

# ASSESSING REAL-WORLD USE OF ENGLISH AS A LINGUA FRANCA (ELF): A VALIDITY ARGUMENT

Sheryl Cooke\*

*British Council, University of Jyväskylä, Finland  
989 Beijing West Road, Shanghai, 200041, PR China*

Received 26 February 2020

Revised 20 March 2020; Accepted 20 July 2020

**Abstract:** Real-world use of English involves speakers and listeners from various linguistic backgrounds whose primary goal is mutual comprehensibility and the majority of conversations in English do not involve speakers from the Inner Circle (Graddol, 2006; Kirkpatrick, 2007). Yet, rather than focusing on comprehensibility, many tests continue to measure spoken performance with reference to an idealised, native-speaker form, weakening the validity of these tests in evaluating authentic spoken communicative competence as it is used in a global lingua franca context and leading to a narrowing of the construct of ELF, or to the inclusion of construct irrelevant factors.

Validation of a test of English as a tool for global communication includes demonstrating the link between the construct (real-world communicative ability in a particular context) and the test tasks and rating criteria (McNamara, 2006), and evidence to support the interpretation of a test score needs to be presented as part of the overall validity argument. First, this paper argues that the context of English use that many high-stakes test-takers aspire to – that of English for Academic Purposes (EAP) – is frequently an ELF context; second, Toulmin’s (2003) argument schema is leveraged to explore what evidence is required to support warrants and claims that a test provides a valid representation of a test-taker’s ability to use ELF. The framework as it relates to the validation of language tests in general is presented and the model is then applied to two tests of spoken English by way of illustration. Although examples are included, the main aim is to provide a theoretical justification for a focus on comprehensibility and the inclusion of linguistic variation in the assessment of ELF and to present a validation framework that can be applied by test developers and test users.

*Keywords:* English as a lingua franca, test validity, comprehensibility

## 1. Introduction

“Speak English!” said the Eaglet. “I don’t know the meaning of half those long words, and I don’t believe you do either!”

— Lewis Carroll, 1865 in *The Adventures of Alice in Wonderland*

English has long been recognised as the major international language for

communication across a range of different domains, from academic conferences to business negotiations, from aviation to the international space station, from the United Nations to popular culture. ‘English’, however, is a broad term that encapsulates a growing range of Englishes, from the forms spoken in the traditional English-speaking countries of the UK, North America and Australasia, to the now established varieties spoken in ex-British colonies such as Singaporean English, South

\* Tel.: +8613818274299

Email: Sheryl.cooke@britishcouncil.org.cn

African English and Indian English, and extending further to the learning and active use of English by speakers from a myriad of different language backgrounds for the purpose of international communication.

Native-speaker (NS) – and particularly ‘standard’ forms of NS English such as Southern British English or general American English – have long held an elevated position in the teaching and learning of English. Quirk (1985) argued that a standard form served the needs of non-native English speakers (NNS) because their communicative purposes were ‘narrow’. Others have attempted to defend or define Standard English: Williams (1980) with a focus on a standard form of US English and Peters (1995) with Australian English. Davies (1999) concluded that *the* Standard form could be considered the ‘language of the educated’.

Berns (2006, p. 723-724) outlines the key assumptions behind NS norms:

“(1) everyone learning English does so in order to interact with native speakers;

(2) the communicative competence learners need to develop is the native speaker’s; and

(3) learning English means dealing with the sociocultural realities of English or the US, that is, British or American ways of doing, thinking and being.”

But evidence from the real-world use of English debunks these assertions: real-world communication in English involves speakers and listeners from various linguistic backgrounds whose primary goal is successful communication. Furthermore, almost 20 years ago, Crystal noted that only a quarter of the world’s English language speakers are NS users (Crystal, 2003); Ethnologue (Paul, Simons & Fennig, 2020) puts the total number of users of English in all countries at 1,268,100,190 of which 369,704,070 use it as

an L1 and 898,396,120 as an L2, with other words, 71% of the world’s speakers of English are not L1 users of the language. Given the international function of English in the geopolitical, economic and academic spheres (Ammon, 2010), it is clear that the majority of conversations in English do not involve speakers from the countries in Kachru’s (1985) well-known Inner Circle (Graddol, 2006; Kirkpatrick, 2007; Nelson, 2011). Learners of English do not, in the majority of cases, have as their goal conversation with native speakers, they do not need to develop native speaker proficiency to achieve their communicative goals, and the sociocultural reality they operate in is diverse, dynamic and more likely to involve cultural and social characteristics of the Chinese, Indians or Brazilians than someone from the Inner Circle. The communicatively successful use of ELF by millions of speakers from varied backgrounds occurs in a range of different domains, both personal and professional (Seidlhofer, 2011).

There has been increasing recognition of the need for language learning and teaching to reflect these realities and support for this has been voiced in the academic community – Jenkins (2000) and Seidlhofer (2011), amongst others. Galloway (2018) explicitly points to the importance of teaching learners to communicate in a global context. There is also a move towards more inclusive course books that reflect the sociocultural reality referred to above, e.g. MacMillan Global course books, and the inclusion of a variety of NS and NNS accents in listening texts, although to what degree these attempts go beyond surface-level recognition of the reality of English use is debatable (Galloway, 2018).

In the field of language assessment specifically, the question of which English to test is particularly pertinent given the

consequences tests have on the lives of individuals, on society more generally, and the washback effect that a high-stakes test has on teaching and learning. If a test-taker is to be tested on a certain variety of English, or a narrow range of varieties, then that test-taker will focus on studying those varieties and seek out exposure to those forms, even where this does not reflect their current or future communicative context. In a chicken-and-egg situation, the varieties that students learn to prepare for the standard-form tests are then used to support continued testing of only standard forms of English because exposure to other forms is limited in the classroom and in textbooks.

The question of whether, and how, to reflect real-world use of English in tests has been discussed by, amongst others, Elder & Harding (2008), Jenkins (2006), McNamara (2014) and Harding & McNamara (2017). Graddol, too, explicitly mentions testing in relation to the new sociocultural reality of English language use: “The way English is taught and assessed should reflect the needs and aspirations of the ever-growing number of non-native speakers who use English to communicate with other non-natives” (Graddol, 2006, p. 87).

This paper presents a theoretical perspective to the challenge of assessing English as a lingua franca (ELF) and test validity is at the core of the discussion. Frameworks used for the evaluation of test validity are presented and Toulmin’s (2003) argument framework is applied to the context of assessing ELF. I will present an argument that centres around two key assumptions:

- the first, central to the contemporary idea of validity in language assessment, is that *a test must reflect the real-world use of English in order to be valid.*
- the second, at the core of the study of the global use of English and the study of

World Englishes, is that *the real-world use of English is not limited to standard forms of the language but includes variety.*

What follows from this is that in order to be a valid assessment, a test must be linked to the domain of use and, in order to demonstrate validity, must present evidence to support the claim that the domain is not only *represented* but *adequately represented* in the test. The sections of this paper that follow consider validity and the link to the intended domain of use, investigate the domain in which candidates taking high-stakes tests for the purpose of academic study in an English-speaking context are likely to function, and use Toulmin’s argument structure as an example of how evidence may be sought, presented and evaluated. Finally, an illustration of how the framework might be applied to two language tests is presented and next steps are suggested.

## 2. Validity

There are various approaches to establishing the validity of a test or assessment system, some more theoretical than others (Messick, 1989; Kane, 2012). Common to many approaches is the investigation of what tasks the test-taker needs to engage in in the real-world situation in which they communicate or intend to communicate. That is, a test must have a demonstrable link to the context – or domain – in which the ability is or will be put to use.

Various scholars have highlighted the importance of the link between the test and the context of use. Mislavy & Yin’s (2012) work on *Evidence Centred Design* outlines a chain of reasoning that starts with domain analysis and then moves on to the crucial stage of domain modelling in which the test construct or ability is articulated: what claims are we making about the test-taker, what evidence do we need to substantiate those

claims, and what tasks we will need to elicit that evidence. Bachman and Palmer (1996) describe this domain as the Target Language Use (TLU) situation. Kane (2012) presents the link between domain of use – observation

of an individual’s performance on a particular task – and the decisions that are made about an individual’s ability as a chain of inferences as illustrated in *Figure 1* below.

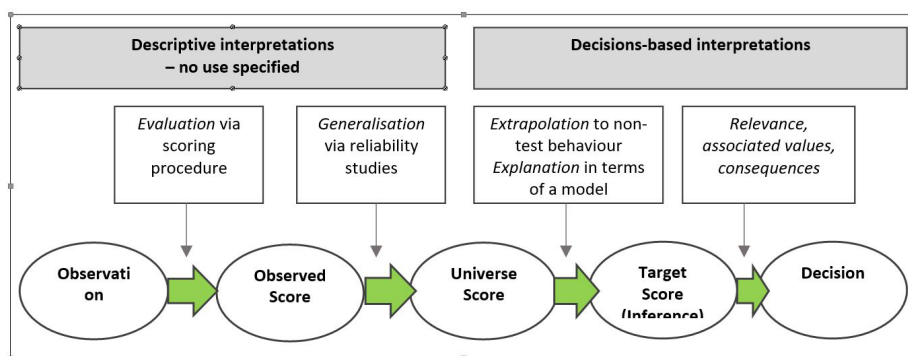


Figure 1: Kane’s chain of inferences (McNamara & Roever, 2006)

In their operationalisation of what are relatively abstract validity theories, O’Sullivan and Weir (2011) connect the domain of use and the claim about an individual’s ability through four questions focusing on test-taker characteristics (Who are we testing?), the construct or ability (what are we testing?), the tasks used to elicit that ability (How are we

testing it?) and the assessment criteria (What system will be use to score it?); the interaction of these questions is presented in *Figure 2* below. They go on to explicitly state:

“Unless we can demonstrate empirically that... they demonstrate a link between the underlying concepts, our test is unlikely to allow us to make valid inferences.” (2010, p. 23).

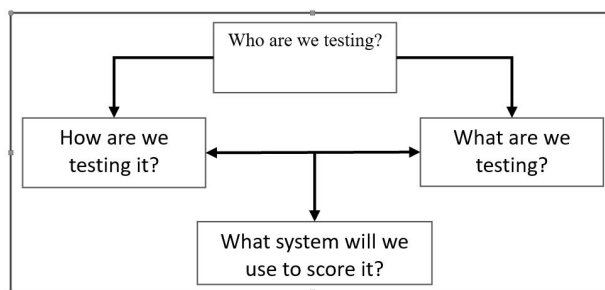


Figure 2: Operationalisation of validity theories (Weir& O’Sullivan, 2010)

What is echoed throughout these approaches and operationalisations of test validity is that the domain of use needs to be reflected in the tests. To put it in a different way: in order to make a plausible decision about whether someone has the ability to perform a certain communicative task in a certain communicative situation, the link between the test and the domain of use

must be demonstrated. If a test is shown to misrepresent or underrepresent the domain to which it purports to link, then test validity is threatened.

Crucial to identifying whether a test is valid is understanding the intended domain of use or the TLU, in Bachman and Palmer’s terms. One argument for continuing to use NS norms in testing is that the domain of use is

characterised by standard forms of English (Berns, 2006 - see above). Is this indeed the case? In the following sections, I attempt to answer two questions:

- What does the domain / Target Language Use situation / underlying construct of English as a lingua franca in an EAP context look like?

- What argument can be developed to demonstrate the link between a test and the ELF construct and what evidence is needed to support this argument?

Finally, the argument structure is briefly applied to two tests by way of example.

### 3. The Domain of Use

#### *English as a lingua franca*

Definitions of ELF are as numerous as the different ways of referring to the broad concept of English, or Englishes, that are used as a common language of communication. The online Cambridge dictionary defines *lingua franca* as follows, using English in the example of use:

lingua franca

noun [C usually singular]

A language used for communication between groups of people who speak different languages: *The international business community sees English as a lingua franca.* (Cambridge Dictionary, n.d.)

Widdowson also refers to this commonality of communicative form between speakers who share no other:

“English as a lingua franca is the communicative use of linguistic resources, by native and non-native speakers, when no other shared means of communication are available or appropriate.” (2013, p.190)

As does Seidlhofer: ELF is “communication in English between speakers with different first languages” (Seidlhofer, 2005, p. 339); later, she describes it as, “any

use of English among speakers of different first languages for whom English is the communicative medium of choice, and often the only option” (Seidlhofer 2011, p. 7).

But to use only these definitions is to simplify the situation and to ignore the rapidly changing dynamics of global communication. Seidlhofer, in 2009, argued for a new perspective on ELF and particularly how it is being influenced by new technologies. Figure 3 below illustrates the traditional view of English as a global form of communication, the new way in which we could or should conceptualise ELF, and the catalysts driving this change.

The left-hand side of the diagram depicts the Circles that Kachru used to describe the Englishes and English use that are by this time, very familiar to most applied linguists. They include:

- the Inner Circle – varieties attributed to mother-tongue or ‘Native Speakers’, typically the UK, US, Australia, Canada, New Zealand and Ireland;

- the Outer Circle – describing varieties that have emerged following decolonisation of the British Empire in the 1950s and 1960s, usually nativised with a corresponding written form and strongly associated with identity in the post-colonial world, such as Indian English, Singaporean English, Nigerian English, Jamaican English; and,

- the Expanding Circle describing the use of English by those who do not fall into the first two Circles – a learned form of English, usually an Inner Circle form, generally assessed in relation to these standard forms, with deviations from these varieties described in terms of errors or fossilisation.

In Seidlhofer’s 2009 argument for the re-evaluation of what defines English for the majority of users she points to two key ‘push’ factors:

- technology-enabled communication allows for increased international contact where we are no longer confined to communicating with people within our own immediate physical environments, but with a range of people from around the world;

- a move towards *communities of practice* rather than physical communities, academic communities being a good example, such as a language assessment and language learning community of practice.

The right-hand section of Figure 3 illustrates how global interaction and communities of practice cut across Kachru's circles, resulting in the need for a different conceptualisation of what constitutes ELF in different settings, a characteristic that Leung & Lewkowitz (2006) and Canagarajah (2007) have also pointed out.

Seidlhofer sums the situation up as follows:

“With the current proliferation of possibilities created by electronic means and unprecedented global mobility, changes in communications have accelerated and forced changes in the nature of communication. And for the time being anyway, it is English as a lingua franca that is the main means of wider communication for conducting transactions and interactions outside people's primary social spaces and speech communities. It seems inevitable that with radical technology-driven changes in society, our sense of what constitutes a legitimate community and a legitimate linguistic variety has to change, too.” (2009, p. 238)

There have been moves towards a definition of the construct of ELF, with corpus linguistics driving much of the outcomes, e.g. the English as a lingua franca in Academic Settings corpus which draws on data from speakers of 51 different first languages, and the Vienna-Oxford International Corpus of English (VOICE) which “seeks to redress the balance [between the predominantly

NNS of English and the NS-referenced linguistic description of the language] by providing a sizeable, computer-readable corpus of English as it is spoken by this non-native speaking majority of users in different contexts”. The move away from seeing English as being an inherently NS-domain is also evident in the most recent CEFR review – the CEFR Companion Volume (2018) has removed all references to ‘native speaker’ in any of the can-do statements. Finally, Jenkins's proposal of a lingua franca Core for phonology is well-known but has been only minimally adopted, partly as a result of a paucity of supporting evidence. Isaacs cautions, “substantially more empirical evidence is needed before the lingua franca Core can be... adopted as a standard for assessment” (2013, p. 8).

The lack of construct definition for ELF and its fluid and dynamic character, in addition to socio-political factors, are possible reasons for the continued reliance on Inner Circle of English in high-stakes tests, even in the face of evidence that standard forms are reflected to a lesser degree than other varieties in the domain of use. For example, the International English Language Testing System (IELTS) refers to NS norms in both the Grammar and Vocabulary and Pronunciation evaluation criteria (my emphasis in bold):

**G&V:** produces consistently accurate structures apart from ‘slips’ characteristic of **native speaker speech**

**Pronunciation:** is easy to understand throughout; **L1 accent** has minimal effect on intelligibility

<https://www.ielts.org/-/media/pdfs/speaking-band-descriptors.ashx?la=en>

Pearson, another major testing organisation, does so, too: “Pronunciation reflects the ability to produce consonants, vowels, and stress in a **native-like manner** in sentence context” (my emphasis in bold) (2011, p. 12).

The Educational Testing Service (ETS) ETS SpeechRater programme, an automated rating system, likewise appears to establish a NS benchmark by using a ‘pronunciation dictionary’, based on NS standards (with some alternative pronunciations) (Xi et al, 2008).

The continued use of predominantly NS varieties in major language tests calls into question the validity of the assessments if they are used to decide whether an individual

is able to function in a context that is not shaped according to NS norms of English communication. Although the construct of English communicative ability is fluid and changing, the wider context of use – the community of practice – will ultimately shape the construct definition; some of these domains are stable enough to offer a more solid description of the ELF construct associated with them. This is what we turn to next.

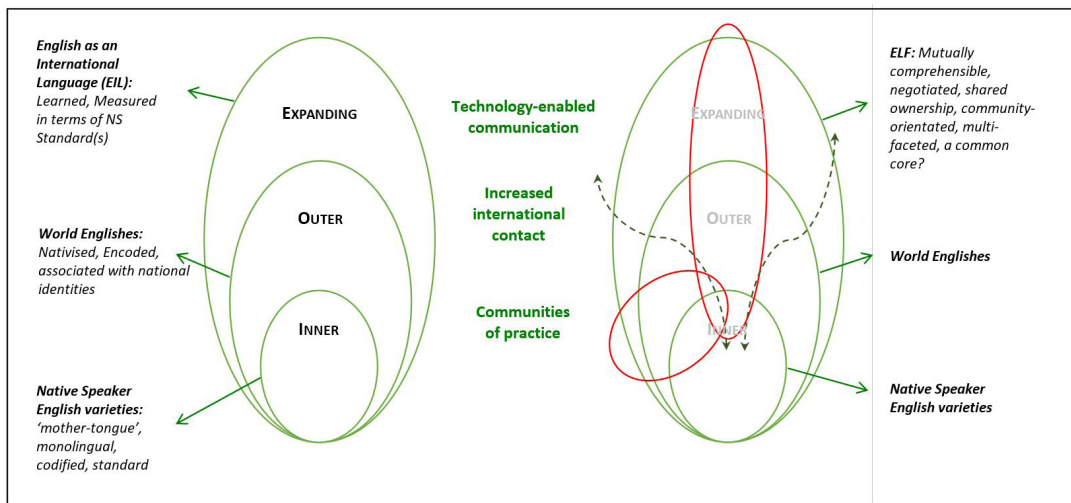


Figure 3: Changes in the conceptualisation of English(es) and English as a lingua franca

### *The domain of use*

As suggested above, communities of practice are diverse, dynamic, and potentially overlapping. In the interests of brevity and the conciseness of an example validation argument, the scope of the discussion in this paper is limited to English for Academic Purposes (EAP), specifically the context of universities where English is the medium of instruction (EMI) and academic discourse. Given the large number of EMI institutions worldwide, the investigation of the domain of use is further limited to data for two major EMI destinations for international students, the UK and Australia. Indeed, EAP is one of the most prevalent uses of high-stakes, international tests of English proficiency.

Before considering whether a test is valid as an instrument to decide whether someone’s language ability is adequate to function in an EAP environment, the domain of use needs to be understood.

The following statistics allow us to better understand the EAP domain of use at universities in the UK and Australia. Figure 4 shows that more than two-thirds of the 2016-2017 cohort of students at UK universities were not from the UK but from a range of backgrounds, both EU and non-EU. Thus, it would follow that someone preparing for post-graduate study in the UK should expect to interact with a variety of fellow students from a wide range of language and cultural backgrounds, some NS, but the majority NNS.

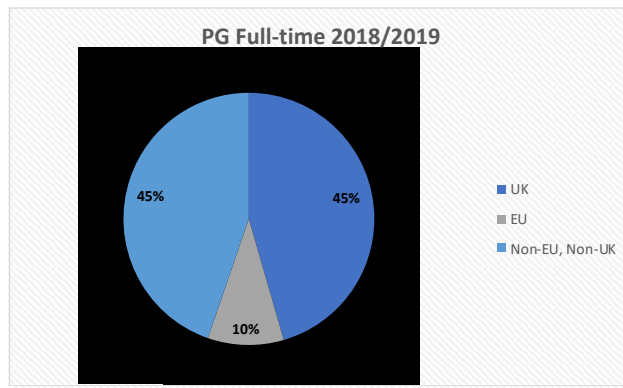


Figure 4: Higher Education student enrolments in Post-Graduate full-time study across the UK by domicile academic year 2018 - 2019.

Lecturers are also key stakeholders in the educational milieu of international students and they need to be understood and communicated with effectively. Universities UK, a collaboration of 137 universities across the UK puts the percentage of international staff working at UK universities at 30% (Universities UK international, 2018); Figure 5 shows that more than 30% of UK academic

staff in 2018/2019 were not from the UK suggesting that, aside from regional UK accents, international students would need to understand and interact effectively with a range of lecturers and tutors from different backgrounds and with a variety of accents.

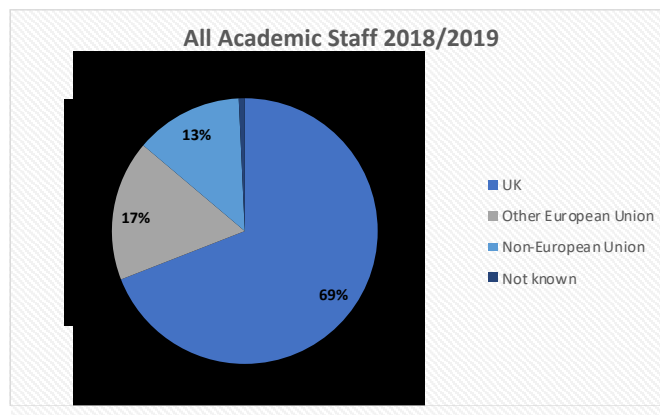


Figure 5: Percentage of academic staff employed at UK Higher Education institutions in the UK.

Figure 6 shows that the picture is no different in Australia where international enrolments are on the increase, meaning that students at university have a strong likelihood of interacting with someone who speaks a non-Inner Circle variety of English, or speaks English as their second

or third language. At post-graduate level, international students make up over 40% of the study body (Figure 7).



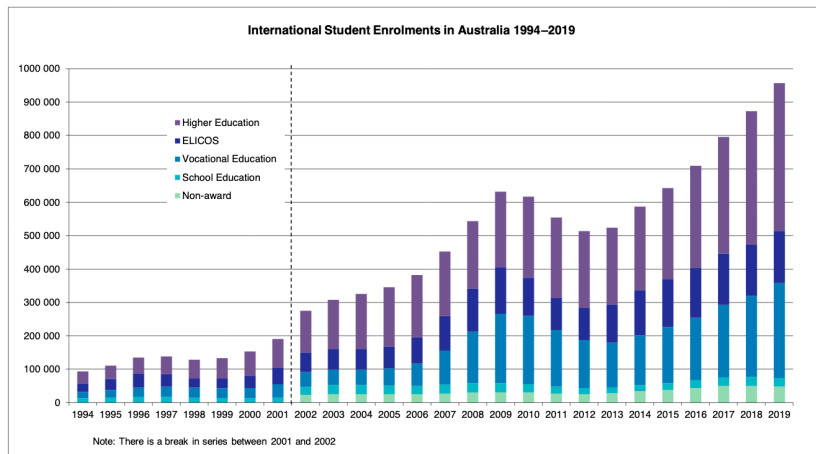
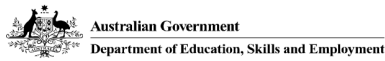


Figure 6: Rise in international student enrolments in Australia 1994 – 2019.

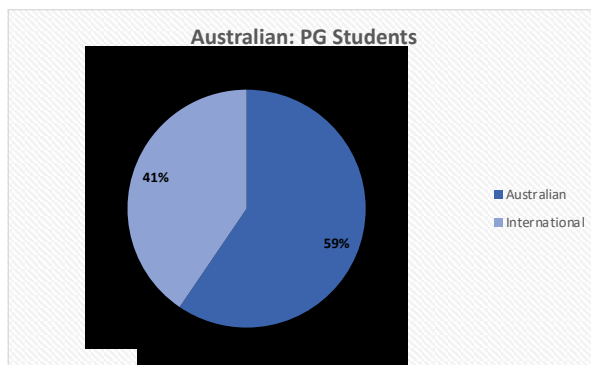


Figure 7: Percentage of on-campus post-graduate students enrolled in Australian Higher Education institutions by origin (possibly 2018 – exact year unclear from available data).

The recognition of this increasing internationalisation of Australian higher education is echoed in the press:

“Because the Government has effectively capped the number of domestic students, international students are becoming an increasing percentage of all students,” Mr Norton said.<sup>1</sup>

It is also investigated by research institutes: for example, the Grattan Institute

reports that in 2018, just under three-quarters of students enrolled in Australian higher education institutions were Australian citizens or permanent residents.

Even this very superficial consideration of the domain of use – tertiary EMI institutions in two traditionally Inner Circle countries – suggests that NS norms and standard forms of English should not be the only varieties to be tested if linguistic preparedness for these contexts is the primary ability being evaluated. Instead, the context of use suggests that we should be evaluating someone’s ability

<sup>1</sup> <https://www.abc.net.au/news/2018-04-18/australia-hosting-unprecedented-numbers-international-students/9669030>

to communicate effectively with a range of speakers from different L1 backgrounds and that, rather than assessing proficiency in relation to a NS norm, test developers should be considering comprehensibility and ensuring that assessment for the situations described above is inclusive of variety as long as the comprehensibility principle is met.

#### 4. A validity argument

##### *The argument framework*

Toulmin's argument schema is a tool for the evaluation of a claim (2003). Conceptualised as framework for the analysis of legal arguments, the schema is useful to guide test-developers in evidence-centred design and to support the validity assertions of their tests. It also provides a useful tool for those analysing the veracity of the validity claims of a language test. The latter is what is especially appealing about the framework in terms of language test evaluation: it helps to identify the types of evidence necessary to support a claim. An example of Toulmin's argument structure as applied in general to tests of spoken English within the ELF context is presented in Figure 8. Note that, in the interests of brevity and as the aim of this paper is to present an *example* of how the argument structure can be applied to the assessment of ELF, the details in Figure 8 pertain only to the evaluation of *spoken performances*. In the following section, an example of how this framework can be applied to evaluating the validity of two well-known English language is presented.

At the core of the argument are the facts necessary to support the overall claim of validity. In the case of ELF, we would want to know that the *Scores on a test of spoken English reflect a speaker's ability to make themselves understood in an international context* such as the EAP domains considered above. The

facts (the grounds) that would act as the basis for this claim are that *the means of performance elicitation and the assessment criteria applied to the performances focus on comprehensibility of and by a range of different L1 speakers*.

It is, of course, necessary to substantiate the facts in order to link the grounds to the overall claim. In our example, this link between claim and underlying grounds exists as long as *the observed performances in the test provide observed scores reflective of an ability to be comprehensible to a wide range of English speakers from different L1 backgrounds* – the warrant. To evaluate whether this condition has been met, concrete evidence needs to be presented. Figure 8 suggests three key areas in which evidence can be presented and according to which we might consider in evaluating the validity of the claim: the task types that are used to elicit the performance (are they reflective of the context of use? – assumption 1); the rating criteria (do they have comprehensibility rather than native-speakerhood as a benchmark? – assumption 2) and, a question related to reliability – are the evaluation criteria applied accurately and consistently? – assumption 3). The data that could be used to substantiate or refute these assumptions are described in Figure 8, both qualitative as well as quantitative data that is necessary to support the overall claim.

Finally, legal-orientated rebuttals presented on the right-hand side of the diagram provide useful jumping-off points for a critical analysis of a language test being used to assess a test-taker's readiness to function in an ELF context:

- are the tasks on the test comprised of only NS linguistic and cultural input?
- is preference for a NS accent evident in the rating scales?
- who are the raters? are they made up only of NSs or is there adequate representation of a

range of proficient users of ELF? This applies to tests with human raters as well as tests where machines ‘learn’ from a pool of human raters.

Figure 8 presents just one example of how Toulmin’s argument structure can be used to evaluate tests from the ELF perspective.

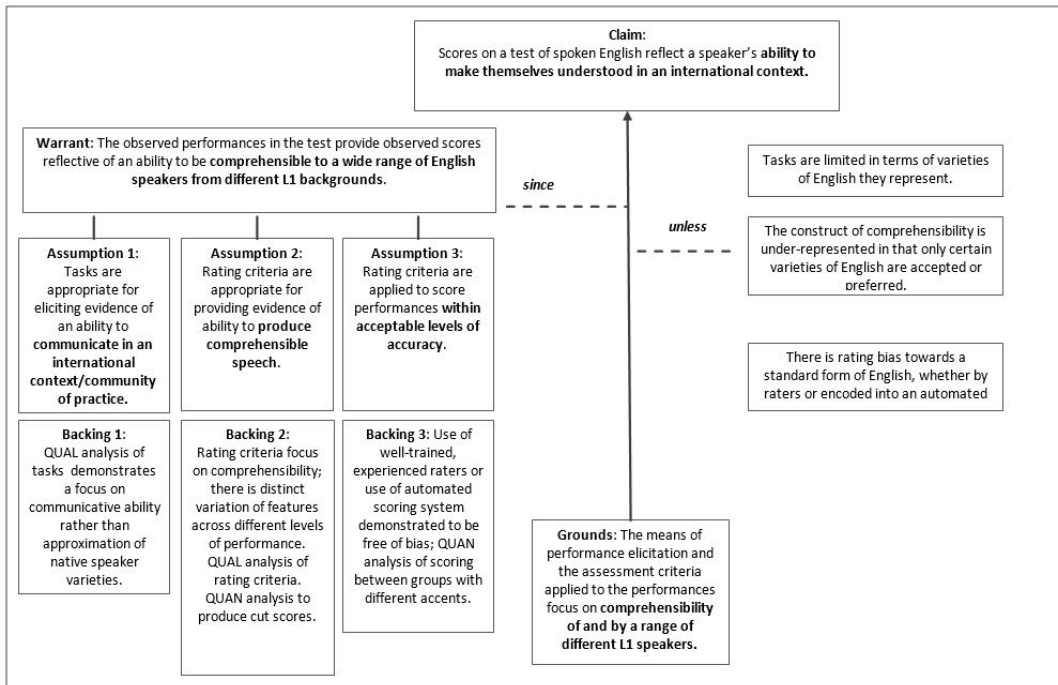


Figure 8: An argument structure for the evaluation of a language test in an ELF context

**Applying the argument framework: an example**

This section presents a brief analysis of two tests to illustrate the potential application of the argument framework presented above. This is by no means intended to be a detailed analysis of any of the tests; rather, the aim is to provide an example of how the strength of the link between test and domain of use can be investigated using the argument framework to establish the validity of a test.

Assumptions:

1. The tests are being taken as predictors of ability to communicate in the domain of use explored above, i.e. EAP in the UK/Australia.
2. The ability under scrutiny is the production of spoken English and, as such, the focus of the mini-analysis are the speaking modules or components thereof.

The two tests under consideration are the International English Language Testing System

(IELTS) and the Pearson Test of English (PTE). These are both high-stakes tests, frequently required for entrance to higher education in the UK or Australia. The tests are different in that IELTS is delivered and rated by humans while PTE is delivered by computer and rating is automated, i.e. Artificial Intelligence (AI) is used to assign scores to spoken performances. Given that this is an illustration of the application of the argument framework above, one specific task was focused on for PTE due to the variety of tasks included in the integrated listening-speaking module; IELTS, however, is somewhat more homogenous in nature as the entire speaking test consists of a 12-14 minute Oral Proficiency Interview (OPI) with a trained interlocutor.

Table 1 below illustrates how the assumptions taken from figure 8 can be supported and rebutted for each of the two

tests. While this illustrative analysis has drawn only on publicly available materials, a more robust analysis by the test developers or those engaged in evidence-based, critical selection of tests for EAP purposes could – and should – include independent research to obtain the necessary evidence.

The table below also shows that some evidence can support the validity of a test while other evidence undermines those claims. It also illustrates how an analysis of test validity must be linked to the purpose for which the test will be used.

Table 1: Application of the argument framework to two language tests – an example

IELTS [Overall speaking module]		PTE [Task 1.5 – Listen and Retell]	
<i>For related citations and sources please see below.</i>			
Warrant: The observed performances in the test provide observed scores reflective of an ability to be comprehensible to a wide range of English speakers from different L1 backgrounds.			
Supporting evidence	Rebuttal	Supporting evidence	Rebuttal
Assumption 1: Tasks are appropriate for eliciting evidence of an ability to communicate in an international context/ community of practice.			
Interaction with an interlocutor, providing a reasonably authentic communication context.	If all interlocutors are NS, the test context is not reflective of an international context and the construct is narrowed.	Task 1.5 is an integrative task that reflects the EAP setting, i.e. listening to a lecture and then speaking to summarise what was heard.	If the listening is always a NS, this does not reflect an international context and the construct is narrowed.
Assumption 2: Rating criteria are appropriate for providing evidence of ability to produce comprehensible speech.			
The analytical rating scales include reference to “intelligibility”. Academic research shows a link between scores and performance in an EAP setting.	The analytical rating scales include reference to “slips’ characteristic of native speaker speech”.	Scoring criteria include “how accurately and thoroughly” meaning is conveyed, i.e. a focus on content.	The scoring description includes reference to “regular speakers” of English; score guide explicitly lists “native-like” as the highest level of proficiency, above “advanced” for both fluency and pronunciation.
Assumption 3: Rating criteria are applied to score performances within acceptable levels of accuracy.			
Several means are in place to ensure consistency in the marking of the writing and speaking tests including robust recruitment and training, standardisation and monitoring, as well as statistical analysis of results.	Standard Error Measurement in human rating of productive skills may allow more tolerance of bias by some raters.	AI scoring removes potential human bias towards different accents, for example.	Educated, proficient or NS speakers score poorly on the test, e.g. reference below.
Sources for above			

IELTS public band descriptors (<https://www.ielts.org/>):  
*Produces consistently accurate structures apart from 'slips' characteristic of native speaker speech. [Grammar and vocabulary criterion]*  
*Is effortless to understand [Pronunciation criterion]*  
<https://www.ielts.org/teaching/examiner-recruitment-and-training>  
<https://www.ielts.org/about-the-test/ensuring-quality-and-fairness>  
*"The clearest finding emerging from this research is the predictive validity of IELTS scores in relation to general language performance."* Ingram & Bayliss: 2007, p. 59

PTE: Listen and Retell <https://pearsonpte.com/the-test/format/english-speaking-writing/re-tell-lecture/>  
*Content is scored by determining how accurately and thoroughly you convey the situation, characters, aspects, actions and developments presented in the lecture.*  
*Pronunciation: Does your response demonstrate your ability to produce speech sounds in a similar way to most regular speakers of the language?*  
*Pronunciation is scored by determining if your speech is easily understandable to most regular speakers of the language. The best responses contain vowels and consonants pronounced in a native-like way, and stress words and phrases correctly. Responses should also be immediately understandable to a regular speaker of the language.*  
*PTE Academic recognizes regional and national varieties of English pronunciation to the degree that they are understandable to most regular speakers of the language.*  
 PTE Score Guide for pronunciation: 5 Native-like 4 Advanced 3 Good 2 Intermediate 1 Intrusive 0 Non-English  
 News article (see references): *Irish vet fails oral English test*

## 5. Next steps

ELF is centred around the concept of mutual comprehensibility. In order to move towards testing of the comprehensibility of spoken English rather than an approximation to NS varieties, several steps need to be taken:

- achieving a better understanding of what constitutes comprehensibility;
- ensuring that research into comprehensibility is not limited to NS assessments of what is comprehensible but includes the perceptions of a range of speakers of English to reflect the real-world communication context;
- encouraging test developers to take active steps to better reflect the ELF context; for example, ensuring that listening tests include a range of accents, and removing reference to 'native-like' speech in rating rubrics;
- guarding against encoding bias into the algorithms of automated assessment systems by not relying only on NS reference points;
- raising awareness amongst test users that 'English' does not only include the Englishes of the Inner Circle and that the goal is to be comprehensible to listeners from a

wide range of different language backgrounds; as consumers of commercial language tests, test users have the power to influence the test developers.

## 6. Conclusion

This paper presented the notion that real-life use of English is not limited to the use of standard forms of English but includes the use of a myriad of Englishes that facilitate common understanding: ELF. The construct definition of such varied, dynamic use of language presents challenges to the language assessment community but does not mean that it can be ignored – the consequences to test-takers and society more broadly are too great. While researchers work on understanding more about the underlying construct of ELF – a crucial component to more reflective, equitable and fit-for-purpose testing – critical questions should be asked about the validity of current language tests to drive test developers in the direction of a more equitable, fair and inclusive evaluation of this global lingua franca.

The question of whether a test is valid is inextricable from its purpose and the context of language use to which the test scores are

linked. Where the communicative environment is peopled by so many different voices from different language backgrounds interacting in English, it is becoming increasingly necessary for language assessment tools to reflect this, and that test users – educational institutes, Ministries of Education, immigration agencies, employers and test-takers themselves – seek the relevant assurance that tests do, indeed, do this. The example of the argument framework presented in this paper demonstrates a powerful tool with which to identify the crucial questions that need to be asked and types of evidence that should be demanded as proof that a test is valid in a real-life context. While Toulmin's framework serves as a robust tool for critically evaluating tests in the context of ELF use, it can also provide a blueprint for test developers keen to be fair and inclusive in the design of their assessments to better reflect the real-life use of English.

## References

- Ammon, U. (2010). World languages: Trends and futures. In Coupland, N. (Ed.), *The handbook of language and globalization* (pp. 101 – 122). Oxford: Blackwell Publishing.
- Australian Associated Press. (2017, August 8). *Computer says no: Irish vet fails oral English test needed to stay in Australia*. <https://www.theguardian.com/australia-news/2017/aug/08/computer-says-no-irish-vet-fails-oral-english-test-needed-to-stay-in-australia>
- Australian Education Network, (2020). University Rankings. <https://www.universityrankings.com.au/international-student-numbers/>
- Australian Government Department of Education, Skills and Employment. (2020). *International student enrolments in Australia 1994-2019*. [https://internationaleducation.gov.au/research/International-StudentData/Pages/InternationalStudentData2019.aspx#Annual\\_Series](https://internationaleducation.gov.au/research/International-StudentData/Pages/InternationalStudentData2019.aspx#Annual_Series)
- Bachman, L.F. and Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Berns, M. (2006). Word Englishes and communicative competence. In B.B. Kachru, Y. Kachru, & C.L. Nelson (Eds.), *The handbook of world Englishes*, (pp 718 – 730).
- Cambridge Dictionary. (n.d.). Lingua franca. In *Dictionary.Cambridge.org*. Retrieved 19 May 2020, from <https://dictionary.cambridge.org/dictionary/english/lingua-franca>
- Canagarajah, S. (2007). Lingua franca English, multilingual communities, and language acquisition. *The Modern Language Journal*, 91(1), 923–939. doi-org.ezproxy.jyu.fi/10.1111/j.1540-4781.2007.00678.x
- Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.
- Council of Europe (2018). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Companion Volume*. <https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989>
- Crystal, D. 2003. *English as a global language (Second Edition)*. Cambridge: Cambridge University Press.
- Davies, A. (1999). Standard English: Discordant voices. *World Englishes*, 18(2), 171–186. doi.org/10.1111/1467-971X.00132
- Elder, C. & Harding, L. (2008). Language testing and English as an international language: Constraints and contributions. *Australian Review of Applied Linguistics*, 31(3), 34.1-34.11. doi:10.2104/ara10834
- ELFA 2008. *The Corpus of English as a lingua franca in Academic Settings*. Director: Anna Mauranen. <http://www.helsinki.fi/elfa> (accessed 26 April 2020).
- Galloway, N. (2018). ELF and ELT teaching materials. In J. Jenkins, W. Baker, & M. Dewey (Eds.), *The Routledge handbook of English as a lingua franca* (pp. 468-480). Routledge.
- Graddol, D. (2006). *English next*. London: British Council.
- Harding, L., & McNamara, T. (2017). Language assessment: The challenge of ELF. In J. Jenkins, W. Baker & M. Dewey (Eds.), *The Routledge handbook of English as a lingua franca*, (pp. 570–582).
- Higher Education Statistics Agency. (2019). Figure 6 - All staff (excluding atypical) by equality characteristics 2017/18. Ref. ID: SB253 Figure 6. <https://www.hesa.ac.uk/data-and-analysis/sb253/figure-6>
- Higher Education Statistics Agency. (2020). Figure 7 - HE student enrolments by level of study, mode of study and domicile 2018/19. Ref. ID: SB255 Figure 7. <https://www.hesa.ac.uk/data-and-analysis/students/where-from>
- Ingram, D. & Bayliss, A. (2007). *IELTS as a predictor of academic language performance, Part 1*. In: IELTS Research Reports, Volume 7, 2007. British Council and IELTS Australia Limited.
- International English Language Testing System (IELTS). *IELTS Speaking: Band Descriptors (public version)*. British Council, IDP: IELTS Australia and Cambridge English Language Assessment. <https://www.ielts.org/-/media/pdfs/speaking-band-descriptors.ashx?la=en>

- Isaacs, T. (2013). Assessing pronunciation. In Kunnan, A. J. (Ed.), *The Companion to Language Assessment Vol. I* (pp. 140–155). doi.org/10.1002/9781118411360.wbcla012
- Jenkins, J. (2000). *The phonology of English as an international language*. Oxford: Oxford University Press.
- Jenkins, J. (2006). The spread of EIL: a testing time for testers. *ELT Journal*, 60(1), 42–50. doi.org/10.1093/elt/cci080
- Kachru, B.B. (1985). Standards, codification and sociolinguistic realism: the English language in the outer circle. In Quirk, R. and Widdowson, H. (eds). *English in the World: Teaching and Learning the Language and Literatures*. Cambridge: Cambridge University Press, pp. 11–30.
- Kane, M. (2012). Articulating a validity argument. In G. Fulcher & F. Davidson (Eds.), *The Routledge handbook of language testing* (pp. 34–47). Oxford and New York: Routledge.
- Kirkpatrick, A. (2007). *World Englishes: Implications for international communication and English language teaching*. Cambridge: CUP.
- Leung, C. and Lewkowicz, J. (2006). Expanding horizons and unresolved conundrums: Language testing and assessment. *TESOL Quarterly*, 40(1), 211–234. doi-org.ezproxy.jyu.fi/10.2307/40264517
- Macmillan Education Limited. (n.d.). *Global Coursebook*. <http://www.macmillanglobal.com/about/the-course>
- McNamara, T. (2006). Validity in language testing: The challenge of Sam Messick's legacy. *Language Assessment Quarterly*, 3(1), 31–51. doi:10.1207/s15434311laq0301
- McNamara, T. and Roever, C. (2006). *Language testing: The social dimension*. Oxford: Blackwell.
- McNamara, T. (2014). Evolution or revolution? *Language Assessment Quarterly*, 11(2), 226–232. DOI: 10.1080/15434303.2014.895830
- Messick, S. (1989). Validity. In R.L. Linn (Ed.), *Educational measurement (3<sup>rd</sup> ed.)* (pp. 13–103). New York: Macmillan.
- Mislevy, R. J., & Yin, C. (2012). Evidence-centred design in language testing. In G. Fulcher & F. Davidson (Eds.), *The Routledge Handbook of Language Testing*. Oxford and New York: Routledge.
- Nelson, C. (2011). *Intelligibility in World Englishes: Theory and Application*. New York: Routledge.
- Norton, A. & Cherastidham, I. (2018). *Mapping Australian higher education 2018*. Grattan Institute. Grattan Institute Report No. 2018-11, September 2018, <https://grattan.edu.au/wp-content/uploads/2018/09/907-Mapping-Australian-higher-education-2018.pdf>
- O'Sullivan, B., & Weir, C. J. (2011). Test development and validation. In B. O'Sullivan (Ed.), *Language testing: Theories and practices* (pp. 13–28). Basingstoke: Palgrave Macmillan.
- Paul, L.P., Simons, G.F. and Fenning, C.D. (Eds.) 2020. *Ethnologue: Languages of the World, Eighteenth edition*. Dallas, TX: SIL International. [www.ethnologue.com](http://www.ethnologue.com) [accessed: May 2020].
- Pearson. (2020). *PTE Academic Score Guide for test takers. Version 12 – April 2020*. Palo Alto: Pearson.
- Peters, P. (1995). *The Cambridge Australian English style guide*. Melbourne: Cambridge University Press.
- Quirk, R. (1985). The English language in a global context. In R. Quirk, R. & G. H. Widdowson (Eds.), *English in the world: Teaching and learning the languages and literatures* (pp. 1–6). Cambridge: Cambridge University Press and The British Council.
- Robinson, N. (2018, April 18). Australia hosting unprecedented numbers of international students. *ABC News*. <https://www.abc.net.au/news/2018-04-18/australia-hosting-unprecedented-numbers-international-students/9669030>.
- Seidlhofer, B. (2005). English as a lingua franca. *ELT Journal* 59(4), 339–341. doi.org/10.1093/elt/cci064
- Seidlhofer, B. (2009). Common ground and different realities: World Englishes and English as a lingua franca. *World Englishes*, 28(2), 236–245. doi.org/10.1111/j.1467-971X.2009.01592.x
- Seidlhofer, B. (2011). *Understanding English as a lingua franca*. Oxford: OUP
- Toulmin, S. (2003). *The uses of argument (Updated edition)*. Cambridge: Cambridge University Press.
- Universities UK International. (2018). *International facts and figures: Higher Education 2018*. [https://www.universitiesuk.ac.uk/policy-and-analysis/reports/Documents/International/International%20Facts%20and%20Figures%202018\\_web.pdf](https://www.universitiesuk.ac.uk/policy-and-analysis/reports/Documents/International/International%20Facts%20and%20Figures%202018_web.pdf)
- VOICE. 2013. *The Vienna-Oxford International Corpus of English* (version 2.0 XML). Director: Barbara Seidlhofer; Researchers: Angelika Breiteneder, Theresa Klimpfinger, Stefan Majewski, Ruth Osimk-Teasdale, Marie-Luise Pitzl, Michael Radeka. <https://www.univie.ac.at/voice/> (accessed 26 April 2020).
- Widdowson, H. G. (2013). ELF and EFL: What's the difference? Comments on Michael Swan. *Journal of English as a lingua franca*, 2(1), 187–193. doi: <https://doi-org.ezproxy.jyu.fi/10.1515/jelf-2013-0009>
- Williams, R. O. (1980) Our Dictionaries. In J. Green (Ed.), *Chasing the Sun: Dictionary-Makers and the Dictionaries they Make*, p. 379. London: Pimlico/Random House.
- Xi, X., Higgins, D., Zechner, K., & Williamson, D. M. (2008). Automated scoring of spontaneous speech using SpeechRater<sup>SM</sup> V1.0. *ETS Research Report Series*, 2008(2), i–102. doi.org/10.1002/j.2333-8504.2008.tb02148.x

# ĐÁNH GIÁ KỸ NĂNG SỬ DỤNG TIẾNG ANH NHƯ MỘT NGÔN NGỮ TOÀN CẦU: LẬP LUẬN VỀ TÍNH GIÁ TRỊ

Sheryl Cooke

*Hội đồng Anh, Đại học Jyväskylä, Phần Lan  
989 Đường Tây Bắc Kinh, Thượng Hải, 200041, Trung Quốc*

**Tóm tắt:** Việc sử dụng tiếng Anh trong thực tiễn thường có sự tham gia của những người có nền tảng ngôn ngữ khác nhau với mục tiêu chính là sự hiểu nhau, và phần lớn các cuộc hội thoại bằng tiếng Anh này không có sự tham gia của những người nói tiếng Anh bản ngữ (Graddol, 2006; Kirkpatrick, 2007). Tuy nhiên, thay vì tập trung vào tính dễ hiểu của lời nói, nhiều bài kiểm tra vẫn đo lường khả năng nói của thí sinh với tham chiếu về một khuôn mẫu bản ngữ lý tưởng. Điều này khiến cho tính giá trị của bài thi bị giảm trong việc đánh giá khả năng giao tiếp nói khi tiếng Anh được dùng như một ngôn ngữ toàn cầu, dẫn tới việc bỏ sót kỹ năng cần đánh giá hoặc đánh giá các yếu tố không liên quan.

Việc xác định một bài thi tiếng Anh như một công cụ để giao tiếp toàn cầu bao gồm chứng minh mối liên hệ giữa kỹ năng cần đánh giá (khả năng giao tiếp thực tiễn trong bối cảnh cụ thể) với các tác vụ trong bài thi và tiêu chí đánh giá (McNamara, 2006). Các bằng chứng hỗ trợ việc giải thích ý nghĩa của điểm số cần được trình bày như một phần của lập luận tổng thể về tính giá trị. Trước hết, bài viết này muốn chỉ ra bối cảnh sử dụng tiếng Anh mà nhiều thí sinh trong các kỳ thi lớn hướng tới với mục đích học thuật thường là bối cảnh trong đó tiếng Anh được dùng như một ngôn ngữ toàn cầu (English as a lingua franca – ELF). Tiếp theo, lập luận của Toulmin (2003) được tận dụng để tìm ra những bằng chứng cần thiết cho việc chứng minh các khẳng định về một bài kiểm tra có tính đại diện tốt cho khả năng sử dụng tiếng Anh như ngôn ngữ toàn cầu. Mô hình được đưa ra và áp dụng vào phân tích minh họa hai bài thi Nói tiếng Anh. Bài viết có mục đích đưa ra minh chứng về mặt thuyết cho sự cần thiết của việc tập trung vào tính dễ hiểu trong giao tiếp và việc tăng cường sự đa dạng ngôn ngữ trong đánh giá ELF. Ngoài ra, tác giả cũng mong muốn đưa ra một mô hình xác định có tính ứng dụng đối với người soạn và sử dụng bài thi.

*Từ khóa:* Tiếng Anh như một ngôn ngữ toàn cầu, tính giá trị của bài thi, tính dễ hiểu