Constant Leung, Jo Lewkowicz and Jennifer Jenkins

English for Academic Purposes: A need for remodelling[1]

Abstract

English for Academic Purposes (EAP) is an established domain of research, teaching, and assessment within additional/second language education. In this article we examine the conceptualisation of English that underpins much of its current thinking and pedagogic practice, and raise questions of validity and claims of 'fit-for-purpose'. In particular we explore issues underpinning EAP assessment and argue that there is a need to reconceptualise the basis of the language model. We propose that given the complex and changing practices in academic communication, there is a good case for broadening the established understanding of Academic English to better reflect target language use. The principles and arguments underlying this discussion are relevant to assessment as well as to EAP more broadly.

Constant Leung, King's College London. constant.leung@kcl.ac.uk
Jo Lewkowicz, King's College London, jo.lewcowicz@kcl.ac.uk
Jennifer Jenkins, University of Southampton. j.jenkins@soton.ac.uk

1. Introduction

It is generally recognised that academic English is a challenge to most students. This recognition is one of the reasons that that the higher education sector in Australia and New Zealand has now instituted a policy of post-enrolment assessment of all first-year students, home and international students alike. (see, for example, Dunworth, 2009, 2010, 2013; Dunworth, Drury, Kralik & Moore., 2013). However,

---

[1] We would like to thank the two anonymous reviewers for their helpful suggestions. Any remaining infelicities are of course our own.

many universities still rely on pre-enrolment large-scale tests offered by internationally recognised testing bodies such as Educational Testing Service, Cambridge English Language Assessment and Pearson's for determining whether incoming students have the necessary language skills to undertake their studies through the medium of English. And, despite the fact that tests such as IELTS and TOEFL have undergone substantial revisions in the last 40+ years, given our current understanding of language use in academic contexts, the question remains: 'Are the readily available and widely used tests of EAP "fit for purpose"?'

This paper explores the above question from two perspectives. It looks at test utility and validity from a theoretical perspective and challenges the way these are currently conceived. It then highlights the lack of fit between the language tested and the way language is actually used in academic contexts. The discussion starts, however, by providing some background to current standardized language test use and practices.

2. Standardized test use and practices

Educational mobility is not new. Davies (2008) points out that as early as the 1950s the presence of students from non-English-speaking countries to the UK was noted and that between the 1950s and 1960s there was a five-fold rise in the number of international students in higher education in the UK. What is new, however, is the numbers of students undertaking at least part of their studies abroad, and the range of places from which they come and to which they travel for education. This extensive mobility has been largely facilitated by the widespread use of English in higher education institutions, well beyond the borders of Anglophone countries. The rise of English medium higher education has brought with it the need for assessing the language proficiency of applicants and has helped the language testing industry to flourish. Tests of academic English abound with perhaps the most widely accepted on the international scene being TOEFL developed by Educational Testing Service (ETS) in the US and IELTS developed by Cambridge English Language Assessment together with the British Council and International Development Program (IDP), Australia. There is an implicit acceptance among test-takers and other test users that these large scale tests are well researched and provide the necessary and relevant information about university applicants for correct decisions to be taken.

Thus, if a candidate attains the minimum score requirement on a standardized test, then both the applicant and the university authorities are given to believe that the applicant has the language skills to function in the university setting. 'English

proficiency scores … are assumed to be definitive sources of evidence about a student's language ability and readiness to study in an English-medium institution' (O'Loughlin, 2008, p.81), or least that is the kind of promotional rhetoric often found in their print and web based publicity material presented by the international test organisations

Yet, it is widely acknowledged that the scores of large-scale standardized language tests such as IELTS and TOEFL are not necessarily reliable and sensitive predictors of future academic performance. The relevant research investigating the relationship between test scores and test-takers' subsequent ability to perform academic tasks at university has shown somewhat contradictory findings. Some studies found little or no statistically significant relationship between test scores and academic performance (e.g. Cotton & Conrow, 1998; Lee & Greene, 2007), while others found a positive but generally weak correlation. Elder, Bright and Bennett (2007) investigated the relationship between academic language proficiency and subsequent academic performance of groups of home and international first-year students across a range of disciplines in a New Zealand university. They reported a correlation of 0.32 between their Grade Point Average (GPA) and test scores in Listening, Reading and Writing (combined), which accounted for 10% (approx.) of the variance of GPA. (Also see Cho & Bridgeman, 2012; Feast, 2002; Woodrow, 2006). Perhaps it is not surprising that language test scores are not very useful predictors of students' future academic performance. After all, it is widely acknowledged that language competence is but one of the many intertwining components of academic participation. Individual investment and interest in academic study, the learning environment, and the capacity to deal with disciplinary content are among the other important factors involved. If language competence were the sole or main contributing factor to academic performance, then there should be little variation among native speakers of English in English-medium schools and universities. We know that this is not the case. For these reasons many university admissions tutors do not restrict themselves to test scores as the sole indicator of English language proficiency; they also take account of other sources of information regarding applicants' English language and academic achievements (cf. Dickins et al, 2011).

It is perhaps pertinent to ask how far the internationally marketed standardized English language tests are tapping into the kinds of language use that correspond to the actual communicative activities in academic target language use (TLU) contexts. In a detailed analysis of the language use of four international students, Paul (2007, p.30) reports that 'entry-level IELTS ratings do broadly predict students' capacity for language production in academic settings. However, when examining language behaviour by these four individuals in different task types across spoken and written genres, three of these students experienced difficulty in language and content with

increased complexity of academic demands'. She also observes that the language features associated with the global band descriptors (and the associated sub-skills) do not appear to reflect the task-specific requirements in different disciplines. For instance, '[i]n many written tasks across disciplines, rhetorical organisation, acknowledgement of sources from the field, positioning and use of field specific vocabulary play a key role in successful completion' (op. cit., p.26). The diverse expectations and practices in academic discourse in different disciplinary areas have, of course, been well documented by research in the field of academic literacies and academic writing (e.g. Lea & Street, 1998, 2006; Lillis & Scott, 2007; Tribble & Wingate, 2013, Wingate 2015).

Another salient question is how far English language tests that are primarily concerned with assessing individual test-takers' knowledge and skills can take account of the complexities engendered by the multi-party nature of academic communication, particularly in relation to the spoken language. As part of their study of IELTS as a predictor of academic language performance, Ingram & Bayliss (2007, p.43) interviewed a group of international students most of whom reported that they understood very little in the early weeks of their first semester: 'All students said that the speed of native-speaker speech in the lectures was problematic'. (These students estimated that their comprehension grew to between 60% and 80% by the end of the semester.) The researchers observe that some of the teaching staff 'made little effort to modify their language despite the fact that at least 25% of the students in their classes were from non-English-speaking backgrounds' and they 'spoke extremely rapidly and failed to repeat key information' (loc. cit.). Another aspect of the complex nature of academic communication is related to classroom participation structures. Participatory talk and public display of knowledge and opinion are valued in most, if not all, English-speaking classrooms. Ingram & Bayliss report that their participant students were reluctant to offer opinions in class for a variety of reasons. The students suggested that the lack of the necessary language repertoire, the reluctance to show oneself up in front of peers and the lack of time for working up a response all contributed to their conduct in class. These are clearly issues directly related to language use in academic communication, but they cannot be easily accommodated within an approach that regards language proficiency as something that resides largely within the individual.

Given the limitations of relevant information that can be gleaned from proficiency tests and the continuing difficulties that incoming non-native English speaker (henceforth NNES) students experience when entering university, it is not surprising that academic institutions are beginning to question the reliance on standardized test scores and to look for alternative provisions to aid students transitioning to university. A cursory glance at the webpages of the English Language centres in UK

universities would bear witness to this. We now turn to consider the possible reasons for this situation from a theoretical perspective.

3. Large-scale standardized assessment frameworks: validity and utility   claims

The educational value and public credibility of any language assessment instrument are closely associated with its claims to validity and demonstrated practical utility. For our present purposes, validity issues are critically important; in real life, validity and utility are related. Newton & Shaw (2014, p.1) open their extensive account of the development of the concept of validity with this statement: 'Validity is the hallmark of quality as far as educational and psychological measurement is concerned, the "single most important criterion" for evaluating a test …'.   Their work also shows that validity continues to be a contested issue in educational assessment generally. In this section we will address some aspects of validity as they relate to large-scale English Language assessment at this time, in respect firstly, of the conceptual/theoretical claims that have been made, and secondly, of real-life utility claims.

Conceptual/theoretical claims

Messick's (1989) unified theory of validity, largely located within a psychometric tradition, has been very influential in language testing in the past 25 years, particularly in relation to English as a foreign/second language. Two facets of his ideas on validity are relevant to this discussion: construct validity and consequential validity, particularly in terms of how they have been taken up and operationalized. 'Construct' is typically glossed in this way: '… we can consider a construct to be the specific definition of an ability that provides the basis for 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 the assessment performance' (Bachman & Palmer, 2010, p.43). Seen in this light, 'construct' is defined prior to test development; it specifies the knowledge/skill/ability to be assessed. It is the basis of test item development. The underlying reasoning is as follows: If we can define the construct for, say, English for Physical Education in English-medium secondary education in an international school setting, then we can develop appropriate assessment tasks and activities to tap into the knowledge and skills involved. Furthermore, the test/assessment scores (indexing performance) yielded can be used to make

trustworthy inference on the test-taker's proficiency and likely future performance in that domain[2].

At the present moment, it is generally accepted that construct validity is established by demonstrating that test scores have been interpreted appropriately (with reference to construct definition) (see, for example, Green 2014). For instance, if spoken academic language proficiency is defined in terms of the ability to understand lectures and seminars, and assuming that the test items tap into these activities, then a resultant test score should be interpreted as an indication of the test-taker's ability in that domain of language use.

This view gives rise to two epistemological problems. Firstly, the defining of construct itself depends on what counts as relevant attributes to be taken into account in both theoretical and empirical terms. Secondly, validity does not refer to the properties of a test, but to the interpretation of test score. On this view, validity (or claims of validity) is conceptually located at some distance from any actual assessment instrument or task. These issues resonate with some of the contentions in the continuing debate on validity in the wider educational assessment field (see Borsboom, Cramer, Kievit, Scholten and Fransić (2009); Newton and Shaw (2014); also the collection in *Educational Researcher*, 36/8 for a representation of this debate). One point of contention stems from the view that the validity of a test should be immediately linked to its properties, i.e. the epistemological and ontological nature of the attributes of a test. Another area of contention is that construct should be established on the basis of the cognitive composition of test items. In other words, construct can be described, if not defined, by analysing and decomposing focal tasks into cognitive processes and resources, e.g. knowledge structures and performance strategies.

These issues are clearly related to the question of what counts as relevant attributes. They also signal that construct is theory-driven (either explicitly or implicitly) and it can be construed in a variety of different ways. Indeed, if we look at tests in a subject such as History it is possible to see, in quite sharp relief, a range of educational concerns regarding subject content (e.g. which period/s of history?) and nature of knowledge (e.g. what kind/s of reasoning?) embedded in the construct. Seen in this light, there can be different ways of construing the notion of construct in English Language assessment.

---

[2] Test performance can of course be influenced by extra-construct factors such as a test-taker's motivation. So the score attributed to a test-taker may include some non-construct-related matter but for reasons of focus we are restricting our discussion to issues related to construct.

In relation to English Language assessment in the past fifty years or so, the dominant psychometric paradigm has informed the way construct is understood: a 'specific definition of an ability that provides the basis for a given assessment or assessment task and for interpreting scores derived from this task' (Bachman & Palmer, 2010, p.43). Procedurally this means that we need to start by identifying a domain of communicative activity (e.g. writing an essay at university). Bachman & Palmer (1996:89) suggest that we should state '... the precise nature of the ability we want to measure, by defining it abstractly'. Once we have this abstractly defined construct we can proceed to seek empirical accounts of how this activity is performed and what abilities are involved. This procedural sequence reflects a psychometric assumption that the defining of a construct has to precede assessment (e.g. Kane, 2004).

There is a good deal of careful discussion on how to go about establishing the abilities involved in a focal activity. Wilson (2005), for instance, proposes using a construct map as a way of representing and specifying the different facets or dimensions within any abstractly defined 'global' construct (e.g. writing an essay may involve the ability to use lexical and grammatical knowledge and to select subject content appropriately). Once these dimensions have been 'mapped', it would be possible to develop test items to tap into the required abilities. The relationship between a test construct and test items should be as close as possible so that the inference drawn between performance (triggered by test item) *and* test construct is defensible. In fact, Wilson (op. cit.) suggests that one may conceptualize a kind of causality going from construct to test items, i.e. the test designer assumes that the test-takers have (some amount of) the construct (understood in terms of ability) in question, and their test performance (triggered by test items) is caused by this putative construct. The measurement (score) reflects the amount of the construct residing in the test-taker.

However, research findings from different branches of language studies have demonstrated amply over recent decades that real life English communication does not comply neatly with such simple, uniform constructs, and, moreover, that this is true of native as well as non-native English users. In terms of spoken English, Crystal & Davy (1975) were the first to demonstrate empirically that native English speakers often speak 'ungrammatically' in respect of 'standard' English and employ extensive stylistic variation, while other scholars such as Carter & McCarthy (1995) and Cheshire (1999) have shown that the grammar of native English speech differs in numerous ways from that of the written channel.

More recently, the use of English in writing as well as speech has been further complexified by the massive spread of electronic communication and the range of grammars and styles employed in email, texting, twitter and the like (see e.g. Baron

2010). Meanwhile, sociolinguists have long been arguing that the notion of 'standard' English is itself an idealised abstraction far removed from the realities of "the diversified and variable data of everyday interaction" (Milroy 1994, p.156), and that rather than being set in stone, the English language, both spoken and (albeit more slowly) written, is subject to constant change. In other words, even in relation to native English users, research has increasingly found that people do not use language in predictable ways, that they do not necessarily follow conventional rules of the kind described by language assessment professionals, and that the English language is, itself, in a state of permanent flux (note, for example, how the Oxford English Dictionary adds and removes words on an annual basis).

Such findings already had the potential to destabilise the English language assessment status quo even before more recent research developments relating to non-native English users entered the frame. In the past twenty years, however, findings from research into both English as a Lingua Franca (henceforth ELF) and multilingualism have begun to call into question more seriously the basis on which the English of non-native users is assessed in domains such as English medium higher education. We will argue on the basis of these two bodies of research, which add substantially to the problems discussed earlier concerning the ambiguity and contradictory findings of studies of standardized tests, that the tests in their present form are not fit for purpose.

Real-life utility claims: The challenge of the changing linguistic landscape and implications for language modelling and norming

In the past fifty years or so, English Language Teaching (henceforth ELT), as a worldwide profession, has tended to assume that a 'native speaker' variety of Standard English (generally either British or American) provides the only appropriate model for pedagogic and curriculum purposes. While the notion of the 'native speaker' has been relatively user-friendly for modelling norms of English, recent developments in the ways in which English is used in many parts of the world have cast doubts on its value for teaching and assessment. Meanwhile, recent developments in research into multilingualism is demonstrating problems at a deeper level by casting serious doubt on the possibility of separating the languages within an individual's linguistic repertoire, and even of the feasibility of identifying boundaries between languages. We turn now to consider these two bodies of research, first English as a Lingua Franca, then critical multilingualism and their relevance to informing present-day assessment practices in relation to academic language.

English as a Lingua Franca

Since the first days of British colonisation in the early seventeenth century, English has been learnt and used by those for whom it is not their native language. And with its more recent spread across the globe into many regions that were never colonised by the British (Kachru's 'expanding circle'; see Kachru 1992), English is regularly the primary shared language of communication, or lingua franca, among speakers from diverse first language backgrounds who may, but more often do not, include native English speakers among their number. Non-native English speaking users of English as a lingua franca, or ELF, nowadays constitute numerically the largest group of English users in the world (see e.g. Crystal 2012). Given the continuing internationalisation of education, business, industry and research, ELF is likely to remain a significant and growing phenomenon involving an ever-larger number of speakers well into the foreseeable future.

This is not the place for an extended discussion of ELF (see e.g. Jenkins, Cogo & Dewey 2011, Mauranen 2012, Seidlhofer 2011 for fuller details). What follows is an outline of the key findings from ELF research that have implications for English language assessment particularly in terms of assessing whether non-native speakers' English is appropriate for entry into programmes of study in English dominant or English medium universities (i.e. universities in, respectively, Anglophone and non-Anglophone settings).

The first point to make is that ELF speakers typically have what Mauranen (2012, pp.28-29) calls 'similects'. As she observes, speakers of English from countries where English is neither the mother tongue nor an official postcolonial language generally use their first language rather than English to communicate with each other, and reserve English to communicate with people from other first languages than their own. On the one hand, this means that speakers from a specific first language such as Japanese tend to share certain characteristics, or at least a certain 'flavour', in their English as a result of the influence of the same first language on their English. But, on the other hand, it also means that there is no community of, for example, Japanese English speakers. Rather, their English develops in parallel with that of each other through their various diverse contacts with speakers of other similects such as Spanish, Polish, Korean. This accounts for both shared features among ELF users from a specific first language and for shared features among ELF users from different first languages that have been found in empirical ELF research.

Turning to these ELF research findings themselves, the challenge they provide to English language assessment is three-fold. Firstly, as described above, ELF users' similects include to a greater or lesser degree certain influences from their L1s that from an ELF perspective are considered no less acceptable than regional features of

native English varieties. Secondly, the findings demonstrate that a number of other features of native English are often replaced in similar ways regardless of speakers' first languages. Examples of such features include: countable use of nouns that are uncountable in standard native English (e.g. 'feedbacks', 'informations', 'softwares'); zero marking of the third person singular in the present simple tense (e.g. 'she think'); and replacement of the voiceless and voiced dental fricatives 'th' (typically with 's' and 'z' or 't' and 'd'). In the first case, ELF research demonstrates that similect features rarely impede successful communication, while in the second case it shows that these usages across L1s are usually highly communicatively effective. Thirdly, ELF research findings highlight the extent to which the use of ELF is emergent (see Baird, Baker, & Kitazawa 2014), and that contingent variability is therefore a defining characteristic of ELF communication. It could thus be described as 'variably variable', with users constantly adjusting the way they speak in order to accommodate to their interlocutors and work with them in the co-construction of meaning as they discover 'online' the extent of their shared linguistic repertoires (see Jenkins 2015). Notwithstanding what we said earlier about the impossibility of pinning down even native speaker real-life language use and imposing a predetermined proficiency template on it, when ELF communication enters the equation, this impossibility is multiplied many times over. And yet this is precisely the communication on which much English language assessment attempts to impose such a template. For the international examination boards responsible for standardized testing have not, to date, incorporated any of these three aspects of ELF communication into their test designs, despite the fact that in many cases, such as university English language entry testing, candidates (if successful) will go on to participate primarily in higher education communities that use ELF.

Multilingualism

We turn now to the body of research that takes a critical approach to multilingualism. The conventional approach to multilingualism is one in which the two or more languages of a multilingual speaker are seen as separate entities that should not be 'tainted' by each other within the speaker's multilingual repertoire. Heller refers to this as "parallel monolingualism" in which "each variety must conform to certain prescriptive norms" (1999, p.271), while Cummins describes it as the "two solitudes assumption" (2005, p.588). More recently, there has been a 'multilingual turn' within applied linguistics that involves a move away from notions such as 'native speaker competence', 'L1 interference', and monolingual approaches to languages more generally, to a focus on language as social practice and multilingualism as a valuable resource rather than a problem.

The newer critical orientation to multilingualism involves a number of heavily overlapping research approaches. These are 'translanguaging' (e.g. García 2009,

García & Li Wei 2014), 'translingual practices', 'code-meshing' (e.g. Canagarajah 2011, 2013), 'polylanguaging'/'poly-lingual languaging' (e.g. Jørgensen 2008, Jørgensen, Karrebæk, Madsen & Møller 2011, Jørgensen & Møller 2014), 'super-diversity' (e.g. Vertovec 2006, 2007), 'metrolingualism' (e.g. Otsuji & Pennycook 2010, 2011), and 'mobile resources' (e.g. Blommaert 2010). For present purposes, we need not be concerned with the conceptual differences across these various approaches. Suffice it to say that the current research in this field as a whole has focused on how speakers make use of all the linguistic resources at their disposal to engage in communication in particular language environments. However, we will look more closely at the phenomenon/approach known as 'translanguaging' as it has particular resonance with the position we are developing here vis-à-vis language modelling and norming, as well as enabling us to take a more nuanced approach to the multilingualism of most ELF users, including all speakers of English as an additional language in higher education settings.

It has long been recognised that in linguistically diverse communities where different languages or varieties of language/s are co-present, speakers often combine elements of all the available languages flexibly in their communication practices. Up until a few years ago such language practices were commonly referred to as 'code-switching'. More recently, in line with a more complexified understanding of the phenomenon and as part of the wider multilingual turn, it has become known by a number of (critical) multilingualism scholars as 'translanguaging'.

The practice of translanguaging, which has much in common with the approaches listed above, includes but goes far beyond, code-switching. As García & Li Wei (2014, p.22) observe:

> Translanguaging differs from the notion of code-switching in that it refers not simply to a shift or a shuttle between two languages, but to the speakers' construction and use of original and complex interrelated discursive practices that cannot be easily assigned to one or another traditional definition of a language, but that make up the speakers' complete language repertoire.

They go on to describe translanguaging as the "bilingual norm" (p. 23), one that creates what Li Wei has called a "translanguaging space" in which "the artificial dichotomies between the macro and the micro, the societal and the individual, and the social and the psycho in studies of bilingualism and multilingualism" is broken down (2011, p.1234). This involves not only a hybrid mix of L1s and L2s but also new creative possibilities. Translanguaging research has a strong educational focus. For instance, Creese & Blackledge (2010), in their work on flexible bilingual pedagogy, look at how elements of Gujarati are used in combination with English in a complementary school in England. A good deal of this body of work is concerned

with such kinds of language practices among young people, particularly in social and school settings. This, in turn, makes translanguaging research especially relevant for the way language use is approached in higher education, including the kinds of language use that are accepted in language entry testing as appropriate for higher education study purposes.

4. The need to reconceptualize principles of testing academic English

We have already alluded to the lack of progress in current, large-scale assessments of EAP to provide the necessary and required information for stakeholders of higher education. In this section we draw our arguments together to show more precisely why we believe there is a need to reconceptualize the language modelling of EAP assessment. Before doing so, however, we consider the way in which such EAP tests are currently conceptualized.

The testing of EAP as distinct from general language proficiency dates back to the 1980s. Since then, although the major tests such as TOEFL and IELTS have undergone substantial change, the basic premise on which they are operationalized is that of communicative competence as modelled on the educated native speaker, with the authority of the native speaker being 'enshrined in the test construct' (McNamara, 2011, p.517). Thus the NNES is judged against an idealised, monolingual native English speaker (henceforth NES) performance in terms of the distance between them, which by implication privileges the NES and views the NNES as to some extent deficient.

What is said to characterise tests of specific purposes, whether they are for occupational purposes such as for airline pilots, or for academic purposes, such as IELTS, is their authenticity in terms of input content and task performance (Douglas, 2000). In other words, in a test of EAP the underlying assumption is that the tasks candidates are to perform share features of communicative language use at university. Candidates are to use their knowledge of subject content to complete the test. But in terms of EAP testing this assumption is fraught with difficulties. For a start, applicants to university will not have the content knowledge prior to their studies. Thus, it would be naïve to expect them to use such knowledge even if it were possible to construct tests that are specific enough to be applicable to all groups of applicants. The argument that the field of EAP seems too broad to realistically test on the basis of specific content knowledge has been well rehearsed (see, for example, Davies, 2008). Early versions of IELTS which had specific-purpose modules were rejected in subsequent revisions because of inherent

problems with comparability across versions and because they were not able to adequately address the specific needs of all university applicants. The need to satisfy everyone has resulted in the use of input texts, for reading as well as listening, of general interest that all candidates are supposed to be able to deal with, thus failing to reflect the specific features of written or spoken language that students of, for example, physics, geography or law are likely to encounter during their studies.

A further but related concern is one of task authenticity. Given the broad scope of EAP, it is impossible to select texts that could be considered authentic for all test candidates. This leads one to conclude that what distinguishes an EAP test from a more general tests of language proficiency is the test tasks and expected test outcomes. But here too there are problems. Much of what goes on in an academic setting is not what is assessed and valued in language assessment. For example, Leung & Lewkowicz (2013) have shown how linguistic accuracy may be less important than getting the message across in real-life classrooms, yet in an language assessment setting it is the accuracy (as articulated by language experts), breadth of vocabulary, fluency and organisation of expressed thought that are rewarded in terms of higher test scores. The value of one's contribution to an ongoing debate or discussion, to be able to co-construct spoken discourse or to negotiate meaning which have been shown to be central to successful participation in academic classrooms (Leung & Lewkowicz, op. cit.) seem to have no place in the test situation. Indeed one of the hallmarks of authentic spoken language is its 'unpredictability' in interactional contexts (Harding, 2014), and yet unpredictability causes turbulence in the mapping of pre-specified construct (see earlier discussion). This is readily apparent in the TOEFL speaking test which is 'tape-mediated' thus giving candidates no opportunity to engage the interlocutor in real-life interaction: essentially each response is a presentation for which the candidate has had minimal time to prepare. And although the IELTS speaking test is 'live', the examiner is required to follow an interlocutor frame (O'Sullivan & Liu, 2006) which again fails to allow genuine real-life interaction of the sort that candidates would participate in during their studies. Norton (2013), in a study of the co-constructed nature of talk in spoken language tests, has shown how examiners may deviate from the script to engage the candidate in a contingent way, e.g. embellishing or paraphrasing information, and how such 'deviations' can impact on the candidates' performance in an unpredictable way.

Performing under test conditions in an EAP test of writing also raises questions of authenticity. Most academic writing requires prior researching of a topic and being able to show an understanding of different points of view as well as expressing one's own stance. When contrasted with language assessment writing, academic writing is different in many ways. But perhaps the most significant difference is the way

learners are taught to develop their argument in answer to an exam-type essay and the way they are expected to do so in answer to an academic paper, which has also been found to vary according to subject area (Wingate, 2015). Thus, students may need to 'unlearn' what they have been taught on exam preparation courses. For example, in a study investigating students' early perceptions of their studies (Jenkins 2014), one participant described how she had been taught to structure paragraphs on her preparation course. But when she structured them in this way in her first piece of writing for her PhD supervisor, the supervisor exclaimed "no, no, no", and told her it was far too mechanical. This shows that faculty expectations of what entails successful writing may be very different from what those preparing to take an EAP test are taught in order to succeed in the exam.

Thus, there appears to be considerable discord between, on the one hand, current understanding of what distinguishes EAP tests from more general language proficiency tests and, on the other hand, the theory underpinning such tests and the way they are operationalized in practice. The EAP test appears to be a misnomer in that it is based on a premise that does not appear to exist. Yet, such tests continue to be constructed on the basis of an a priori notion of an EAP construct that fails to account for real life (authentic) use of language in the world of academia, drawing inferences on potential future success of candidates from what may be less than appropriate performances. There still appears to be a tacit acceptance that given the alternatives, the current large-scale EAP tests provide the most convenient means of gatekeeping – a point we take issue with.

If assessments of incoming students are to be valid then the emergent linguistic forms and pragmatic uses in ELF and translanguaging communication need to be taken into account when developing English Language curriculum and assessment frameworks, as they impact on the ways in which English is used. Teachers working in English-speaking universities in places such as Hong Kong and English-medium programmes in European universities would readily grasp the educational value of this argument. As contemporary societies become increasingly diverse, these emerging phenomena are likely to form part of our 'normal' language landscape. The implications for assessment are clear: any attempt at assessing speakers' proficiency in a context- or domain-sensitive way should take into account situated language practice/s. The corollary of this is that the standardised large-scale international academic English Language tests need to be re-conceptualized in the form of smaller and more bespoke assessments that can supple-ly meet local needs. At this time there is an urgent need for further conceptual and technical discussion on the relationship between emergent linguistic forms and practices and the basic tenets of psychometrics (including construct) and other approaches to language assessment. This work has barely begun (cf. Jenkins & Leung 2014, Jenkins & Leung in press).

5. Concluding remarks

We end by suggesting a way forward. There is evidently a need to move away from the large-scale testing solely premised on the monolithic notion of standardized English as used by the idealised user. Although we acknowledge that international testing organisations such as ETS and Pearson recognize the need to incorporate diverse accents and styles of English into current test design to make them more internationally acceptable (see, for example, Gu & So, 2015), we argue that accepting a wider range of language varieties does not go far enough. Focussing on language varieties (e.g. Australian or Singaporean English) is perhaps missing the point. The key issue is that the use of English for academic purposes is influenced by different disciplines and institutional practices. We believe that for EAP assessment to be authentic and to provide truly meaningful information about university applicants, the currently used all-purpose, large-scale 'fit-for-all' model of assessment needs to be replaced by local assessments that would be both context and domain specific.

We suggest that such a move calls for the following. Firstly, there is an urgent need to rethink communicative effectiveness drawing on insights from research into ELF and multilingualism, particularly translanguaging, as suggested above, but also from an expanded view of communication that acknowledges the significance of student volition and agency in linguistically and ethnically diverse contemporary settings (see Kramsch, 2006, 2010 for a discussion on symbolic competence). Focus should be placed on transcultural communication among both NES and NNES rather than simply on the accuracy of the language used by NNES which, in turn, would suggest that not only NNESs should be assessed: NES should also be required to demonstrate an awareness of the nature of transcultural communication.

Language modelling for assessment purposes also needs to be more nuanced, conscious of variability over time, and informed by the context and situation in which it is to be used. Given that the interactional situations for candidates applying to study in, for example, Singapore or the Philippines and those applying, for example, to New Zealand or South Africa would be very different. The assessments should be informed by the specific situations and language mix that the candidates are likely to encounter. Proficiency criteria should be informed by the demands of the subject area; for instance, those of law or clinical medicine where accuracy is of paramount importance, whereas those of business studies or physics may be very different. Underpinning this is an acknowledgement that there are a number of different proficiencies, linguistic and non-linguistic, English and non-English, rather than a universal proficiency of English scale along which a learner moves as s/he gains mastery of a monolithic version of the English language.

References

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

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

Baird, R., Baker, W., & Kitazawa, M. (2014). The complexity of ELF. *Journal of English as a Lingua Franca 3*(1), 171-196.

Baron, N. (2010). *Always On. Language in an Online Mobile World.* . New York: Oxford University Press.

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

Borsboom, D., Cramer, A.O.J., Kievit, R.A., Scholten, A.Z., & Franić, S. (2009). The end of construct validity. In R. W. Lissitz (Ed.), *The concept of validity: Revisions, new directions, and applications* (pp. 135-170). Charlotte, NC: Information Age Pubishing, INC.

Canagarajah, A. S. (2011). The plurilingual tradition and the English language in South Asia. *AILA Review, 22*, 5-22.

Canagarajah, A. S. (2013). *Translingual Practice*. London: Routledge.

Carter, R., & McCarthy, M. (1995). Grammar and the spoken language. *Applied Linguistics 16*(2), 141-158.

Cheshire, J. (1999). Spoken standard English. In T. Bex & R. J. Watts (Eds.), *The Widening Debate* (pp. 129-148). London Routledge.

Cho, Y., & Bridgeman, B. (2012). Relationship of TOEFL iBT® scores to academic performance: Some evidence from American universities. *Language Testing, 29*(3), 421-442.

Cotton, F., & Conrow, F. (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* (pp. 72-115): International English Language Testing Systems.

Creese, A., & Blackledge, A. (2010). Translanguaging in the bilingual classroom: A pedagogy for learning and teaching? *Modern Language Journal, 94*(1), 103-115.

Crystal, D. (2012). A global language. In P. Seargeant & J. Swann (Eds.), *English in the World. History, Diversity, Change.* Abingdon: Routledge.

Crystal, D., & Davy, D. (1975). *Advanced Conversational English*. London: Longman.

Cummins, J. (2005). A proposal for action: strategies for recognizing heritage language competence as a learning resource within the mainstream classroom. *Modern Language Journal*(89), 585-592.

Davies, A. (2008). *Assessing Academic English: Testing English proficiecny, 1950-1989: The IELTS solution*. Cambridge: Cambridge University Press.

Douglas, D. (2000). *Assessing language for specific purposes*. Cambridge [England]: Cambridge University Press.

Dunworth, K. (2009). An investigation into post-entry English language assessment in Australian universities. Journal of Academic Language and Learning, 3(1), A1–A13

Dunworth, K. (2010). Clothing the Emperor: Addressing the Issue of English Language Proficiency in Australian Universities. The Australian Universities' Review, 52(2), 5–10

Dunworth, K. (2013). Discussion paper 2: In-course student English language development. Five Years On: English Language Competence of International Students Discussion Papers. Melbourne: International Education Association of Australia, pp. 39–67

Dunworth, K., Drury, H., Kralik, C. and Moore, T. (2013). Degrees of Proficiency: Building a Strategic Approach to University Students' English Language Assessment and Development. Final Report. Sydney: Australian Government, Office of Learning and Teaching. www.olt.gov.au/project-degrees-proficiency-building-strategic-approach-university-studentsapos-english-language-ass Retrieved 23 May 2014.

Elder, C., Bright, C., & Bennett, S. (2007).The role of language proficiency in academic success: Perspectives from a New Zealand university. *Melbourne Papers in Language Testing*, 12(1), 24-58.

Feast, V. (2002). The impact of IELTS scores on performannce at university. *International Educational Journal, 3*(4), 70-85.

García, O. (2009). *Bilingual education in the 21st century: A global perspective*. Chichester, West Sussex.

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

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

Gu, L., & So, V. (2015). Voices from stakeholders: What makes an academic English test 'international'? *Journal of English for Academic Purposes 18*, 9-24.

Harding, L. (2014). Communicative language testing: Current issues and future research. *Language Assessment Quarterly, 11*(2), 186-197.

Heller, M. (1999). *Linguistic Minorities and Modernity: A Sociolinguistic Ethnography*. London: Longman.

Ingram, D., & Bayliss, A. (2007). *IELTS as a predictor of academic language perofrmance Part 1*. London: British Council.

Jenkins, J. (2014). *English as a lingua franca in the international university: The politics of academic English language policy*. London: Routledge.

Jenkins, J. (2015). Repositioning English and multilingualism in English as a lingua franca. *Englishes in Practice, 2*(3), 49-85.

Jenkins, J., Cogo, A., & Dewey, M. (2011). Review of developments in research into English as a lingua franca. *Language Teaching, 44*(3), 281-315.

Jenkins, J., & Leung, C. (2014). English as a Lingua Franca. In A. Kunnan (Ed.), *Companion to Language Assessment (Part IV)* (pp. 1605-1616). Hoboken, NJ: Wiley -Blackwell.

Jenkins, J., & Leung, C. (In press). Assessing English as a Lingua franca. In E. Shohamy (Ed.), *Encyclopedia of Language and Education: Vol. 7 Language Testing and Assessment*. New York: Springer.

Jørgensen, J. N., Karrebæk, M., Madsen, L., & Møller, J. (2011). Polylanguaging in superdiversity. *Diversities, 13*(2), 23-37.

Jørgensen, J. N. (2008). Polylingual anguaging around and among children and adolescents. *International Journal of Multililingualism, 5*(3), 161-176.

Jørgensen, J. N., & Møller, J. S. (2014). Polylingualism and languaging. In C. Leung & B. Street (Eds.), *The Routledge companion to Englis Studuies* (pp. 67-83). Abingdon, Oxon: Routledge.

Kachru, B. B. (1992). Teaching World Englishes. In B. B. Kachru (Ed.), *The Other Tongue: English across Cultures (2nd edition)* (pp. 355-365). Urbana, Illinois: University of Illinois Press.

Kane, M. (2004). Certification testing as an illustration of argument-based validation. *Measurement, 2*(3), 135-170.

Kramsch, C. (2006). From communicative competence to symbolic competence. *The Modern Language Journal, 90*(2), 249-252.

Kramsch, C. (2010). The symbolic dimensions of the intercultural (Plenary speech). *Language Teaching* Retrieved October, 2015, from http://journals.cambridge.org/action/displayFulltext?type=1&pdftype=1&fid=7931790&jid=LTA&volumeId=-1&issueId=&aid=7931788

Lea, M. R., & Street, B. V. (1998). Student writing in higher education: an academic literacies approach. *Studies in Higher Education, 23*(2), 157-172.

Lea, M. R., & Street, B. V. (2006). The 'Academic Literacies' model: theory and applications. *Theory Into Practice, 45*(4), 368-377.

Lee, Y.-J., & Greene, J. (2007). The predictive validity of an ESL placement test: A mixed methods approach. *Journal of Mixed Research Methods, 1*(4), 366-389.

Li, W. (2011). Moment Analysis and translanguaging space: Discursive construction of identities by multilingual Chinese youth in Britain. *Journal of Pragmatics, 43*(5), 1222-1235.

Li, W. (2011). Moment Analysis and translanguaging space: Discursive construction of identities by multilingual Chinese youth in Britain. *Journal of Pragmatics, 43*(5), 1222-1235.

Li, W. (2011). Moment Analysis and translanguaging space: Discursive construction of identities by multilingual Chinese youth in Britain. *Journal of Pragmatics, 43*(5), 1222-1235.

Leung, C., & Lewkowicz, J. (2013). Language communication and communicative competence: A view from contemporary classrooms. *Language and Education, 27*(5), 398-414.

Lillis, T., & Scott, M. (2007). Defining academic literacies research: Issues of epistemology, ideology and strategy. *Journal of Applied Linguistics, 4*(1), 5-32.

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

McNamara, T. (2011). Mananging learning: Authority and language assessment. *Language Teaching, 44*(04), 500-515.

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

Milroy, L. (1994). Sociolinguistics and second language learning. In G. Brown, K. Malmkjær, A. Pollitt & J. Williams (Eds.), *Language and Understanding* (pp. 153-168). Oxford: Oxford University Press.

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

Norton, J. (2013). Performaing identies in speaking test: Co-construction revisited. *Language Assessment Quarterly, 10*(3), 309-330.

O'Loughlin, K. (2008). The use of IELTS for university selection in Australia: A case study. IELTS Research Report Vol. 8. Retrieved September, 2015, http://www.ielts.org/pdf/Volume8_Report3.pdf

O'Sullivan, B., & Liu, Y. (2006). *The impact on candidate language of examiner deviation from a set interlocutor frame in the IELTS Speaking Test. IELTS Research Reports 6*. Melbourne: IELTS Australia Pty Limited, British Council.

Otsuji, E., & Pennycook, A. (2010). Metrolingualism: fixity, fluidity and language in flux. *Journal of Multilingualism, 7*(3), 240-254.

Otsuji, E., & Pennycook, A. (2011). Social inclusion and metrolingual practices. *International Journal of Bilingual Education and Bilingualism, 14*(4), 413-426.

Paul, A. (2007). IELTS as a predictor of academic language performance, Part 2. *IELTS research reports vol. 7*, from Retrieved September, 2015 from http://www.ielts.org/pdf/Volume7_Report4.pdf

Rea-Dickens, P., Kiely, R. , & Gu, Y. (2011). Uses and impact of test scores in university admissions processes: the language test as the "hard" criterion. In B. O'Sullivan (Ed.), *Language Testing: Theories and Practices* (pp. 262-281). Basingstoke: Palgrave Macmillan.

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

Tribble, C., & Wingate, U. (2013). From text to corpus – A genre-based approach to academic literacy instruction. *System, 41*(2), 307–321.

Vertovec, S. (2006). *The emergence of super-diversity in Britain. Working Paper No. 25*. University of Oxford: Centre on Migration, Policy and Society.

Vertovec, S. (2007). Super-diversity and its implications. *Ethnic and Racial Studies, 30*(6), 1024-1054.

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.

Woodrow, L. (2006). Academic success of international postgraduate Education students and the role of English proficiency. *University of Sydney Papers in TESOL, 1*, 51-70.