*. Princeton, NJ: Educational Testing Service.

Ducasse, A.M. & A. Brown (2009). The role of intereactive communication in IELTS Speaking and its relationship to candidates' preparedness for study or training contexts. *IELTS Research Reports* Volume 12 www.ielts.org.

Edwards, J. (1994). *Multilingualism*. London: Routledge.

Ekiert, M. (2017). What Interlanguage Analysis reveals about L2 referent tracking. Talk given in UCL IoE Centre for Applied Linguistics Research Seminar Series, 20 November 2017.

García O. & W. Li (2014). *Translanguaging: Language, bilingualism and education*. Houndmills, Basingstoke: Palgrave Macmillan.

Gee, J. P. (2004). *Situated language and learning: A critique of traditional schooling*. London: Routledge.

Gray, J. (2010). The branding of English and the culture of the new capitalism: Representations of the world of work in English Language textbooks. *Applied Linguistics* 31.5, 714 – 733.

Green, A. (2014). *Exploring language testing and assessment*. London: Routledge.

Harding L. & T. McNamara (2018). Language assessment. The challenge of ELF. In J. Jenkins, W. Baker & M. Dewey (eds.), *The Routledge handbook of English as a Lingua Franca*. Abdingdon, UK: Routledge, 570–582.

Harrington, M., & T. Roche (2014). Identifying academically at-risk students in an English-as-a-Lingua-Franca university setting. *Journal of English for Academic Purposes* 15, 37–47.

Hyland, K. (2004). *Disciplinary discourses: Social interactions in academic writing*. Michigan: Michigan University Press.

Hyland, K. (2016). Academic publishing and the myth of linguistic injustice. *Journal of Second Language Writing* 31, 58–69.

Hyland, K., & L. Hamp-Lyons (2002). EAP: issues and directions. *Journal of English for Academic Purposes* 1.1, 1–12.

Ingram, D., & A. Bayliss (2007). *IELTS as a predictor of academic language performance Part 1*. Retrieved from London:

James, M. (2006). Assessment, teaching and theories of learning. In J. Gardner (ed.), *Assessment and learning* . London: Sage, 47–60.

Jenkins, J. (2000). *The phonology of English as an International Language*. Oxford: Oxford University Press.

Jenkins, J. (2015). Repositioning English and multilingualism in English as a lingua franca. *Englishes in Practice* 2/3: 49–85. De Gruyter Open. Available for free download at https://www.degruyter.com/view/j/eip.

Jenkins, J. 2018. Not English but English-within-multilingualism'. In S. Coffey & U. Wingate (eds.), *New directions in foreign language education*. Abingdon, UK: Routledge, 65–78.

Jenkins, J. & C. Leung (2014). English as a Lingua Franca. In A. Kunnan ed. 2014. *The companion to language assessment*. Chichester: Wiley Blackwell, 1607–1616.

Jenkins, J. & C. Leung (2017). Assessing English as a Lingua Franca. In E. Shohamy, I. Or & S. May 2017, *Language testing and assessment*, 3rd ed. Vol. 7 of S. May ed. *Encyclopedia of language and education.* Heidelberg: Springer, 103–117.

Kachru, B.B. (1985). Standards, codification and sociolinguistic realism: the English language in the outer circle. In R. Quirk & H.G. Widdowson (eds.): *English in the world: Teaching and learning the language and literatures.* Cambridge: Cambridge University Press, 11–30.

Kane, M. (2006). Validation. In R. Brenman (ed.), *Educational measurement,* 4th ed. Westport, CT: American Council on Education and Praeger, 17–64.

Kane, M. (2010). Validity and fairness. *Language Testing* 17.2, 177–182.

Kramsch, C. (2010). The symbolic dimensions of the intercultural (Plenary speech). *Language Teaching.* Retrieved from

http://journals.cambridge.org/action/displayFulltext?type=1&pdftype=1&fid=7931790&jid=LTA&volumeId=-1&issueId=&aid=7931788

Kramsch, C. & A. Whiteside (2008). Language ecology in multilingual settings: Towards a theory of symbolic competence. *Applied Linguistics* 29.4, 645–671.

Kunnan A.J. (2018). Evaluating language assessments. New York & London: Routledge.

Lea, M. R., & B. Street (2006). The 'Academic Literacies' model: theory and applications. *Theory into Practice* 45.4, 368–377.

Lee, Y.-J., & J. Greene (2007). The predictive validity of an ESL placement test: A mixed methods approach. *Journal of Mixed Research Methods* 1.4, 366–389.

Leung, C. (2011). Language teaching and language assessment. In R. Wodak, B. Johnstone, & P. Kerswill (eds.), *The Sage handbook of sociolinguistics*. London: Sage, 545–564.

Leung, C. (2013). The 'Social' in English language teaching: Abstracted norms versus situated enactments. *Journal of English as a Lingua Franca* 2.2, 283–313.

Leung C. & J. Jenkins (2018). Farewell to the phantom of standardization. Paper given in the colloquium *Transition, mobility, validity: English as a (multi)lingua franca perspectives on language assessment*. Language Testing Research Colloquium (LTRC), Auckland, July 4-6, 2018.

Leung C., J. Lewkowicz & J. Jenkins (2016). English for Academic Purposes: a need for remodelling. *Englishes in Practice* 3.3, 55–73. De Gruyter Open. Available for free download at: https://www.degruyter.com/view/j/eip

Leung, C., & A. Scarino (2016). Reconceptualizing the nature of goals and outcomes in language/s education. *Modern Language Journal* 1000(S1), 81–95.

Li, W. (2017). Translanguaging as a practical theory of language. *Applied Linguistics* open access 00/0:1-23, doi:10.1093/applin/amx039.

Lillis, T. & M. Scott (2007). Defining academic literacies research: Issues of epistemology, ideology and strategy. *Journal of Applied Linguistics* 4.1, 5–32.

Lippi-Green, R. (1997). *English with an accent*, 1st ed. London: Routledge.

Maringe F. & J. Jenkins (2015). Stigma, tensions and apprehension. The academic writing Experience of international students. *International Journal of Educational Management* 29.5, 609–26.

Mauranen, A. (2003). The corpus of English as a Lingua Franca in academic settings. *TESOL Quarterly* 37.3, 513–27.

Mauranen, A. (2012). *Exploring ELF. Academic English shaped by non-native speakers*. Cambridge: Cambridge University Press.

McNamara, T. (2001). Language assessment as social practice: Challenges for research. *Language Testing* 18.4, 333–349.

McNamara, T. (2005). 21st century shibboleth: Language tests, identity and intergroup conflict. *Language Policy* 4.4, 351–370.

McNamara, T. (2014). 30 years on – evolution or revolution? *Language Assessment Quarterly* 11, 226–232.

McNamara T. 2017. A challenge for language testing: The assessment of English as a Lingua Franca. Department of Education, University of Oxford, 11 December 2017.

McNamara, T., J. Morton, N. Storch & C. Thompson (In press). Students' accounts of their first year undergraduate academic writing experience: Implications for the use of the CEFR. *Language Assessment Quarterly.*

Messick, S. (1989). Validity. In R. L. Linn (ed.), *Educational measurement* (3rd ed.). New York: ACE-NCME, MacMillan, 13–103.

Murray, N. (2018). Language education and dynamic ecologies in world Englishes. In E.L. Low & A. Pakir (eds.) *World Englishes. Rethinking paradigms*. Abingdon, Oxon: Routledge, 47–63.

Newton, P. & S. Shaw (2014). *Validity in educational and psychological assessment*. London: Sage.

Piller, I. (2016). *Linguistic diversity and social justice*. Oxford: Oxford University Press.

Pitzl, M-L. (2018, in press). *Creativity in English as a Lingua Franca: Idiom and metaphor*. Berlin: De Gruyter Mouton.

Rawls, J. (2001). *Justice as fairness: A restatement (edited by Erin Kelly)*. Cambridge, Mass.: Harvard University Press.

Schulz M. (2006). Controlled chaos: European cities do away with traffic signs. *Spiegel Online International*, 16 November 2006.

Seidlhofer, B. (2001). Closing a conceptual gap: the case for a description of English as a lingua franca. *International Journal of Applied Linguistics* 11.2, 133–58.

Seidlhofer, B. (2011). *Understanding English as a Lingua Franca*. Oxford: Oxford University Press.

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics* 10, 209–31.

Shohamy, E. (2001). Democratic assessment as an alternative. *Language Testing* 18.4, 373–391.

Shohamy, E. (2006). *Language policy: Hidden agendas and new approaches*. London: Routledge.

Shohamy, E. (2011). Assessing multilingual competencies: Adopting construct valid assessment policies. *Modern Language Journal* 95.3, 418–429.

Shohamy, E. (2017). Critical language testing. In Shohamy, Or & May (eds.) 2017, 441-454.

Shohamy E., I. Or & S. May (eds.) 2017. *Language testing and assessment* (3rd ed). In S. May (ed.) *Encyclopedia of language and education*. Heidelberg: Springer.

Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.

Van Parijs, P. (2011). *Linguistic justice for Europe & for the World*. Oxford: Oxford University Press.

Vertovec, S. (2007). Super-diversity and its implications. *Ethnic and Racial Studies* 30.6, 1024–1054

Weir, C. J., & B. O'Sullivan (2017). *Assessing English on the global stage: The British Council and English language testing, 1941-2016*. London: Equinox.

Wilson, M. (2005). *Constructing measures: An item response modelling approach*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Wingate, U. (2015). *Academic literacy and student diversity: The case for inclusive practice*. Bristol: Multilingual Matters.

Wingate, U. (2016). Academic literacy across the curriculum: Towards a collaborative instrumental approach. *Language Teaching, First View* https://doi.org/10.1017/S0261444816000264.

Yen, D. & J. Kuzma (2009). Higher IELTS score, higher academic performance? The Validity of IELTS in predicting the academic performance of Chinese students. *Worcester Journal of Learning and Teaching* 3, 1–7.