Chapter 2: A step toward the assessment of English as a lingua franca

Gary J. Ockey and R. Roz Hirch

Iowa State University

Abstract

This conceptual paper describes the need for L2 English tests to consider English as a lingua franca (ELF) in their development. After discussing what is meant by ELF, it describes an oral communication placement test developed at a large Midwestern university in the United States. The test is then analyzed based on a framework designed to determine the extent to which a test can be considered to have adhered to ELF principles. It is argued that for the most part the test does appear to assess ELF. However, it does not completely adhere to ELF principles, and sometimes, because there is little agreement on ELF principles, it is difficult to determine the extent to which the test actually does assess ELF. It is recommended that researchers in the field come to agreement on what can be considered critical aspects of ELF, which will make it possible for language assessment researchers to better design their assessments to include ELF.

Introduction

Changing demographics and globalization, among other factors, have led to the increasing use of English as a lingua franca (ELF), which we broadly define as a communication context in which at least one user in an English-mediated communication has a different first language (L1) than other users (Canagarajah, 2006; Jenkins, 2009; Mortensen, 2013; Seidlhofer, 2011). Based on this definition, it is becoming more common to find ELF than non-ELF contexts. It follows that oral assessments designed to determine the degree to which a test taker has the necessary oral

proficiency to function in a given context should be designed to assess ELF. Because the ELF context has become so common, there has been growing criticism directed at the language assessment community for not targeting ELF when designing L2 English assessments (e.g., Jenkins, 2016; Canagarajah, 2006; Leung & Lewkowitz, 2017; McNamara, 2018). These criticisms may not be completely accurate, however, since some L2 assessments already take into account many ELF concerns. In fact, a move within the language testing community to consider the ideas underlying ELF can be traced at least back to when Bachman argued that L2 assessments should assess communicative language ability (Bachman & Clark, 1987; Bachman, 1990). The assessment of communicative language ability removed the focus from assessing a rule-governed linguistic system devoid of context to using the language successfully within a given context and for a particular purpose. While neither Bachman nor others in the language assessment field discussed the concept of ELF at the time, this move toward communicative language testing was a critical step toward future ELF testing, as many language assessments have increasingly morphed into assessments that integrate aspects of ELF, whether intentionally or not. An example of a test with many ELF features is in current use at Iowa State University (ISU). ISU's oral communication placement test was designed with ELF as a guiding principle; however, this proved to be a challenge, in part because there is little agreement on what ELF is and almost no practical guidance on how ELF could be used to inform L2 test design. This chapter was inspired by the challenges of designing an ELF university placement test. It begins with a discussion of how ELF has been defined. Next, it describes the design of ISU's oral communication placement test. This is followed by a discussion of the extent to which ISU's oral communication placement test can be considered an ELF test. The paper concludes with

implications and future directions for ELF and assessment researchers and developers who aim to develop ELF assessments.

Constructs and Historical Perspective

English as a Lingua Franca

English as a lingua franca (ELF) is, in many ways, related to two other concepts in English as a second language: World Englishes (WE) and English as an International Language (EIL). The three are frequently mentioned together and share many attributes; indeed, ELF is sometimes seen as growing out of WE. All three reference Kachru's (1992) concentric circles to varying degrees. However, there are distinctions. In WE, communication occurs based on local conventions, so the focus is on shared linguistic components (Canagarajah, 2006; Leung, Lewkowicz, & Jenkins, 2016). EIL is also concerned about local use of English, but rather than being about linguistic features, it focuses on the needs of users (Brown, 2014). ELF, on the other hand, focuses on negotiating communication between speakers of different L1s. This section will explore the different phases of ELF which has led to its current definition, as well as setting out the definition for ELF that will be used in the rest of this chapter. This will be followed by ELF criticism of language testing and language testers' responses to these criticisms.

What is English as a Lingua Franca?

Jenkins (2015) identifies 3 phrases of the evolving definition of ELF. The first began in the 1980s with her own research in pronunciation and mutual intelligibility. Other researchers subsequently focused on additional features such as grammar and lexis and identifying features shared among English speakers from different L1s. Phase 1 closely resembled WE; the break would come in the 2000s. In phase 2, researchers relied less on Kachru's circles, as the circles

were criticized for having too great a focus on native identity (Jenkins, 2015). More significantly, researchers dispensed with lists of commonalities, paying greater attention to the ways in which English speakers of different L1s negotiate meaning (Seidlhofer, 2006). Jenkins (2015) also proposes a third phase, which she dubs "English as a multilingual franca"; this phase has many features of the second phase, except that English is given less precedence. In phase 3, English is one of many languages that speakers have at their disposal, and speakers may draw on those other languages to communicate, which essentially becomes a form of translanguaging. While this is an interesting development, we feel the third phase focuses more on communication generally (Guzman-Orth, Lopez, & Telentino, 2019), and would therefore require a test of communication more than a test of English language. For that reason, our definition is drawn from Jenkin's phase 2.

There are two essential points that definitions of ELF in phase 2 have in common: 1) at least one speaker in a communication situation has a different L1 from other speakers, which may be English or another language (e.g. three Chinese L1 speakers and one English L1 speaker); and 2) the speakers choose to use English in the communication context (Canagarajah, 2006; Jenkins, 2009; Mortensen, 2013; Seidlhofer, 2011). Identity is also an important concept. ELF is not a natural language in that no one identifies as an ELF speaker; instead, speakers identify as being a member of their L1 communities. Since native English speakers (NESs) therefore cannot be native ELF speakers, "native English" should not take precedence nor serve as a model for ELF (Davies, 2009; Elder & Davies, 2006; McNamara, 2011).

Criticisms of Tests and of ELF

The use of native English speakers (NESs) as a standard for testing is one of the main criticisms ELF researchers make of current practices in language testing. Using a NES standard

assumes that, when a test taker is in real-world English language situations, the person with whom they speak will also be a native English speaker. In fact, given the number of nonnative English speakers (NNESs) in the world today, we know that this will likely not be the most frequent situation, even within an English-speaking country (Brown, 2014; Jenkins & Leung, 2017; McNamara, 2011; Newbold, 2015; Ockey & Wagner, 2018). The reliance on the NES standard also suggests that the NNES is in some way deficient or even incapable of achieving the level of an NES (Leung et al., 2016; McNamara, 2011). Another criticism of current language tests expressed by ELF researchers is their focus, which is usually on formal aspects such as lexis and grammar. Instead, tests should focus less on the rules of communication and more on whether speakers are able to mutually understand each other (Brown, 2014; McNamara, 2011). Similarly, ELF practitioners would like to see more emphasis placed on pronunciation and dialects in the testing situation (Jenkins & Leung, 2017; Leung et al., 2016; McNamara, 2011).

While most assessment researchers agree that these are valid criticisms, they also have concerns about ELF and its application to tests. For language testers, one of the greatest complications of incorporating the ideas supported by ELF researchers is that it is difficult to pin down a practical definition for assessment purposes (Elder & Davies, 2006). Often, ELF ideas are focused on political concepts rather than linguistic ones, so assessment practitioners, who rely on construct definitions, can find it difficult to fit ELF principles into a practical testing situation (Davies, 2009; Newbold, 2015). Furthermore, there may be consequences for psychometrics, which generally rely on a stable definition of the construct being tested; considering that ELF is highly context oriented and describes shifting communication situations, many of the psychometric bases of large-scale, standardized tests may need to be reconsidered (Jenkins & Leung, 2017). It is also important to take into account the effects on test takers, many

of whom may not be familiar with ELF, or who may for many reasons be unprepared for diverse dialects (Brown, 2014). Finally, as Brown (2014) points out, testing is much more than standardized tests such as TOEFL or IELTS, which are largely the target of ELF researcher criticisms – indeed, the focus of this chapter is on a test type that is commonly administered at universities throughout the world: placement tests. The next section briefly describes the form and content of such a test that is currently in use at Iowa State University and is representative of a type seen at numerous universities. This description is expanded in the third section, which examines the test in relation to an ELF framework to suggest that this placement test may function as a form of ELF assessment.

The English Placement Test of Oral Communication (EPT OC)

General description. The English Placement Test of Oral Communication (EPT OC) is a face-to-face assessment of oral communication ability that aims to determine the extent to which test takers have the academic oral communication skills necessary to be successful at the university. Students who do not obtain a passing score on the test are placed into appropriate English language courses for their level, which are designed to give them the requisite skills to communicate effectively in the university.

Test design and development. The course and test were designed based on a needs analysis of second language English users at the university and the English required for them to be successful in their content courses and in navigating their way through the university. Content and ESL instructors as well as international students provided feedback on the students' English needs at the university as part of the needs analysis. The course and the test were developed simultaneously and shared the same constructs/objectives. The test tasks were selected from the range of activities completed in the course.

Test construct. Aligned with the course objectives, the EPT OC aims to assess the following: Interactional Competence, the ability to respond appropriately in various contexts – for example, to effectively interact with professors, instructors, peers, and other university staff; Fluency, the ability to use the language fluently; Comprehensibility, the ability to produce language comprehensible to English users at the university; Vocabulary/Grammar, the ability to use academic vocabulary and grammar effectively and appropriately (Ockey & Li, 2015).

Test administrators/raters. The test is administered and rated by graduate students and instructors, who teach the Oral Communication courses for test takers who do not pass the EPT OC, as well as other English courses offered at the university. All raters have advanced English language abilities, have taught English language courses, and are working on or possess advanced degrees in the field of applied linguistics. They come from a variety of backgrounds. For example, during the fall 2018 administration, raters had 11 different L1s, were from 11 different countries, and represented diverse cultures as well as subcultures and dialects within these cultures and L1s.

Test takers. The test takers are students who have been accepted to the university but may need further English language support courses as they begin their undergraduate or graduate journey. To limit the resources needed for the test, students who have demonstrated success in an English-medium academic context (e.g., have a bachelor's degree, master's degree, or high school diploma from an English-medium school) or have sufficient scores on a standardized English test (e.g., SAT, ACT, GRE, TOEFL iBT, IELTS, PTE Academic) are exempted from taking the test and the oral communication class. Test takers can choose to not take the test if they prefer to take the oral communication class. If they desire to be exempted from the class and do not satisfy any of the other exemption criteria (Iowa State University English placement test

(ISU EPT): https://apling.engl.iastate.edu/english-placement-test), they can choose to take the test. This results in a test taker population from many parts of the world, who received high school or university degrees in non-English medium universities and have TOEFL iBT scores between 71 and 99 (or their IELTS or PTE Academic equivalent); seventy-one is the minimum TOEFL iBT score for university acceptance, and exemption from taking the speaking class is given to students who have TOEFL iBT scores of 100 or above.

Procedures and test tasks. The EPT OC is video recorded, takes approximately 20 minutes, and is composed of three tasks: a scripted one-on-one interview with a test administrator, a retell to two test administrators and a peer test taker, and a paired discussion with the peer test taker. After brief introductions with a test administrator, the scripted interview task is administered. For this task, test takers are asked three questions. First, they are shown a picture and asked to describe it; second, they are asked to respond to a hypothetical question associated with the situation in the picture; and third, they are asked to talk about what they would do in the hypothetical situation posed in the second question. Students are asked to provide 60-second responses to each of the three questions.

For the retell task, test takers listen to a 30-second recording supporting one side of a two-sided issue. The speakers on the recordings are judged to have strengths of accents based on Ockey and French's (2016) Strength of Accent scale, which was designed to make it possible for language assessments to include diverse speech varieties, regardless of the speaker's L1, which would not unfairly influence a test taker's score on a test. Speakers' strengths of accent on the EPT OC are of two types: 1) "The speaker's accent was NOT noticeably different than what I am used to and did NOT require me to concentrate on listening any more than usual; the accent did NOT decrease my understanding." Or 2), "The speaker's accent was noticeably different than

what I am used to but did NOT require me to concentrate on listening any more than usual; the accent did NOT decrease my understanding." A speaker's strength of accent is judged by highly proficient L1 and L2 listeners familiar with the local speech variety. After listening to the position, one test taker is asked to retell the speaker's view and reasons for this view in their own words. Students then listen to another 30-second recording from a speaker who expresses the opposing view on the issue and the other test taker is asked to retell that speaker's view.

The third task is a paired discussion task, in which the two test takers are asked to discuss and defend the positions they summarized in the retell task. Test takers are given four minutes for their discussion. After the discussion begins, the test administrators listen quietly away from the two test takers engaged in the discussion. Further details about the test and example test tasks can be found at the ISU EPT website: https://apling.engl.iastate.edu/english-placement-test

The content of the items is based on topics that test takers might encounter at the university. Examples of content for the one-one-one oral interview include explaining how one might deal with: losing a library book; plagiarizing a course paper; cheating on an exam; and working on a group project with an uncooperative group. The prompts encourage test takers to talk about these issues based on their own cultural experiences. Examples of items for the retell and paired discussion include taking a position on topics such as: the value of group work; the importance of a part-time job while studying; and the usefulness of online classes.

Evaluation criteria. Test takers' performances are evaluated independently by the two trained raters who administer the test. Ratings are based on a 4-point analytic scale, with subscales of pronunciation, interactional competence, fluency, and grammar/vocabulary (See Appendix A for Oral EPT rubric). Test takers are assigned two sets of scores by the test

administrator with whom they talked during the one-on-one oral interview, and one by the test administrator who listened to them during the retell and paired discussion tasks.

Raters have the option of requesting a test taker be given another opportunity to take the test if it is felt that the test was in some way unfair (e.g., a test taker's partner dominates the paired discussion task), or that another trained rater evaluate the video-recording of the test, if the rater feels biased against assigning a fair rating. Scores are adjusted for rater severity and prompt difficulty with Many-facet Rasch Measurement techniques (Eckes, 2015; McNamara, Knoch, & Fan, 2019). Test takers assigned scores of 3 or higher on the rating scales, after taking into account the standard error of measure of the test, are assigned passing scores. Typically, this means a score of roughly 2.75. This is a weighted score based on doubling the importance of scores for interactional competence and comprehensibility as compared to fluency and grammar/vocabulary.

Critical Issues

A framework for designing or analyzing a test based on ELF principles does not exist, but Brown (2014) developed a "criteria for locally defined EIL curriculum development that could equally well apply to testing locally defined EIL" (p. 10). This framework was originally developed for assessment of EIL, which is different from but related to ELF, as described in the introduction above; Brown's framework therefore served as a useful starting point to build an ELF framework because there is a great deal of overlap between the two. It should be noted that Canagarajah (2006) also contains a framework for assessing EIL and influenced the development of the framework presented in this chapter, but it focused on classroom assessment, whereas Brown's framework was about non-classroom tests such as the placement test we evaluate in this chapter. Changes to Brown's framework were made based on differences between EIL and ELF,

Canagarajah's (2006) framework, current research on ELF, and ELF researchers' criticisms of English assessments (Table 1). Additionally, the framework is divided into two parts, the first being attributes of the test taker that an ELF test should measure (5 points), and the second being qualities that the test should have (2 points). The aim of this 7-point framework was to provide guidelines that could be used to determine the degree to which an assessment could be considered an ELF test. While there may not be any tests perfectly designed to assess ELF, we contend that some tests do embody many aspects of ELF. To this end, the EPT OC test, described above, was judged according to each of the criteria to see to what extent it can be considered an ELF test. Each point in the framework below begins with a criterion based on ELF principles, followed by a more detailed definition of the term and its relation to ELF. Each point ends with an evaluation of the EPT OC according to that criterion.

(Place Table 2.1 about here)

English rhetorical sensitivity. The first of the test taker traits listed, rhetorical sensitivity is required in all language situations (Bachman & Palmer, 1996); learners should always be aware of the genre they are in and respond to the situation appropriately, whether speaking or writing (Canagarajah, 2014). ELF researchers point out that, while native English rhetorical structures are generally well taught, rhetorical structures from other cultures are not always as widely known (Elder & Davies, 2006; Jenkins, 2011). This also means that people from two different L1s may be accustomed to different approaches to rhetoric as well as having knowledge of native English rhetoric. The difference between ELF and other testing situations is that ELF speakers should be able to draw on non-English rhetoric as well as their knowledge of English rhetoric – which may also differ between them – as part of their communication. This description is abstract; a practical application might be a situation such as an argument. English has its own

rules for structuring arguments that differ from those of Chinese; learners should be able to draw from both if they choose.

Part of rhetorical sensitivity, as it is presented in ELF literature, are pragmatics and sociolinguistics, which ELF scholars tend to closely associate together. This connection is clear in Canagarajah (2006) when he writes "We have to focus more on proficiency in pragmatics. Sociolinguistic skills of dialect differentiation, code switching, style shifting, interpersonal communication, conversation management, and discourse strategies are important for shuttling between English varieties" (p. 233). An ELF view of pragmatics is one that takes into account cultural differences and the need to negotiate these in the course of any given interaction (Newbold, 2015). One of the interesting consequences of this view of pragmatics is that it has less focus on idiomaticity, since idioms would likely be more of an impediment to communication than most grammar errors (Cogo & Dewey, 2011; Hall, 2014; Kim & Billington, 2014; Newbold, 2015; Prodromou, 2010). Thus, an ELF test would highlight sensitivity to differences in dialect, culture, and identity, while downplaying the use of traditional native-like fluency indicators such as idiom use.

The EPT OC paired discussion task, in which two students, usually with different cultural backgrounds and L1s, discuss a topic, shows this type of rhetorical sensitivity. Students are able to structure their arguments as they like and often need to negotiate meaning based on the structure of each other's arguments because their rhetorical styles may differ. Rhetorical style is not indicated as part of the construct; instead, responding appropriately and effectively negotiating meaning are included in the test construct as part of the Interactional Competence subscale. At the same time, the raters also have various cultural backgrounds and L1s and make judgements of the test takers' language abilities while having their own culturally-influenced

approaches to rhetoric. Raters are furthermore instructed not to judge according to an NES standard or expect or require test takers to use idiomatic expressions. Instead, the focus is on how well test takers adapt to the conversation with another L1 speaker. Given these features of the test, it is likely that the EPT OC assesses English rhetorical sensitivity in line with ELF standards.

International communicative competence. An ELF view of communicative competence in speaking can be broken down into 3 aspects: intelligibility, accommodation to variety, and correction. Intelligibility is what Jenkins (2015) began with in 1980. In ELF research, intelligibility tends to be summarized as how well a person is understood, and is therefore related to the act of speaking (Chopin, 2015; Elder & Davies, 2006; Isaacs & Trofimovich, 2012). Accommodation to variety is the flip side of intelligibility in that it is how well a person can understand other speakers, particularly those with unfamiliar dialects, and thus relates to listening (Canagarajah, 2014; Jenkins & Leung, 2017; Ockey & Wagner, 2018). Correction refers to repairs, but these are not grammatical; they are, instead, communication repairs and are similar to Canale and Swain's (1980) strategic competence (Canagarajah, 2014; Chopin, 2015; Newbold, 2015). Thus, an ELF definition of communicative competence in a speaking context can be viewed as how well the test taker can be understood by the other speaker (intelligibility), how well the test taker can understand a variety of speech types (accommodation), and in cases where one of those breaks down, how well the test taker can make corrections.

Communicative competence is a substantial component of the EPT OC rating scale in all three categories:

Intelligibility. The construct of the test includes the subscale, Comprehensibility, which refers to the degree to which the speech variety of the test taker requires effort to understand, as judged by the raters. The rating scales do not make reference to native-like speech.

Accommodation: How well a person can understand another speaker is part of the EPT OC construct. It manifests itself in the Interactional Competence rating subscale, which has been shown to require "Active Listening" (Ducasse & Brown, 2009; Galaczi, 2014; May, 2011) since an appropriate response is often based on understanding another speaker. In fact, raters are trained to judge the degree to which it is necessary to accommodate to the needs of a partner. For example, successful accommodating may require slowing one's speech or using more simple vocabulary or sentence structure to communicate with a partner with a different speech variety or proficiency level.

Correction. The ability to correct miscommunication is also part of the EPT OC construct and is indicated in the interactional competence rating subscale. Because test takers speak with both a highly proficient test administrator and a peer who could have rather limited English proficiency, both of whom are likely to speak different English speech varieties, communication break downs are common. The ability to correct these miscommunications as judged by the rater as part of the test construct.

Context sensitivity. Within test use, ELF is concerned with context: where the test is being delivered, who is delivering it, and who is taking it, among other concerns (Brown, 2014; Canagarajah, 2006; Leung et al., 2016). Each context may have its own norms and expectations; some contexts may be more flexible and allow for more negotiation than others. In contexts that allow for more flexibility, too strict an adherence to an NES standard could lead to bias from an ELF perspective (Canagarajah, 2014). For example, delivering a business proposal to a company

in an English-speaking country may require stricter adherence to NES standards than a conversation with classmates in a university in an English-speaking country would. In the former case, a large-scale standardized test with formal linguistic features may be appropriate (Chopin, 2015), while the latter situation may be better served by a test that contains more ELF elements (Jenkins & Leung, 2017).

However, most ELF researchers would suggest that ELF is necessary in all assessment situations to varying degrees. Brown (2014) laid out 8 language constituencies to consider: the local community where the test will be administered; test takers, which may be from the local community or elsewhere; test content; test proctors; test raters; the community that will be affected by the decision (which may be different from the local community); the purpose of the decision being made with the test; and the people who are making the decisions. When any of these Englishes is different – which would presumably happen in almost all language assessment situations – then there is a potential for unfairness (Brown, 2014). Looking at these 8 different constituents might be a useful tool to assist test developers with deciding on the degree to which ELF should be incorporated into a test.

The design of the EPT OC includes placing test takers into different contexts and judging their abilities to navigate these different contexts. The one-on-one individual interview with an examiner is meant to be a formal situation, and test takers are expected to recognize this formality and respond appropriately to it. The paired discussion with a peer test taker is meant to be a somewhat informal situation. In this context, test takers are expected to use less formal language. In neither context is the test taker expected to conform to an NES norm. Rating scales refer to concepts such as effectiveness, appropriateness, and comprehensibility. Using these two

tasks makes it possible for test takers to use more standard English, as they would with an instructor, or less standard English, as they would with a group of classmates.

The EPT OC takes into account all eight of the constituents suggested by Brown (2014). Most test takers are from outside of the local community, while the raters/test administrators are mostly from outside the local community but familiar with the local speech variety and culture. They can judge a test taker's oral communication ability based on both an insider and an outsider perspective. This is important since test takers will need to be able to communicate with both of these groups of English users in the university setting. The test content is based on topics and genres commonly encountered at the university that afford students opportunities to talk about their values and cultures.

Motivation. Brown (2014) highlighted 2 aspects of use: the ways that test takers use English, and the way that the test is used. In ELF, considerable attention is given to characteristics of the speakers, especially their first languages and the fact that they are choosing to communicate in English. However, less attention is paid to why they are speaking English – what is the speakers' motivation? And in a testing situation, what are the motivations of the stakeholders? Motivation is important in ELF because it can affect the degree of negotiation that may be needed or expected of the speakers. Motivations can take many forms, either locally, such as for communication with residents who speak a language or for advancement in a company or similar purposes, or globally, such as for business, travel, immigration, or even things like online gaming (Brown, 2014). Another related issue is consideration for the different experiences and the meaning that learning English has for the test takers, which could vary considerably depending on their L1 and their country of origin (McNamara, 2011). Learners are affected by not only the differences or similarities of their L1 to English, but also by the varying

experiences their countries have had in contact with English language-speaking countries, and the cultural beliefs and customs of their countries as well (McNamara, 2011). The personal and cultural experiences of the test taker will likely affect their motivation for learning English.

From an assessment researcher's perspective, "motivation" is closely tied to the concept of assessment validity; is the test valid for the use being made of it? In the case of the EPT OC, the immediate motivation for test takers is to be able to take content classes with English as the language of instruction; hence, the test should assess whether learners are able to communicate in an English-medium classroom. A needs assessment was conducted in the development phase that included discussions with various stakeholders regarding classroom communication needs. Additionally, the test is aligned with an English course to ensure that students who do not meet the minimal requirements will still have the opportunity to achieve their learning goals. It should also be noted, however, that learners may have motivations beyond taking content courses in English that are still tied to their success in university. For example, students may want to graduate from an English-medium university to get a better job in their home country or to emigrate to an English-speaking country. The EPT OC cannot be used as an assessment for those purposes, nor would anyone suggest it should; it can only assess learners for the immediate goal of communication within the university and the surrounding community. Large-scale tests are mostly blind to any of the test taker's motivations, and it is a difficult component to understand – in fact, knowing a test taker's motivation may mean discouraging them from taking a test on ethical grounds if the use of the test is not valid.

Grammatical appropriacy. Grammar is frequently a sub-construct to be measured in most assessment situations; the rubrics for TOEFL's, iELTS's, and the CEFR's speaking tests each have a descriptor band that includes grammatical accuracy, and both iELTS and CEFR

mention native speakers at least once in theirs. Because these rubrics are frequently adapted for use in other testing situations such as university placement, test takers' performances are likely to be compared to some standard of native-like grammar (Chopin, 2015). Despite the reliance assessments often have on NES grammar, ELF researchers seem hesitant to do away with grammar entirely, since some form of grammatical knowledge is needed to communicate (Canagarajah, 2014). ELF researchers suggest that grammar should be viewed as something "emergent, not preconstructed. As [speakers] collaborate with each other in attaining their communicative objectives, they construct certain norms that make their interaction possible" (Canagarajah, 2014, p. 770). There are clearly several problems with this description for testing purposes, the first being that this definition is too fluid for defining a construct (Davies, 2009; Newbold, 2015). Furthermore, this idea of grammar as something that may be constructed over time is more appropriate as a long-term classroom goal than an aspect of assessment, which has been observed in other studies (Newbold, 2015; Prodromou, 2010). Canagarajah (2018) suggests that assessors should reject structuralist definitions of grammar by acknowledging that a test taker's grammar need not be perfect by NES standards, but should be sufficient to achieve mutual comprehension. Adopting Canagarajah's (2018) view of grammar, rubrics would require two changes in the grammar category: 1) grammar would play a less significant role in scoring, and may indeed be combined with another category; and 2) "grammatical errors" would be defined in the context of the conversation as instances where grammar interfered with comprehension, not as deviations from prescriptivist rules.

The EPT OC includes in its construct grammatical accuracy, but accuracy is not defined based on an NES norm; instead, it is based on the judgement of the diverse population of raters. During rater training sessions, examples of speakers who do not have accurate grammar based on

an NES norm but are nonetheless good at communicating are provided as examples of speakers that should be assigned passing scores. Raters are trained to focus on the degree to which what the test taker says is understandable. The use of grammatical structures that lead to breakdowns in communication are judged negatively, while the use of comprehensible grammar forms that do not conform to NES standards are judged much less negatively.

Nevertheless, just as ELF researchers do not have a shared view of how to approach grammar, it is not clear what exactly the role of grammar is and should be for the EPT OC. For example, in the picture description task, a test taker might attempt to describe a picture, which includes two bottles on a table. If the test taker said, "Two bottles are on the table," the sentence would be acceptable from both a grammatical accuracy point of view and an ELF point of view. However, if the test taker said, "Two bottle on table," the discourse would be considered inaccurate from a grammatical accuracy point of view, but probably appropriate from an ELF perspective. The EPT OC does include grammatical accuracy in scoring because instructors of both ESL and content courses base grades on it; excluding it would not align with the real-world needs of the students. However, to also align with ELF standards, the focus is on communication, rather than accuracy. Thus, following ELF standards for rating grammar would result in a rather high overall score (probably passing), but would not result in the maximum point value.

It should be noted that despite this training, it is apparent from observations of and conversations with raters that they do pay attention to NES norms when evaluating grammar structures during the assessment. However, it is also clear that they value comprehensible grammatical forms over incomprehensible ones. In short, raters seem to struggle with how to evaluate grammar beyond its comprehensibility and accuracy when compared to an NES norm.

One further element that should be pointed out is that, although grammar is given its own category (along with vocabulary), the EPT has been designed so that grammar would not be given as much weight as other categories; therefore, it is given half the weighting of interactional competence and comprehensibility in the final EPT OC score. Thus, from an ELF perspective, grammar should not play an overly important role in assessing the test takers.

Relevance. As described above, oral communication tests that involve an element requiring negotiation best fit ELF; this section and the next look at appropriate characteristics of tasks and tests as a whole. While it might be possible to adapt currently-administered tests, this may be insufficient; several aspects need to be considered that may change the structure of a test (Elder & Davies, 2006). Testers need to consider the context of the test to identify appropriate types of Englishes (Brown, 2014). For example, in a university placement test, consideration should be given both to the variety of dialects likely to be encountered in the local community as well as in the university. Local dialects in a rural area will probably be more homogeneous than those in a major metropolis. The test should also have a variety of tasks that are performance-based and interactive, allowing for social negotiation and pragmatic competence relevant to the testing context; discrete items are therefore considered less likely to be useful for ELF assessment (Canagarajah, 2006; Chopin, 2015; Elder & Davies, 2006; Jenkins & Leung, 2017; Newbold, 2015). At the same time, receptive skills should be considered; in the case of speaking, listening will undoubtedly be a component of the task, so different dialects may need to be built into the test (Brown, 2014; Ockey & Wagner, 2018). Finally, the materials used to elicit test taker responses should be relevant to the assessment context (Elder & Davies, 2006; Newbold, 2015). In a university placement situation, this could mean including topics that are not only relevant to students as a whole but also to students at that particular university.

The EPT OC was designed to assess the ability to negotiate meaning and pragmatic competence; the paired discussion task was created with precisely this purpose in mind. Receptive skills are also important on the EPT OC, in all three tasks. Test takers must comprehend other speakers and a recorded input (in the Retell task) to be successful. Finally, the materials for the test were created specifically for the targeted test takers. The prompts were all created (and continue to be created) based on a needs analysis of the types of communication commonly encountered at the university. Furthermore, the topics (as described above) are drawn from situations that students may experience in a university situation or may be related to current and relevant circumstances at the university.

Fairness. One of the subtler ways in which NNESs are judged against NESs is in the use of item writers and raters who are native English speakers. Incorporating NNESs in assessments is an issue of fairness, but also the reality of the world today; there are more NNESs in the world than there are NESs, so it is simply practical to include a variety of voices (Davies, 2009; Leung et al., 2016). In the case of item writing, there are many aspects that can be affected, both in the topics that are used and the language of the text (Elder & Davies, 2006). It is not enough to have different accents; the words and phrasing should be authentic to the accent, which requires having writers from the dialects being tested. Similarly, raters should be advanced English speakers from a variety of L1s (including NESs, and also potentially different from the test takers' L1) because ELF requires that a speaker be intelligible, and not only to NESs (Jenkins & Leung, 2017; Leung et al., 2016).

ELF researchers recommend extensive training on non-standard forms, particularly of the rubric elements described above, to break habits of attempting to conform to native standards (Canagarajah, 2006; Davies, 2009; Elder & Davies, 2006). Of course, since studies have found

that there is little difference between NES and NNES ratings, it may not be necessary to have

NNESs (Brown 2014). A counterargument to this is that other studies have found that there may

be qualitative differences – for example, Zhang & Elder (2011) found that, although speech

ratings were similar, NESs focused more on the descriptors related to content and ideas, while

NNESs focused more on appropriateness and completeness. Thus, while the scores may be

similar, the meaning of them may not be; it is possible that the effect of these differences could

be more pronounced in an ELF test.

At present, most language tests are required to be taken only by NNESs; the exceptions

are some tests for academic English, such as those that place students into academic writing

classes. One of the recommendations for an ELF test is that NESs should also be required to take

the test (Canagarajah, 2014; Chopin, 2015; Jenkins & Leung, 2017; Leung et al., 2016;

McNamara, 2011; Newbold, 2015). Proponents argue that this is an issue of social justice; when

non-native speakers have to take a test that native speakers do not have to take, it can make

NNESs feel that they are somehow different or even deficient (Chopin, 2015; McNamara, 2011).

The EPT OC includes NNESs and NESs at every stage of test development,

administration, and scoring. Prompts are created by groups of item writers and then go through

numerous stages of feedback about the content and the language. The aim is to create prompts

that will be of interest and relevance to the test takers with language that is considered

appropriate by advanced speakers from various L1 and L2 English speech varieties. After scripts

have been written, speakers for the listening input are encouraged to revise what they will say in

accordance with their speech variety. Test takers encounter multiple speech varieties on the test,

including the oral interviewer, who is an advanced English speaker, two different recorded

speakers for the retell task, one male and one female who are both advanced English speakers

with different speech varieties, and a peer with a lower level of English proficiency from an L2 speech variety. Given that much of this language is not scripted, a test taker will encounter a variety of Englishes. Raters are advanced English speakers from a variety of L1s; it is highly unlikely that a test taker would be assessed by two raters with the test taker's own speech variety. Moreover, it is highly likely that at least one of the raters and recorded speakers will be an L2 user of English. Raters go through extensive online and face-to-face training, which underscores the need to rate based on effectiveness, clarity, and comprehensibility rather than targeting a native-like norm.

Requirements for taking the oral communication class or passing EPT OC are based on the lack of demonstrating sufficient English proficiency to be likely to be successful at the university. Students who have graduated from an English-medium high school or university are exempted regardless of their L1 or country of origin. Sufficient scores on various English assessments, both ones commonly taken by L1 and L2 English speakers, are also grounds for exemption. Likewise, being a citizen of the country where the university is located is not grounds for exemption. The exemption criteria are designed to reduce the resources needed for testing (by not requiring students to take the test who are very likely to pass it anyway) but ensure reasonable equity in who is exempted from taking it.

Conclusions, Implications, and Future Directions

Given the widespread use of ELF, it is pretty clear that most English tests should be guided at least to some degree by ELF principles. Criticisms by ELF researchers that English tests are not considering ELF issues are no doubt to some extent legitimate. On the other hand, ELF researchers should recognize that many L2 English tests have developed in ways that suggest ELF principles are, whether intentional or not, part of their design. Many of the same forces,

such as globalization, increased use of English among non-native speakers, and communicative language teaching that have influenced ELF have also affected language test design.

This paper suggests that ELF tests can not only exist but that to one degree or another already do. For example, the ISU oral communication English placement test has been shown to have many ELF features. However, this paper also makes clear that assessment of ELF may conflict with some language assessment principles. This paper further indicates the need for a clear framework that could be used to determine the extent to which a test can be considered an ELF test and even more importantly, a framework that can guide language assessment test developers who aim to create ELF tests. The adaptation of Brown's (2014) EIL framework used in this paper was at best a rudimentary start to something that could be used for this purpose. Researchers will, no doubt, point out many limitations to this framework, which should help to move research in this area forward. However, before a defensible framework can be fully developed, researchers need to come to a shared agreement about what the critical aspects of ELF are. This will make it possible to better target these aspects in a useful ELF test development framework.

Efforts to adopt a clear definition of ELF that could help guide language assessment developers is an important future direction for both ELF and language assessment researchers. Such agreement could lead to the development of an ELF framework that could be used to guide ELF test development and raise awareness of the degree to which English tests in current use already follow many ELF principles. Once this is sorted out, it may be discovered that ELF tests, to one degree or another, already exist. While it is unlikely that Bachman foresaw the future trend toward ELF assessment when he was writing about communicative competence in the

1980s and 1990s (e.g., Bachman, 1990; Bachman & Clark, 1987), it is likely that his vision of

communicative language assessments aimed the field toward the assessment of ELF.

References

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford UP.

Bachman, L. F., & Clark, J. L. (1987). The measurement of foreign/second language proficiency. *The ANNALS of the American Academy of Political and Social Science*, *490*(1), 20-33.

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

Brown, J.D. (2014). The future of world Englishes in language testing. *Language Assessment Quarterly*, 11(1), 5-26.

Canagarajah, S. (2006). Changing communicative needs, revised assessment objectives: Testing English as an international language. *Language Assessment Quarterly*, *3*(3), 229-242.

Canagarajah, S. (2014). In search of a new paradigm for teaching English as an international language. *TESOL Journal*. *5*(4), 767-785.

Canagarajah, S. (2018). The unit and focus of analysis in lingua franca English interactions: In search of a method. *International Journal of Bilingual Education and Bilingualism*, *21*(4), 1-20.

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

Chopin, K. (2015). Reconceptualizing norms for language testing: Assessing English language proficiency from within an ELF framework. In Y. Bayyurt & S, Akcan (Ed.). *Current*

perspectives on pedagogy for English as a Lingua Franca. (pp. 193-204). Berlin: De Gruyter Mouton.

Cogo, A., & Dewey, M. (2006). Efficiency in ELF communication: From pragmatic motives to lexico-grammatical innovation. *Nordic Journal of English Studies*, *5*(2), 59-93.

Davies, A. (2009). Assessing world Englishes. *Annual Review of Applied Linguistics*, *29*, 80-89.

Ducasse, A. M., & Brown, A. (2009). Assessing paired orals: Raters' orientation to interaction. *Language testing*, *26*(3), 423-443.

Eckes, T. (2015). *Introduction to Many-Facet Rasch Measurement. Analyzing and evaluating rater-mediated assessments*. 2nd Revised and Updated Edition. Series: Language Testing and Assessment. Volume 22. Frankfurt am Main: Peter Lang.

Elder, C., & Davies, A. (2006). Assessing English as a lingua franca. *Annual review of applied linguistics*, *26*, 282-304.

Galaczi, E. (2014). Interactional competence across proficiency levels: How do learners manage interaction in paired speaking tests? *Applied Linguistics*, *35*(5), 553–574.

Guzman-Orth, D., Lopez, A., Tolention, F. (2019). Exploring the use of dual language assessment tasks to assess young English learners. *Language Assessment Quarterly*, *16*(4-5), 447-463.

Hall, C. J. (2014). Moving beyond accuracy: from tests of English to tests of 'Englishing'. *ELT Journal*, *68*(4), 376-385.

Iowa State University English Placement Test. (2020). https://apling.engl.iastate.edu/english-placement-test.

Iowa State University English Placement Test Rating Scale: https://apling.engl.iastate.edu/alt-content/uploads/2016/04/EPT-Oral-Communication-scale-updated.pdf

Isaacs, T., & Trofimovich, P. (2012). Deconstructing comprehensibility: Identifying the linguistic influences on listeners' L2 comprehensibility ratings. *Studies in Second Language Acquisition*, *34*(3), 475-505.

Jenkins, J. (2009). *World Englishes*. 2nd ed. London: Routledge.

Jenkins, J. (2011). Accommodating (to) ELF in the international university. *Journal of Pragmatics*, *43*(4), 926-936.

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

Jenkins, J. (2016). International tests of English: Are they fit for purpose? In H. Liao (Ed.), *Critical reflections on foreign language education: Globalization and local interventions*. Taipei: The Language Training and Testing Center.

Jenkins, J., & Leung, C. (2017). Assessing English as a Lingua Franca. In E. Shohamy, I. G. Or, & S. May (eds.), *Language Testing and Assessment*. (3rd ed. pp. 103-117). Switzerland: Springer.

Kachru, B. B. (Ed.). (1992). *The other tongue: English across cultures*. Urbana: University of Illinois Press.

Kim, H., & Billington, R. (2016). Pronunciation and comprehension in English as a lingua franca communication: Effect of L1 influence in international aviation communication. *Applied Linguistics*, *39*(2), 135-158.

Leung C. , Lewkowicz J. (2017) Assessing Second/Additional Language of Diverse Populations. In E. Shohamy, I. G. Or, & S. May (eds.), *Language Testing and Assessment*. (3rd ed. pp. 343-358). Switzerland: Springer.

Leung, C., Lewkowicz, J., & Jenkins, J. (2016). English for Academic Purposes: A need for remodeling. *Englishes in Practice, 3*(3), 55-73.

May, L. (2011). Interactional competence in a paired speaking test: Features salient to raters. *Language Assessment Quarterly*, *8*(2), 127-145.

McNamara, T. (2011). Managing learning: Authority and language assessment. *Language Teaching*, *44*(4), 500-515.

McNamara, T. (2018). *A challenge for language testing: The assessment of English as a Lingua Franca*, Language Assessment Research Conference, Ames, Iowa, March 21-23.

McNamara, T., Knoch, U., & Fan, J. (2019). *Fairness, justice, and language assessment*. Oxford: Oxford University Press.

Mortensen, J. (2013). Notes on English used as a lingua franca as an object of study. *Journal of English as a Lingua Franca*, *2*(1), 25-46.

Newbold, D. (2015). Engaging with ELF in an entrance test for European university students. In Y. Bayyurt & S, Akcan (Ed.). *Current Perspectives on Pedagogy for English As a Lingua Franca*. (pp. 205-222). Berlin: De Gruyter Mouton.

Ockey, G. J., & French, R. (2016). From One to Multiple Accents on a Test of L2 Listening Comprehension. *Applied Linguistics, 37*(5), 693-715.

Ockey, G.J., & Li, Z. (2015). New and not so new methods for assessing oral communication. *Language Value, 7*(1), 1-21.

Ockey, G. J., & Wagner, E. (2018). *Assessing L2 listening*. John Benjamins.

Prodromou, L. (2010). *English as a Lingua Franca. A Corpus-based analysis*. London: Continuum.

Seidlhofer, B. (2006). English as a Lingua Franca in the Expanding Circle: What it isn't. In R. Rubdy & M. Saraceni (Eds.), *English in the world: Global rules, global roles*. (pp. 40-50). London: Continuum.

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

Zhang, Y., & Elder, C. (2011). Judgments of oral proficiency by non-native and native English speaking teacher raters: Competing or complementary constructs? *Language Testing*. *28*(1), 31-50.

Appendix A. Iowa State University Oral English Placement Test Rubric

(Also available online at: https://apling.engl.iastate.edu/wp-content/uploads/sites/221/2016/04/EPT-Oral-Communication-scale-updated.pdf)

EPT ORAL COMMUNICATION ABILITY RATING SCALE

		Fluency ● Speaking rate ● Repetition/ self-correction and pauses ● Ability to speak naturally (e.g., effective use of fillers and markers)	Interactional competence ● Appropriateness of response to a given situation	Comprehensibility/ Pronunciation ● Individual sounds/word levels ● Stress, linking, rhythm, and intonation ● Listener effort to understand	Grammar/ Vocabulary ● Accuracy & range of grammatical structures ● Accuracy & range of vocabulary
P A S S	4	● Speech is **almost always** at an appropriate pace. ● Speech has **very rare** repetitions, self-corrections, or unnatural pauses. ● Speech is **almost always** natural (e.g., effective use of fillers and markers).	Response is **almost always** appropriate in any given situation, for example: ● initiating and expanding on own ideas ● connecting own ideas to a partner's ideas ● expanding on a partner's ideas ● making relevant comments ● taking turns appropriately ● asking appropriate questions ● (dis)agreeing politely ● answering questions in an appropriate amount of time	● Speech is **almost always** clear with well-articulated individual sounds and accurately pronounced words. ● Speech shows good control of stress and intonation; words in an utterance are almost always accurately and effectively blended. ● Speech variety does **not** require focused listening and does **not** interfere with comprehension.	● Speech **almost always** shows a range of accurate grammatical structures. ● Speech **almost always** shows a range of accurate use of academic vocabulary.
	3	● Speech is **usually** at an appropriate pace. ● Speech may have **a few** repetitions, self-corrections, or unnatural pauses. ● Speech is **mostly** natural (e.g., effective use of fillers and markers).	Response is **usually** appropriate in any given situation, for example: ● initiating and expanding on own ideas ● connecting own ideas to a partner's ideas but may not fully expand on a partner's ideas ● making relevant comments ● taking turns appropriately ● asking appropriate questions ● (dis)agreeing politely ● answering questions in a somewhat appropriate amount of time	● Speech is **usually** clear with well-articulated individual sounds and with accurately pronounced words. ● Stress and intonation patterns **may not be completely accurate**, but this **does not interfere** with communication; words in an utterance are accurately and effectively blended. ● Speech variety **may** require focused listening, but is completely comprehensible.	● Speech **usually** shows a range of accurate grammatical structures. ● Speech **usually** shows a range of accurate use of academic vocabulary.
N O T P A S S	2	● Speech is **generally** at an appropriate pace. ● Speech **may have some** repetitions, self-corrections, or unnatural pauses. ● Speech is **generally** natural (e.g., a little misuse of fillers and markers).	Response is **generally** appropriate in any given situation, for example: ● initiating but may not expand on it very well ● speaking without completely connecting own ideas to a partner's ideas ● making relevant comments ● taking turns appropriately ● may ask questions that are not completely appropriate ● may not (dis)agree completely appropriately/politely ● may not answer questions in a completely appropriate amount of time	● **A little** mispronunciation of individual sounds and words might be present and may **slightly interfere** with communication. ● Stress and intonation patterns may be present and may **slightly interfere** communication; words are accurately and effectively blended in an utterance to some extent. ● Speech variety **requires** focused listening and may result in **slight lack** of comprehensibility.	● Speech **generally** shows a range of grammatical structures, and **accuracy** may not be completely consistent. ● Speech **generally** shows a range of academic vocabulary. Some **errors** in vocabulary may be present but rarely hinder communication.
	1	● Speech is **often too** fast or slow. ● Speech **may have frequent** repetitions, self-corrections, or unnatural pauses. ● Speech **may not be quite** unnatural (e.g., some misuse of fillers and markers).	Response is **often** not appropriate in any given situation, for example: ● rater may assume a speaker cannot understand questions or what a partner says ● may not initiate and develop topics ● may not contribute much to the discussion ● may respond minimally and irrelevantly to a partner ● may not ask appropriate questions ● may not (dis)agree politely ● may not answer questions in an appropriate amount of time	● **Mispronunciation** of individual sounds and words may **often interfere** with comprehensibility. ● Stress and intonation patterns may be missing and may **often** cause difficulty for comprehension; words may not be accurately and effectively blended in an utterance. ● Speech variety **requires** focused listening and **may substantially** interfere with comprehensibility.	● Speech **often** presents a range of grammatical structures; grammatical **errors** may **usually present**. ● Speech **often** shows a range of academic vocabulary. Some **errors** in vocabulary may be present and **hinder** communication to **some extent**.

Table 2.1. Framework for evaluating the extent to which a test can be considered ELF

Criteria	Description	
English rhetorical sensitivity	How effectively the test evaluates the test taker's ability to...	- use awareness of linguistic and cultural differences to identify and use the appropriate rhetorical style to communicate effectively with English speakers from cultures other than their own in a variety of contexts
International communicative competence		- comprehend and be comprehended by advanced speakers of English from any culture
Context sensitivity		- show respect for and confidence in local varieties of English while exemplifying the capacity to contribute to the international body of information
Motivation		- communicate in ways that are appropriate for personal or cultural goals
Grammatical appropriacy		- communicate in a way that is grammatically mutually comprehensible to speakers of English from any culture
Relevance	Qualities of assessment	- should include materials and activities based on local and international situations that are recognizable and applicable to the students' everyday lives
Fairness		- should be based on an inclusive model that incorporates NNESs and NESs in every aspect of the assessment