Assessing English as an International Language for Government Officials

By Douglas Meyer

A thesis submitted to the
Victoria University of Wellington
in fulfillment of the requirements for the degree of
Doctor of Philosophy in Applied Linguistics

Victoria University of Wellington
2017

Acknowledgements

Abstract

The function of English as an international language (hereafter called EIL) is firmly entrenched in many aspects of modern life, such as academia, aviation, business, education, international relations, science, and tourism. This trend has led high-stakes English language proficiency tests to serve increasingly critical gate-keeping roles for people seeking admission to tertiary institutions, employment overseas, and career advancement. However, while language tests focus on language proficiency, they do not measure the ability over and above core linguistic competence that enables and enhances successful international communication. As more and more English communication occurs between people from different language and cultural backgrounds, research and testing of these skills is warranted.

This thesis identifies receptive intelligibility (listening skills), productive intelligibility (speaking skills), intercultural competence, and strategic competence in negotiating meaning as key abilities that enhance the use of EIL. An investigation of these four abilities sets the groundwork for their conceptualization and operationalization into a test battery aimed at government officials from developing Asian and African countries who have gone through English-language training at Victoria University of Wellington. Results from classroom observations and a survey of the participant's target language use tasks serve to highlight the contexts and purposes for which they use English, as well as providing the contexts for developing test tasks and items and conceptualizing skills used to improve international communication. The four abilities identified above were then operationalized into testable items for an online test battery consisting of four constructs and a total of 76 items. The purpose of this test battery is to serve as a low-stakes assessment tool for making decisions regarding the international communicative competence of government employees who use English for international communication.

The validation of the test battery uses the Bachman and Palmer (1996) approach for determining test usefulness, which is shaped by six test qualities. These qualities form a validation argument supported by evidence for construct validity, reliability, authenticity, interactiveness, practicality, and impact. Results

from the validation of the test battery show that the correlation between the four constructs (receptive intelligibility, productive intelligibility, intercultural competence, and strategic competence) is positive, however each of the theoretical constructs was found to be much deeper and broader than the test battery is able to measure. Intercultural competence was found to be the least related to intelligibility and strategic competence, suggesting that intercultural problems are less immediate barriers to successful international communication.

The discussion focuses on the major issues with the test battery, such as construct under-representation, internal consistency problems and the logistical challenges of online testing. This thesis concludes with the methodological limitations of the test battery and suggestions for broadening and deepening test items to account for the complexity and multidimensionality of these four constructs. While the test needs further development, it is a first step into assessing the dynamic and fluid nature of English as an international language.

Table of Contents

List of abbreviations

A conscious effort has been made to minimize the number of abbreviations used in this thesis. The abbreviations that are used are listed here.

ASEAN   Association of Southeast Asian Nations

EIL   English as an International Language

ELTO   English Language Training for government Officials

EU   European Union

IELTS   International English Language Testing System

ILR   Interagency Language Roundtable

TOEFL   Test of English as a Foreign Language

TOEIC   Test of English for International Communication

Table of Figures

List of tables

Chapter 1:   Introduction

<div style="text-align: right; font-size: large;">1</div>

1.1   Background and context

The beginning of the 21st century has given rise to an unprecedented surge in human interconnectivity, catalyzed by the Internet, increased migration, and the dynamics of a global economy. People from all walks of life are communicating with each other unlike anything seen before in human history, and the most popular medium for this international communication is English. In this postmodern era, English has emerged as the dominant international language, spoken to some degree by an estimated two billion people (Crystal, 2003; Graddol, 2007) around the world. This estimate, intended to include *all* learners of English, should be regarded with some cautious skepticism. The global spread of English is the result of centuries of colonization, war, and trade (first by the British and then continued by the Americans). The result of these international activities has seen English become the de facto language of aviation, academia, business, diplomacy, music, pop culture, science, technology, and tourism.

With this explosion of English language users comes a pivotal shift in demographics. It is more recently estimated that non-native speakers now outnumber native English speakers by a ratio of 4:1 (Crystal, 2012, p. 155), raising the issue of the ownership of English and potentially marking the beginning of a shift in the standards and norms of English language use. The approach used in this thesis takes this demographic shift into account with regard to the increasing number of international and intercultural contexts and how English as an international language may be assessed in the future.

1.2   Motivation for this research

My motivation for pursuing this research stems from nearly twenty years experience teaching English in South Korea and Japan. My approach was to instruct students in *proper* English, with the aim of creating language learners who follow native speaker norms. Over the years, I eventually trained to become

a language proficiency test examiner for one of the major language testing companies in Japan. During one particular test, an African English speaker entered the interview room for a one on one speaking test. As the African candidate spoke, I pondered why a fluent native-speaker of English would be sitting a language test. She earned the highest scores warranted by the scoring rubric, as she was effectively a native speaker in my opinion. However, I later learned that senior examiners, who found African English to be 'non-standard', downgraded her score substantially.

As a result, I became interested in the parallel notions of Standard English and language assessment. I believe these two notions are too narrowly defined to accommodate the diverse contexts where English is used internationally. There is simply much more going on in successful international communication than can be addressed either by the limitations of Standard English or measured by standard language proficiency tests. Therefore I began this research with the goal of establishing a broader view of communicative competence for international contexts by attempting to measure abilities above and beyond the assessment of language proficiency.

1.3   Research objectives

Rather than identifying and answering a number of research questions, this thesis has three research objectives: the first aim is to outline a new approach for assessing international communicative competence in English. This new approach to assessment will require a substantial shift in the minds of educators and testers (Canagarajah, 2006), who for years have been following a pedagogy of language instruction that places the educated adult native speaker at the peak of language development. Successful international communication requires more than linguistic skills, and this will be discussed in Chapter 3.

The second aim is to see the development of theoretical constructs (argued to be beneficial for successful international communication) operationalized into an online test battery for English as an international language (hereafter called EIL). The test takers are mid-level government officials from thirteen developing nations who use English both in the workplace and in international contexts.

The third aim of this research is the empirical validation of the online test battery and a discussion of the results. The Bachman and Palmer (1996) *six qualities* approach (reliability, validity, authenticity, interactiveness, impact, and practicality) was determined to be a suitable method for estimating the degree of usefulness of the online EIL test battery, which is an exploratory assessment tool rather than a more formal test with real-life decisions and consequences.

1.4   Research importance

This research project contributes to our knowledge of assessment by addressing four main points that frame the debate around the importance of EIL and the need for a fresh look at assessing international communicative competence: 1) the problematic native speaker standard, 2) the changing demographics of global language users, 3) increasing awareness of EIL, and 4) the impact of high-stakes assessment. First, a look at the argument challenging native speaker standards (on which current testing is based) is persuasive. There is an assumption in second language acquisition that the learner's goal is to achieve native-like competence rather than using English to suit one's immediate needs in given contexts. Bi- or multilingual English speakers do not seem to follow native speaker norms when they speak with each other (Jenkins, 2006c; Seidlhofer, 2004). The reliance on native speaker norms raises the question of whether current language testing actually meets the needs of the test takers and opens up new avenues for research in testing EIL.

Second, native speakers of English are increasingly outnumbered by the growing masses of people who draw upon the power of English as a means for advancement and wider communication. Estimating the number of people worldwide who speak English (to some degree) is complicated, and section 2.2.1 will examine the range of some rather large estimates. For the vast majority, English is an additional language. Migration is one driver of language spread, and in the forty year period between 1960 and 2000, international migration has doubled (Graddol, 2007). Affordable commercial air travel, the opening up of former Communist block countries, and increasing flows of migrant workers has also facilitated this unprecedented movement of humanity. As a direct result of increased migration, interactions between people from different countries have

increased accordingly, resulting in a pressing need for international communication skills in English.

This research is also important in terms of increasing awareness of EIL when considering the problematic native speaker standards and changing demographics of language users. Educators should expose students to different varieties of English so that they might understand other EIL users better. Native speakers should reflect upon the manner in which they speak English in order to improve communication between people from different countries. An investigation into the abilities that enhances international intelligibility is certainly warranted.

The mid-level government officials who are the participants in this research are ideal subjects for a study on EIL. For many of these ministry officials, English is a vital skill in the postmodern era of international cooperation and interconnectivity. English serves as the working language of many international organizations, serving the functional purpose of broader communication. Care in using this shared language must be taken to avoid any misunderstandings or miscommunications, which might result in serious consequences for international relations.

Finally, this research addresses a complex gap in high-stakes English assessment. While many of the current major English proficiency tests claim to be international, this is essentially true only in the sense that there are hundreds of test centers worldwide. The native speaker standards of a small handful of countries remain at the heart of assessment, ignoring the reality of global English and the varieties that naturally develop over time. We live in a world where intercultural communication is on the rise, yet the core of English language assessment remains fixed on a standard that is used by a minority of language users. The test battery developed in this thesis attempts to measure abilities that repair and/or enhance international communication, abilities which are arguably needed by native speakers themselves in the post-modern world. On these grounds, this research takes the next step from theorizing about the functionality of EIL (and how it might be assessed) to actually *conceptualizing* and *operationalizing* such a test. To the knowledge of this researcher, no other

work has attempted to actually test the EIL abilities which rise above and beyond basic language proficiency.

1.5 Research organization

This thesis has three main sections covering the conceptualization, operationalization, and validation of a test battery of EIL abilities. Chapter 2 will lay the foundations by introducing a number of the central terms surrounding EIL, terms which are often overlapping and may create confusion unless unpacked and clarified. Terms such as ESL, EFL, TOEIC and IELTS are introduced, but efforts are made both to distinguish these terms and to avoid the excessive use of acronyms. Chapter 2 addresses the changing demographics of English language users worldwide, a change which has long warranted a re-evaluation of testing standards, or at least the recognition that more communication takes place between non-native speakers of English than native speakers (Crystal, 2012; Graddol, 2007). Chapter 2 also links the need for English language skills to the government officials described in this thesis by discussing the language policies of international organizations and the importance of clear EIL communication in diplomacy.

The conceptualization of EIL abilities is discussed in Chapter 3, following a review of the literature relating to communicative competence. The theoretical evolution of communicative competence from the early linguistic native-speaker model to recent models of international communicative competence is traced, leading to the identification of four core constructs of EIL abilities. These theoretical constructs are also based on classroom observations, a survey of EIL users, and a top-down literature-informed approach. These constructs, defined in the following chapter, are argued to be necessary for successful international communication.

Chapter 4 is the operationalization of EIL abilities into concrete testable items. This chapter has two objectives: first it will outline the design statement and blueprint of an EIL test battery, identifying who the test is for, the mandate for the test, and the operational definitions of the constructs. The second objective of this chapter is the operationalization stage of test development, which will see the theoretical constructs become measureable test items, with

examples provided. Moreover, the online platform used for delivering the test will be described, along with scoring and relevant technical aspects.

Chapter 5 covers the empirical validation of this test by means of statistical analysis and supporting evidence that aim to answer the overriding question of how useful a test is for its intended purpose. Bachman and Palmer's (1996) framework of six test qualities will be used to determine test usefulness in the case of the EIL test battery. These qualities are reliability, validity, authenticity, interactiveness, impact, and practicality. The first test quality is statistical reliability, which will be analyzed using the internal consistency estimate (Cronbach's alpha). Second, evidence will be presented to support the construct validity of the EIL test battery. Third, the authenticity of the test will show to what extent the score interpretations can be generalized beyond the test performance itself, in support of the overall construct validity. Next, interactiveness will discuss the ways in which the test taker's language strategies, topical knowledge, and affective schemata are engaged by the test tasks. Then the section on the impact of the test battery will conclude that as a theoretical test of EIL abilities, little or no impact is expected on the test takers. The final test quality will examine to what degree the EIL test battery is practical, largely in terms of the time and money required for the continued development and administration of such a test. These qualities taken together will support the EIL test battery as a low stakes test suitable for moderate-level decisions about test taker's abilities to use English successfully in international contexts.

The Discussion Chapter opens with a brief review of the research before addressing validity concerns that threaten the use of the EIL test battery for its intended purpose. This is followed by a summary of the major findings from the test battery results, as well as possible explanations for these results. The Discussion Chapter then outlines potential real-world implications of this research on both classroom teaching and testing.

Chapter 7 begins by highlighting the theoretical and methodological contributions made by this research, as well as the limitations it faces. This is followed by a discussion of future directions the EIL test might take. These future directions are only stepping-stones, and it is hoped that this foray into the theoretical assessment of EIL is strengthened and built upon by others.

Chapter 2: Changing perspectives 2

This chapter is organized in to six sections that shape the arguments that form the foundation for this thesis. First, Chapter 2 opens with a review of the background of English as an international language, followed by a discussion of the key concepts and terms that underlie the premise for this research. Several overlapping concepts such as world Englishes, English as a lingua franca, English as an international language, and Standard English need to be defined and distinguished as a prelude to the arguments presented in the following chapters. Second, Chapter 2 will briefly discuss the spread of English before focusing on the changing demographics of English language users. The groundswell of English language users worldwide is now challenging the existing concept of Standard English and how it is assessed. Next, high-stakes English language testing will be critiqued in light of the changing demographics now confronting test developers. The case will be made that current language proficiency testing is primarily concerned with upholding the standard norms at the native English speaking 'center' rather than adapting to the needs of the people the test should serve. Fourth, a look at the language policies of the Association of Southeast Asian Nations (ASEAN) and the European Union (EU) will show the completely different approaches towards using English as a working language of these international bodies, and how these language policies have fueled the demand for English language skills in these countries. The following section on diplomacy describes the test takers and why they are considered ideal for research into testing EIL. The final section will make a distinction from existing work on EIL and similar topics such as English as a lingua franca.

2.1 Clarifying key terms and concepts

Given the depth of the global spread of English and the diverse multilingual contexts that exist, it is not a surprise that a number of terms are used to describe the same phenomenon. Smith (1976) was the first to coin the term *English as an international auxiliary language* with the aim of describing the *denationalization* of English (away from the U.K. and U.S.). This term was later shortened to English as an international language by Smith (1983). However, it

would be several years later that the topic gained interest among scholars, who coined several new terms in the literature: English as a medium of intercultural communication (Meierkord, 1996), World Standard Spoken English (Crystal, 1997), Global English (Bolton, 2002), International English (McKay, 2002), English as a global language (Crystal, 2003), Lingua Franca English (Canagarajah, 2007b), English Lingua Franca, (Mesthrie & Bhatt, 2008), and even *Globish* (Nerriere & Hon, 2009). The two dominant terms emerging from this academic quandary are presently English as a lingua franca and English as an international language.

This goal of this section is to distinguish and separate the most commonly used terms: EIL, world Englishes, and English as a lingua franca. Traditional concepts such as the Three Circles of English use, Standard English, and native speakers will also be introduced and critiqued from a modern EIL perspective. Finally, the issue of accents and intelligibility will be introduced in terms of the role they play in successful international communication.

2.1.1 English as an international language

EIL in this thesis refers to the English spoken by all speakers of English in international contexts, in what Canagarajah calls the "...heterogeneous global English speech community, with heterogeneous English, and different modes of competence" (2006b, p. 211). As much as possible, the approach taken herein aims to avoid the traditional native/non-native speaker comparisons often found in the EIL literature. Moreover, comparisons between English speakers from different countries are not a research objective since a broader approach is being taken. Inner circle, outer circle, expanding circles (see section 2.1.4) of English use distinctions which are arguably no longer relevant in this modern age of mobility, where the flow of goods and people across linguistic and national borders is an expected consequence of globalization. In effect, this thesis aims to level the playing field and argue that it does not matter who one is, or where they are from, but rather attempt to measure how effectively anyone can wield the English language in international contexts.

Much of the literature on EIL is devoted to explaining what EIL is *not*, rather than clearly identifying what exactly it *is*. This is a valid criticism, since

EIL is far too dynamic to be hammered down into a concrete descriptive language code. Likewise, a prescriptive approach towards EIL is not a useful objective given the polymorphous nature of the English language and the concept of a monolithic standard model of English (Canagarajah, 1999; B. B. Kachru, 1994). The focus of EIL herein will be on the *functional* use of EIL, which extends above and beyond the traditional four skills (reading, writing, speaking, listening) of language proficiency exams. Seidlhofer (2011, p. 25) agrees, arguing that EIL use should be "…functionally and not formally defined; it is not a variety of English but a variable way of using it". Seidlhofer (2003, p. 12) goes on to argue that the functionality of EIL is *econocultural* fact, positing the motivation for learning English as both utilitarian and idealistic. Although language proficiency certainly is a major component of the EIL skill set, it is not the only component of successful international communication. Chapter 3 will examine in more detail what these theoretical components might be, and argue for a shift towards functionally relevant constructs of EIL.

2.1.2   World Englishes

The term *world Englishes* is used to describe the dozens of local or indigenized varieties of English around the world (primarily in former British colonies) over the decades. World Englishes are used in the legal/political system, in education, business and finance, and as a common shared language *intra*-nationally. Some examples include the variety of English used in Australia, India, Kenya, Philippines, South Africa, and Singapore. In these countries (and others), world Englishes fully belong to the people who use it daily as their first or additional language, whether in a standard or localized form (B. B. Kachru & Smith, 1985). These varieties, with both marked and subtle differences, are not the result of faulty learning by non-native speakers, but rather "from the nature of the process of microacquisition, language spread, and change" (Brutt-Griffler, 2002). In other words, world Englishes is a natural result of language evolution over time to suit the different contexts in which the language finds itself.

EIL differs from world Englishes in two ways: first, world Englishes is used for local purposes in local contexts, while EIL aims at broad international communicability. For example, the English used by a New Zealander talking with

friends they grew up with would be an example of world Englishes. However, that same New Zealander would have difficulties being understood if they spoke in exactly the same manner to Chinese in Asia or Brazilians in South America. Second, world Englishes are relatively stable in their contexts, whereas EIL is much more fluid and dynamic as people attempt to find common communicative ground. People using EIL in international contexts tend to avoid regional varieties of English, which may contain unknown words and expressions for the listener.

2.1.3   English as a Lingua Franca

Modern day English as a lingua franca has been defined various ways: first, as "a 'contact language' between persons who share neither a common native tongue nor a common (national) culture, and for whom English is the chosen *foreign* language of communication." (Firth, 1996, p. 240). Second, "…as interactions between members of two or more different linguacultures in English, for none of whom English is the mother tongue." (House, 1999, p. 74). Third, "any use of English among speakers of different first languages for whom English is the communicative medium of choice" (Seidlhofer, 2011, p. 210). Other definitions do exist, but are similar to the above concepts.

Despite the acknowledgment that non-native speakers constitute the largest group of English users, research into English as a lingua franca only started in the 1990's (Jenkins, 2006a). English as a lingua franca tends to focus on a phonological lingua franca core (Jenkins, 2000), lexicogrammar (Seidlhofer, 2004), pragmatics (Meierkord, 1996), and corpra (Mauranen, 2003). However, English as a lingua franca has been criticized (see Holliday, 2005; Prodromou, 2007) for promoting a systematic codification (standardizing) of English as an international variety in its own right, or even a new standard form that learners should speak. Codification of English as a lingua franca has been difficult to achieve, given its fluid and dynamic nature. Moreover, such codification has led to criticism that it may represent a new standard that has been perceived as deficient, lacking, or as a sub-standard learner variety. This criticism is misguided and rejected according to those in the lingua franca field (Dewey & Jenkins, 2010; Firth, 2009; Trudgill, 1999). English as a lingua franca has been

rejected as a linguistic variety by Berns (2009), Canagarajah (2007a), and Friedrich and Matsuda (2010).

The question of whether or not native speakers of English are included in such definitions is answered by Jenkins (2004), who concedes that English as a lingua franca communication includes native speakers but notes that given modern linguistic demographics, such interactions are in the minority. However, when native speakers are in English lingua franca conversations, it cannot be assumed that their variety of English is ideal; native speakers should adopt accommodation strategies to ensure a high degree of intelligibility (Dewey & Jenkins, 2010). Something that is often lost in the discussion, however, is that English as a lingua franca is simply an expression used by applied linguists – most people just think that they are just using ordinary English, as noted by Mackenzie (2014). This ordinary English notion lends itself not to any particular variety of English, but rather to the *purpose* for which English is used (e.g. academia, commerce, entertainment, or travel). English as a lingua franca has not yet been formalized and institutionalized "...so it is not possible to teach and learn it. Therefore it seems that teachers as well as learners have to concentrate on the *function* of ELF and on the structure of an ENL [English native language] variety for successful international communication. (Gnutzmann, 2005, p. 21 emphasis added).

To emphasize the differences between the terms EIL and English as a lingua franca, the latter is usually referred to when discussing non-native speakers (Jenkins, 2004, 2006a). However, Jenkins does admit confusion in EIL and English as a lingua franca terminology in the literature, noting, "...to add to the confusion, both terms are currently in use" (2006a, p. 160). Although native speakers of English can be included in English lingua franca conversations (Cogo & Jenkins, 2010) such interactions are eclipsed by the sheer number of non-native speaker interactions. In this thesis, both native and non-native speakers of English are considered EIL users when using English in international or intercultural contexts.

Long before modern day research began into English as a lingua franca, the notion of *where* English was used and by *who* was divided into a series of circles, essentially labeling billions of people in terms of first, second, or

additional language English language users. This notion was first put forward by Kachru (1985, 1986) and is discussed in the following section as part of the conversation on international English use.

2.1.4 Circles of English use

Although Kachru's (1985, 1986) three-circle model of English is largely out of date, it helps to explain the original hegemony behind high-stakes testing institutions (see section 2.3.1) based in the United Kingdom and United States. This model places largely monolingual countries where English is the first language in the center as the norm providers of English. In other words, the population's use of English is widely considered within the norms of correctness and appropriateness insofar as language planners and curriculum developers are concerned (Bruthiaux, 2003). The center of this three-circle model includes Australia, Canada, New Zealand, the US and UK.

A larger outer circle identifies nations – such as India, Nigeria, and the Philippines – where English is widely used as a second language for *intra*national functions such as administration, education, business, and as a common language. These countries are said to be norm developing, according to Kachru (1985, 1986). English was once a foreign tongue in these lands, but due to history and decades of colonialism, it is now firmly rooted in society.

The final expanding circle identifies countries where English is largely a foreign language, with no strong historical or sociopolitical connection to that country (Russia, Brazil, China, Saudi Arabia). English has not colonized the linguistic landscape, nor has it been given an official or administrative status. English is recognized as a useful international language, and may be taught in schools, but as a *foreign* language. Very generally speaking, much of the population in the expanding circle speaks little if any English.

Kachru's (1985, 1986) three-circle model has been criticized as "…a 20th century construct that has outlived its usefulness…" (Bruthiaux, 2003, p. 161) since people and language functions have not been constrained in any way by these three circles. Mass migration and mass learning of English has resulted in a situation where "…the lines separating these circles have become more permeable." (McKay & Bokhorst-Heng, 2008, pp. 29–30). Jenkins (2003)

laments both the way the Three Circles model is linked to geography and places of birth, as well as the fact that the so-called inner circle implies that native speakers have a central role to play in English development (when in fact their influence is on the decline). While the term *inner circle* is considered to be from an out-of-date model, it remains a popular shorthand way to refer to five different English-first language countries: Australia, Canada, New Zealand, the United Kingdom, and the United States.

Other circular models include McArthur's (1987) *Circle of World English* and Görlach's (1990) *Circle model of English*, both of which placed a 'standard' or 'international' English at the center and other varieties of English spreading outward. Neither model can explain what exactly standard or international English is, only that it exists at the center. Continuing efforts to round off models for English language use, Modiano (1999a, 1999b) attempts to adapt to the global spread of English with his centripetal circles of international English and English as an international language models, the former following quickly on the heels of the latter. The former model is problematic in that in addition to proficiency, speakers in the center should have no strong regional accent or dialect (Modiano, 1999a), which is unusual since all speakers have an accent. Although Modiano does recognize the importance of intelligibility in EIL, he identifies strong regional accents as a barrier to intelligibility, leading one to ask which accents are intelligible to whom. The second Modiano EIL model moves away from the intelligibility approach and suggests a common core serve as the basis for EIL, without identifying what these core elements are. Moreover, Modiano implies that all native speakers are also competent EIL users, something Jenkins finds "...patently untrue." (2003, p. 21).

There is little doubt that attempting to pigeonhole or model something as large and complex as EIL is a daunting task. Scholars on this topic worldwide have widely varying views, ranging from traditionalist (Quirk, Widdowson, & Cantù, 1985) to anti-imperialist (Phillipson, 1992) to contemporary (Canagarajah, 2006; Jenkins, 2000; McKay, 2002). Most of these scholars can only agree to disagree. This thesis will not attempt to model EIL or suggest prescriptive measures for speaking a core or Standard English. Rather it aims to recognize EIL for what it is: a language in the state of flux as used by mobile EIL

users who must negotiate different varieties and abilities of English and exercise strategic competence to minimize communication breakdown.

Attempts to describe the complexity of English language use extend beyond the historical and geographical circular models described above. The following section shows how *the order* in which additional languages are acquired has led to further labeling of language users into pigeonholes of stereotypes and expectations that do little to clarify the situation.

2.1.5   English as a first, second, foreign, or other language?

Numerous terms describe the English-language environment around the world. Traditionally, such terms refer to a historical relationship with colonial powers and the degree to which English has penetrated a society. The most common terms are English as a second language (ESL) and English as a foreign language (EFL), differentiated by the context in which the learning takes place.

In addition to the two terms above, English as an *additional* language is sometimes used to circumvent the problem of numbering which languages are learned in which order and in which context. These labels form a complex fuzzy web that while generally understood by educators, becomes convoluted under examination. These terms, which have been used for decades, "…are often overlapping in definition and neither EFL or ESL seems adequate to fully describe the present state of English language usage. These terms have been a source of awkward explanation for us within the field." (Smith, 1983, p. 13).

In the case of using EIL in international contexts, where English is learned is largely unimportant. Moreover, whether it is learned as a first, second, third or other language matters little since research shows (see Smith & Rafiqzad, 1979) that even some native speakers of English are not as widely intelligible as previously believed. Although some people may have considerable advantages learning English in their particular environment, international contexts have a way of creating a level playing field among speakers. Those with greater language variety awareness, openness and flexibility – albeit with moderate proficiency – may excel at international communication while a highly proficient – but inflexible –native speaker may struggle. If this is the case, one may wonder why *proficiency* is held with more regard than *flexibility* in using English.

Table 2.1 brings together these terms into a summary, comparing the main features of the terms discussed thus far: ESL, EFL, and English as a lingua franca.

Table 2.1: Comparing ESL, EFL and English as a lingua franca

English as a second language (ESL)	• Students are regularly exposed to English inside and outside of their classrooms, with ample opportunities to use English. • English is a colonial language. • The proficiency of the average speaker tends to be higher. • English is used mainly for domestic purposes, such as administration, education, law, and commerce.
English as a foreign language (EFL)	• Students tend to use English only in the classroom, with very few opportunities to use English in their daily life. • English is a foreign language, not widely used in society. • Proficiency of an average learner tends to be lower. • Language education aims at the native speaker norms of usage. • Students lacking native speaker norms are considered to be in deficit. • Code mixing and switching are considered as first-language interference.
English as a lingua franca (ELF)	• Meaning tends to be negotiated. • Meaning is given priority over accuracy. • Focus is on accommodation and adapting to audience. • Part of world Englishes. • Code mixing and switching are considered as bilingual resources.

(Jenkins, 2006b; Kirkpatrick, 2010; Seidlhofer, 2011)

Table 2.1 compares and contrasts these terms and the connection to the context in which they are used. Moreover, the various perspectives of English language use are revealed, such as whether native speaker norms are to be adhered to or whether a more flexible approach to communication is used. Jenkins (2006b) contrasts EFL teaching with the more international approach of English as a lingua franca, whereby differences in language output (compared to native speakers) are regarded as a deficit or error to be corrected. First language influences on English are considered interference, and language code switching is regarded as a handicap. These key differences are a matter of perspective, and require a paradigm shift in order to address properly.

This multiplicity of different types of Englishes reveals a confusing state of affairs within the discipline of world English studies. In developing the names of language varieties, the context and conceptualization of these names leads to either a descriptive or revisionary definition. In Seargeant's (2010) categorization of the names of Englishes, one can see how the different discourses of world Englishes have evolved different categories for similar or overlapping terms. The literature on EIL is a perfect example, with many scholars using the term interchangeably with English as a lingua franca, clouding discussions on the issue. On the other hand, a term such as 'world Englishes' is a less contentious term with widespread agreement about its meaning and application.

In Table 2.2, Seargeant (2010) attempts to clear the waters by providing a taxonomy of six categories of English based on fundamental distinctions in the way the term is conceptualized in current academic discourse. These categories are 1) function: for what purpose is the variety used? 2) Community: who speaks the variety? 3) History: how did the variety evolve over time? 4) Structure: what are the structural features of the variety? 5) Ecological model: how does the variety fit within a model of other varieties? 6) Multiplex: varieties that do not fit neatly within the taxonomy but remain alternatives terms for English itself.

A taxonomy of the names of English

Category	Names
Function	English as a second language; English as a foreign language; English as an additional language; English as an international language; English as a lingua franca; International English; Word Standard Spoken English.
Community	Metropolitan standards; Regional dialects; Social dialects; Immigrant Englishes; Native/non-native varieties; Global English.
History	Language-shift Englishes; Colonial standards; Indigenized Englishes.
Structure	Pidgin Englishes; Creole Englishes; Hybrid Englishes.
Ecology	Inner circle varieties; Outer circle varieties; Expanding circle varieties; World Englishes; New Englishes.
Multiplex	World English; English language complex.

(Seargeant, 2010, p. 101)

This taxonomy reveals how the complex terminology used in world Englishes is connected to the society, culture, as well as the functions served by the variety. Chapter 3 will describe the approach used in defining EIL, following the unpacking and examination of more related terms that help shape central notions around EIL. One of the most problematic terms when discussing EIL is that of native speaker, a controversial but unavoidable topic entrenched in the EIL literature.

2.1.6 Native Speaker

In simple terms, a *native speaker* of English refers to a person for whom English is the first language learned; a mother tongue used for daily communication. Rampton (1990) summarizes the features many people associate with native speakers, such as the idea that the language of a native speaker is inherited at birth, and an individual either is, or is not a native speaker, with no middle ground allowed. Cook (1999) adds to this summary a number of native speaker characteristics, such as a subconscious knowledge of the rules, a range of language skills, an intuitive grasp of meaning, fluency, creativity, and identification with a specific language community.

The concept of native speaker can also be linked to identity, where one is from and who they are. Graddol (1997) considers the term native speaker as a means of locating certain individuals in native-speaking countries at the center of the English-speaking world and the source of models of correctness. However, if it can be argued that since English is an international language, it is in a sense de-nationalized, and therefore difficult to connect to a country, identity or speech community. Rather than national identity and proficiency, Davis (2003) argues that a native speaker is more of an attitudinal and symbolic notion; attitudinal in that the concept of native speaker is more of an outlook on language use, and symbolic in that the native speaker is held up as an ideal model of language use.

As can be seen, the definition of a native speaker of English is problematic and far from simple. Any comparisons of non-native to native speakers – who themselves vary considerably in performance (Hamilton, Lopes, McNamara, & Sheridan, 1993) is unnecessary. What is important to EIL is expertise in using

English for *effective international* communication rather than expertise in using accurate forms, extensive vocabularies, and an innate sense of correctness. Second language acquisition research is far too concerned with the constructs of native-like competence and strong integrative motivation to fit in with a native-speaker community (see Sridhar & Sridhar, 1994), ignoring the *function* English serves as a formal register, a means to access information, or to even simply pass an exam or get a promotion at work. With no "…satisfactory characterization of the term 'native speaker', then it is foolish to accept the construct of native speaker as a model of competence" (McKay, 2002, p. 31). This is not to argue that native speakers have no role in the development and assessment of EIL, but rather to argue that the time has come for a broader approach to what constitutes effective English language communication. This time is actually overdue, given "…there is clearer evidence that English is used more in multinational contexts by multilingual speakers than it is in homogeneous contexts of monolingual speakers" (Canagarajah, 2006a, p. 232).

2.1.7 Standard English

Standard English is a topic closely linked with the concept of native speaker despite the argument that both terms are problematic. For example, Strevens (1981) links Standard English to a dialect of *global currency*, ignoring the fact that there is no dialect or accent associated with Standard English. Trudgil (1999) defines Standard English as a variety spoken by an educated native speaker without defining 'educated', resulting in some sort of idealized construct (Widdowson, 2012). McArthur (1998) narrows this educated speaker of Standard English further, identifying his/her origins as middle-class and having completed secondary school.

Despite these problems, the Longman Dictionary of Applied Linguistics (Richards & Schmidt, 2013) describes the standard variety of language as that which is widely used by educated native speakers, as represented in the media, literature, education system, and grammar books. The norms for Standard English "…are the linguistic forms that are *actually used* by institutions and individuals that have power and/or influence…" (Lowenberg, 2012, p. 87 original

emphasis). However, people tend to use language as it suits them, regardless of what powerful institutions might wish.

McNamara (2000) agrees, proposing that English language assessment reflect linguist forms that are actually used, for example, in university contexts. McNamara underscores the gap between Standard English testing and the *functions* of English language use, encouraging readers to consider the example of

> ...the TOEFL test, used primarily for selection of international students to universities in the United States. Given the diversity of varieties of English, both native and non-native, typically encountered in the academic environment there, it might be argued that it is responsible to include examples of those varieties in the test rather than to include only samples of the standard variety. (McNamara, 2000, p. 73).

Another problem with Standard English is that English lacks any official body setting and prescribing language norms, unlike the Real Academia Española in Spain, the Académie française in France, and over 100 other language-regulating bodies that exercise varying degrees of influence over their perspective languages. These regulatory bodies have generated tension between those who seek stability of the code through some form of convergence and the forces of linguistic diversity that are inevitably set in motion when new demands are made on a language that has assumed a global role of such immense proportions.

Language usage naturally varies from country to country, resulting in national dictionaries and innovative norms. It is this innovation that is interesting for EIL researchers, who tend to regard such flexibility and openness for English language use as a natural consequence of the spread of English. Kachru explains that "This sociolinguistic fact must be accepted and its implication recognized. What we need now are new paradigms and perspectives for linguistic and pedagogical research and for understanding the linguistic creativity in multilingual situations across cultures" (1985, p. 30). While this call for new paradigms and perspectives is over 30 years old, the conflicting notions

of a fixed Standard English code and a flexible international and intercultural communication present considerable challenges for researchers and teachers alike.

2.1.8   Speaking with an accent

Accent is defined as the pronunciation, stress patterns, intonation and rhythm of speech (Y. Kachru & Nelson, 2006). Like the previously discussed concepts of native speaker and Standard English, accent also carries with it controversies regarding acceptability, superiority, and desirability. Literature on accented English (Munro & Derwing, 1995; Munro, Derwing, & Morton, 2006; Piske, MacKay, & Flege, 2001) is focused on non-native speakers in an English-speaking community, and the perceptions and implications of accented English. Such discussions regarding accented speech raise the question: accented from *whose* perspective? One has to ask exactly what manner of pronunciation constitutes a foreign accent, and to what degree such a perceived accent might differ from what is expected by a native speaker.

The issue of foreign accents is largely irrelevant "…since everybody has a pronunciation of their language, everybody has an accent." (Bauer, 2002, p. 3). It's part of our identity as both an individual and a speaker, revealing to others a little about who we are and where we might be from (although with postmodern day migration at an all-time high, even this is becoming increasingly difficult). Accent has as much bearing on the concept of intelligibility as does the concepts of Standard English and native speakers. The only issue at hand is intelligibility, and the degree to which one is intelligible to a wide range of EIL users, such as at an international conference or on the floor of the United Nations.

2.1.9   Contrasting language users and learners

This thesis will be making a clear distinction between language *users* and language *learners*. A language user is an active speaker with "…developed and stable linguistic…interactional competence." (Firth, 1996, p. 239). A language learner, however, is presented in the literature in terms of *foreigner talk* or *interlanguage talk*, and are portrayed as having "…incomplete or deficient communicative competence, putatively striving for the 'target' competence of an idealized 'native speaker.'" (Firth, 1996, p. 241). Assumptions made of language

learners are critiqued in traditional second language acquisition literature by McKay (2002) and Sridhar & Sridhar (1994), who observed the following description of language learners in the literature:

1. The learner's target is native-like competence
2. Learners have enough venues of input to acquire this native-like competence
3. The functions served by the language in the local community are not studied
4. The learner's first language is considered in terms of how it interferes with English production rather than facilitating it in a multi-lingual setting

The functional purposes for which most EIL users need English do not require native-like levels of fluency. In addition, their fellow interlocutors are statistically more likely to be other non-native speakers, according to the demographic changes in English language use discussed in section 2.2.1. EIL users are able to function effectively in the vast majority of conversations by collaborating and negotiating meaning fluidly (as opposed to following Standard English forms).

A language user, as defined by the Common European Framework of Reference for Language (*Common European Framework of Reference for Languages*, 2001, p. 28) considers a language user to be:

...a person who as an individual and as a social agent develops a range of competences, both general and in particular communicative language competences. They draw on the competences at their disposal in various contexts under various conditions and under various constraints to engage in language activities involving language processes to produce and/or receive texts in relation to themes in specific domains, activating those strategies which seem most appropriate for carrying out the tasks to be accomplished.

Terms such as *learners*, *foreigners*, and comparisons to native speaker standards serve only to divide and compartmentalize people back into a circular models of language users, notions of us and them, or correct and incorrect forms of communication. The focus should be concerned with language users, and how effectively they are able to wield the English language for successful communication in international contexts. These contexts are proliferating the postmodern era, fueled by both globalization and the opening up of borders to the flow of people and goods, making successful international communication increasingly important.

2.2   Demographic change in a mobile world

The world has undergone considerable changes in a relatively short time. Globalization has brought people together for a variety of reasons: economic trade, military alliances, scientific research, and even entertainment. English has been the common language that has cemented these relationships, an inevitable result of its spread around the world. While the impact of this spread is still being assessed, the sheer number of English users paints an interesting picture for the future of the language. The following section will briefly review the spread of English from a colonial language to an international language, and argue that the ownership of the English language has slipped the leash of its colonial masters. Moreover, rather than just using English to communicate with native speakers, people from everywhere have appropriated English as a tool to serve their own purposes both intra-nationally and internationally. This section will demonstrate the need to develop a truly international approach to language proficiency testing, which is brought into the discussion in section 2.3.

2.2.1   Language explosion

In the past four centuries, Jenkins estimates the number of English language users has "...increased from a mere five to seven million speakers to somewhere between one-and-a-half and two billion." (2003, p. 2), which represents a nearly fiftyfold increase. This growing expansion of English language use was further cemented following World War II by the rise of another English-speaking superpower, the United States of America. By the time of the

British Council's 50ᵗʰ anniversary conference in 1984, Quirk estimated that there were approximately 300 million native speakers of English plus another 400 million using it as a second or foreign language (Quirk et al., 1985).

Today English is an official or semi-official language in over 60 countries, and has been given a prominent place in another 20. Crystal (2010) observes that English is the dominant language of books, newspapers, air traffic control, academia, business, sport, advertising, music, and diplomacy. Crystal goes on to report that in any given year, up to 1 billion students are learning English worldwide.

Early claims by Beneke (1991) that an estimated 80% of spoken English interactions involve non-native English speakers is frequently cited (see Gnutzmann, 2000; House, 2002; Jenkins, 2005; Seidlhofer, 2004). Unfortunately, Beneke provides no basis for the estimate, no data, and no reference. However, as data became available from India, China and the former Communist Block nations by the turn of this century, more recent and more accurate estimates are made possible. Early estimates place the number of English language users at near the two billion mark (Crystal, 2008; Graddol, 2003). More recent data from the British Council (2013, p. 2) refines earlier claims by declaring that "English is spoken at a useful level by some 1.75 billion people worldwide…" but fails to define what is meant by 'a useful level'. Nonetheless, the claims by Graddol (2003) and Crystal (2012) that English first-language users are outnumbered by English second-language users by a ratio of three or four to one remains unchanged. While such large estimates can and should be taken with a fair degree of caution, the sheer pace of this upward trend is difficult to deny. The result of this growing trend, catalyzed by the forces of globalization, has resulted in a power shift within the English language. Claims by the British Council (2013) that English is the "…the UK's greatest gift to the world" might be somewhat of an overstatement given that the language was thrust upon millions, who have nonetheless now appropriated English for their own purposes.

2.2.2   The ownership of English

As a result of these demographic changes, the ownership of English has been questioned. The language ownership issue was raised as far back as the early 1980s, with the launch of the flagship journal *World Englishes* (edited by Kachru and Smith). The issue was raised again by Widdowson (1994) and Norton (1997), who continued the argument that English belongs to everyone. The spread of English over the past few centuries has made the language a global commodity, and resulted in a number of varieties in local speech communities and countries, such as Australia, New Zealand, India, South Africa, and Singapore to mention but a few. The biggest boom in English users comes from countries with no historical or colonial link to the language, countries such as China, Russia, and Indonesia, and Brazil. As noted in section 2.2.1, the language explosion of English

> ...has taken an interesting turn: the native speakers of this language seem to have lost the exclusive prerogative to control its standardization; in fact, if current statistics are any indication, they have become a minority. This sociolinguistic fact must be accepted and its implication recognized. What we need now are new paradigms and perspectives for linguistic and pedagogical research and for understanding the linguistic creativity in multilingual situations across cultures. (B. B. Kachru, 1985, p. 30).

This sentiment is echoed by Graddol who, in his work for the British Council on the future of English, concludes that "Native speakers may feel the language 'belongs' to them, but it will be those who speak English as a second or foreign language who will determine its future"(1997, p. 10). This future is heavily influenced not only by the number of English-language users, but also by their migration around the world. The following section discusses the recent increase in migration and its impact on the need for migrants to seek certification for language proficiency skills as a stepping-stone to better opportunities sought abroad.

2.2.3   International migration, tourism, study, and business

In the forty year period between 1960 and 2000, international migration has doubled to 175 million, or roughly 3% of the global population (Graddol, 2007). This unprecedented movement of humanity is facilitated by affordable commercial air travel, the opening of previously closed markets such as the former Soviet Union, Eastern Europe, and China. People on the move include refugees, immigrants, international students, business people, migrant workers, and even asylum seekers. Migrants seek better opportunities as skilled workers not only in countries where English is the first language, but wherever the labour market demands. According to more recent information from the United Nations Department of Economic and Social Affairs Report (2016), there are a total of 244 million international migrants worldwide, with 47 million of them living in the US and over 160 million living in the EU and Asia. With English as the primary language of the US, and official working language of both the European Union and ASEAN, the linguistic implications for migrants cannot be ignored.

Regarding the situation in the European Union, Graddol remarks that "…it seems that the function of English is changing…European citizens learn and use English in order to communicate with European nationals from *all countries* and not just with native speakers" (Graddol, 2003, p. 160 emphasis added). Two results of mass migration are that major population centers are becoming increasingly multilingual (e.g. London, New York) and that increasing numbers of people are using English as a shared common language in regions like the E.U. and Southeast Asia.

For the same reasons mentioned at the start of this section, the tourism industry has seen unprecedented growth, with the number of international travelers expected to reach 1.18 billion by the year 2020 (Harris, Griffin, & Williams, 2002, p. 25). These numbers are confirmed by Graddol (2007, p. 29), who cites information from the World Tourism Organization that in 2004, nearly three-quarters of international travellers were tourists going from one non-English-speaking country to another non-English-speaking country (Figure 2.1).

Figure 2.1: International tourist encounters in English
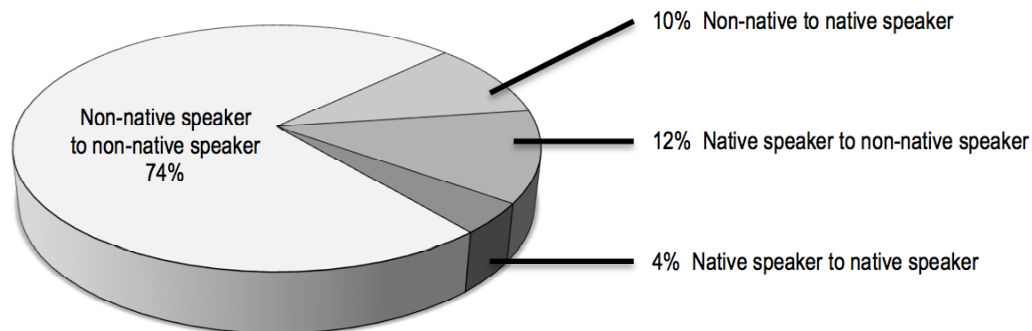
(Graddol, 2007, p. 29)

In addition to increasing migration and tourism, students are also on the move with trends showing that 46% of the 2-3 million international students are enrolling at English-speaking destination countries such as the US and the UK (Graddol, 2007). More recent estimates by the German-based International Consultants for Education and Fairs suggest that up to 5 million students enrolled overseas, which represents a tripling of international student enrollment since 1990 (ICEF Monitor, 2015). Educational institutions in English-speaking countries aim to attract overseas students to generate revenue, while the overseas students clearly place great value in obtaining a degree at a prestigious western university. This makes the English language a major investment for hundreds of thousands of international students each year

International students are well aware that language skills (in addition to numeracy and computer literacy) are critical skills for the international workplace. An international 2012 survey of over 500 major international companies supported this notion, with more than two-thirds of business executive respondents indicating that they consider English language knowledge essential for international success (Economist Intelligence Unit, 2012).

Without a doubt, speaking English has become an extremely useful skill for migrants, tourists, students, and business, serving as the metaphorical grease that facilitates this movement of humanity, or *transcultural mobility* (Appadurai, 1996). More than ever before, people from different linguistic and cultural backgrounds are able to meet and communicate. While such communications

26

are largely between non-native speakers using EIL, there are a number of other linguistic demographics are hard to ignore. Graddol's (2007) predictions for the coming decades also include:

- The number of people learning English is increasing.
- Their average age is decreasing.
- National ethnic and linguistic profiles are changing.
- Global migration is higher than ever before.

These unprecedented levels of human migration create two closely linked needs: English language skills and the certification of English language skills. In other words, people ranging from migrant works to business people need English to communicate to wider audiences. In cases involving an employer, educational institution, or immigration, a certification of English language abilities is often required. This usually takes the form of an officially recognized English language test, which has led to the development of several major English language testing bodies. The following section argues these testing bodies are taking full-advantage of the booming demand for English-language skill certification, making the testing business a major industry in its own right.

2.3 Challenging times for English language testing

Dealing with issues like English as an international language, world Englishes, and the ownership of the language is no easy task for the major English language proficiency testing companies. There have been calls for testing companies to accommodate the variety of language users and to recognize that excellence in English is no longer an inner circle native speaker prerogative (Canagarajah, 2006; A. Davies, Hamp-Lyons, & Kemp, 2003; Elder & Davies, 2006; Hu, 2012; Jenkins, 2006c). This section will introduce the three main English-language testing institutions and the power they hold as gatekeepers. Next the literature calling for language testing changes to be made will be reviewed and connected to the need for a fresh approach towards testing English in the postmodern era of international communication.

2.3.1 The big three language tests

English has become a skill – like numeracy or computer literacy – for occupations with an international orientation. Proficiency skills are usually broken down into speaking, listening, reading, and writing. There are several English-language tests on the market, but the top three major English language proficiency exams are:

- The test of English as a foreign language (TOEFL iBT)
- The test of English for international Communication (TOEIC)
- The International English Language Testing System (IELTS)

These three major tests are large-scale high-stakes English language proficiency tests. According to their respective websites, these three tests are administered in 130 to 150 countries to millions of test takers annually. As high-stakes, the IELTS, TOEIC, and TOEFL iBT can and do have a considerable impact upon the future lives of test takers. Such tests often serve as gatekeepers, certifying candidates who require the approved English-language qualification for life-changing events. Examples include admission and placement in tertiary institutions, qualifying for scholarships, opportunities for employment/advancement, acquiring visas for immigration purposes, and as part of medical licensing-body qualifications. These requirements, opportunities, and qualifications have been a driving force for a boom in demand for language testing services. This is exemplified by the IELTS claim that 2013 was the 18th year of consecutive growth in numbers of test takers ("IELTS Media Centre 2013," n.d.).

As a result, English-language proficiency testing has become a big business. In the U.K. alone, English-language education is recognized as a lucrative growth industry, so much so that on a 2005 visit to Beijing, British Chancellor Gordon Brown noted that the Chinese thirst for English presented a huge opportunity for Britain, a country then boasting a £1.3 billion English-teaching industry (Power, 2005). Maintaining the inner-circle prestige of English, maintaining this ownership over the language, is very big business indeed. Language testing, an inevitable outcome for most language courses, is

also big business. Table 2.3 shows the number of candidates who take these English-language proficiency tests, and the amount of money spent by test takers who require certification by these gatekeeping-testing organizations, and is an estimate based on publically available data gathered from their respective websites.

Table 2.2: Test fees for the big three language tests

Test	Average cost of test (USD)	Estimated number of candidates annually	Total
TOEIC	$85	5 million	$425 million
TOEFL iBT	$190	2 million	$380 million
IELTS	$160	2.5 million	$400 million

Sources: ("IELTS Media Centre 2013," n.d.; "TOEFL iBT: About the Test," n.d.; McCrostie, 2009).

Clearly the testing industry is heavily invested in their product. Likewise, test takers are equally invested in studying for these high-stakes tests since results will have a significant impact on the future lives. This raises the question of why testing companies would be interested in a different approach (in accommodating EIL) when their product already enjoys increasing demand worldwide as it stands. For listening tests, the question is also raised about which varieties to include and why. With millions of test takers preparing for these high-stakes tests each year, such decisions are critical, and will have considerable washback effect on language instruction, teacher hiring choices, and test preparation. If the inner circle bias is removed (or reduced) in the TOEIC, TOEFL, and IELTS tests, English-language learners would suddenly be exposed to a much wider selection of English varieties, and English native speaker teachers might suddenly find fewer employment opportunities in the English-language job sector.

In addition to the inner circle bias in these high-stakes language tests, there is also the underlying assumption that the test takers are seeking to live and work in an English-speaking country. The websites of IELTS, TOEIC, and TOEFL emphasize the fact that results from their tests are accepted by tertiary institutions and immigration organizations in the U.S., U.K., Canada, Australia, and New Zealand. The TOEFL iBT test is advertised as "...an important part of

your journey to study in an English-speaking country." ("IELTS Media Centre 2013," n.d.; McCrostie, 2009; "TOEFL iBT: About the Test," n.d.). However, the case was made (in section 2.2.3) that people are migrating at an unprecedented rate, not just to English-speaking countries, but wherever opportunities for a better life exist. The IELTS website also brands itself as a doorway to English-speaking countries, as highlighted in Figure 2.2. And while the IELTS test makes efforts to avoid cultural bias, it is exclusionary in terms of which varieties of English are acceptable, stating on its website "All standard varieties of native-speaker English, including North American, British, Australian, and New Zealand English are accepted"("IELTS - Why accept IELTS scores?," n.d.). In a single sentence, it is made quite clear that only the English varieties from a handful of inner circle countries are considered appropriate or standard for testing purposes.

Figure 2.2: Inner circle assumptions from the IELTS website

Australia
Achieve your dreams to go to
Australia with IELTS

Canada
Open doors to Canada with IELTS

UK
Live, work and study in the UK
with IELTS

Source: ("Home of the IELTS English Language Test," n.d.)

The British Council, as a major stakeholder in the IELTS test, has a vested interest in strengthening the position of the UK in the international education community, and thus "serves both to deliver and reinforce discourses which support native-speaker language norms" (Hall, 2010, p. 326). Attempts to accommodate non-natives by IELTS has so far been limited to hiring proficient non-native speaking examiners for the oral and written test components (Taylor, 2002). Hu notes that similar attempts to involve non-native speakers in the TOEFL test was "a limited attempt to accommodate to non-native speakers of English without toppling the dominance of native-speaker norms in the tests" (2012, p. 130).

2.3.2 Calls for change in language testing

In order for change to occur, two major concerns must be addressed. First, the inner circle/native speaker/Standard English bias needs to be overcome – not just in testing, but also with regard to student perceptions of native speaker prestige as well. The second concern is the acceptance of accommodations that should be considered with the ultimate goal of creating a truly international test of English. These two points (discussed in the following section) underline the two major themes underlying calls for change in language testing.

2.3.2.1 *The inner circle bias in language testing*

Traditionally, English-language testing has focused exclusively on the two main 'standard' varieties: American and British English (Davies, 2009; Hamp-Lyons & Davies, 2008). However, the combined populations of these two countries account for less than twenty percent of the estimated two billion people studying and using English as an international language today, making the case that the American and British standards are not a representative sample. Jenkins (2006c, p. 44) maintains that if English is to be used globally to facilitate communication among all countries, then:

> ...there is nothing international about deferring to the language varieties of a mere two of the world's Englishes, whose members account for a tiny minority...nor is there reason to suppose that the study of British or American English will promote international understanding.

Anyone in favour of standardized English testing is likely to disagree, since the challenge of codifying EIL is nearly impossible; by definition it is fluid and dynamic. Standard English, however, is well codified and reasonably fixed in place. For this reason, major English language testing organizations cling to standard American/British English despite the changing linguistic environment around them.

Davidson concludes "There is a well-established and legitimate concern that large, powerful English language tests are fundamentally disconnected from the insights in analysis of English in the world contexts. These exams set forth linguistic norms that do not necessarily represent the rich body of English varieties spoken and used in contact situations all over the world" (2006, p. 709). Davidson has held this opinion since 1994, when he acknowledged the imperial power of major international tests of English, "…tests which are clear agents of the English variety of the nation where they are produced. These tests maintain their agency through the statistical epistemology of norm-referenced measurement of language proficiency, a very difficult beast to assail" (Davidson, 1994, p. 120). Similar arguments about the bias in language testing have been made (see Canagarajah, 2006; Davies et al., 2003), reinforcing the idea that large language testing institutions in the U.S. and U.K. seek to maintain their monopoly on the growing global market of test takers. This monopoly is upheld by English native-speaking countries, leading Rampton to conclude that "…the supremacy of the native speaker keeps the U.K. and the U.S. at the center of ELT [English language teaching]…" (Rampton, 1990, p. 98).

However, changing demographics and the rapid pace of globalization has led to a "fundamentally changed landscape of EIL [which] requires a critical examination of the established assessment practices" (Hu, 2012, p. 123). While this critical examination has been ongoing for over thirty years, at present we have yet to see concrete proposals for assessing EIL competences. Instead, within the literature one can only find a number of useful suggestions and accommodations recommended for future English language tests. These accommodations are considered in the following section.

2.3.2.2 *Making accommodations for EIL in testing*

In order to move away from the strict adherence to native speaker norms, a series of accommodations can be made to integrate a broader approach to EIL. While Standard English is firmly codified – and thus easier to test – an EIL approach must answer the challenging question of what EIL abilities entail. Nunn (2011) synthesizes several articles on EIL and proposes what he regards as a list of EIL competences (to be discussed in section 3.1.2). This list includes a

general knowledge of world Englishes, the ability to operate in a number of contexts and speech communities, and the ability to apply learning experiences from one context to another. Competent EIL users possess skills that allow them to be more prepared for unexpected or unpredictable intercultural contexts. They can compensate for the gaps in their knowledge, and adapt to the listener as a means of increasing intelligibility. Likewise, EIL users are active listeners, and share in the responsibility for mutual understanding.

Nunn goes on to argue that "…EIL communication is always intercultural communication" (2011, p. 25) and identifies a number of teaching objectives that can be derived from Byram's (1997) model of intercultural communicative competence. The following list comprises teaching objectives, which can also be considered EIL intercultural characteristics (or abilities). As summarized by Nault (2006), this list includes:

- Attitudes – curious, open-minded and flexible
- Knowledge – of different social groups and their practices
- Skills of interpreting – of a document or event from another culture
- Skills of discovery – the ability to learn a new culture and adapt in real-time
- Critical cultural awareness – the ability of users to critically evaluate perspectives and practices of not just other cultures, but their own as well

In the case of EIL, the role of culture is unique in that there is no single culture that can serve as a basis for teaching cultural content, especially an Inner Circle country (McKay, 2002, p. 82). McKay goes on to promote EIL as a platform for intercultural education by asking language learners to reflect on how cultural knowledge might affect interactions with others. This notion builds on earlier concepts by Kramsch (1993), who suggested the following two goals for culture in the language classroom:

1. Sphere of interculturality – learners should reflect on their own culture in addition to a target culture.

33

2. Teaching culture as difference – recognizing that national identities constitute a variety of characteristics (such as region, ethnicity, social class), and are not monolithic.

In addition to reflecting on one's own culture, a general attitude of openness to other cultures and forms of speaking is also an integral element that facilitates communication and intelligibility in international contexts. In addition to this attitude, Canagarajah (2006a) identifies a sociolinguistic sensitivity and language awareness in EIL users, as well as a healthy dose of patience, tolerance, and even humility in EIL communication. He repeats the call to create some distance from native English speaker standards, arguing for the need to view English competence from the perspective of the participants of a conversation. This call for the involvement of EIL users in any kind of new assessment paradigm is gaining ground in the EIL literature (Elder & Davies, 2006; House, 2003a) and addresses the issue of the *comparative fallacy* (Cook, 1999) whereby deference is given – perhaps unfairly – to the prestigious native English speaker varieties.

Canagarajah calls for prioritizing communicative competence over form, and recognizes the primacy of "…a range of other skills, abilities, and awareness that enable multilingual speakers" (2006a, p. 928). He concludes that language awareness, strategic competence, and pragmatic competence are essential for understanding EIL proficiency.   This alignment of linguistic resources and the collaborative nature of communication suggest that the skill of negotiating meaning in EIL communication is critical. When these skills have been found lacking (i.e. communication breakdown), it is often the native English speakers who have failed to negotiate meaning, since they tend to regard their own norms as universal (House, 2003a).   In this case, we are reminded of the comparative fallacy, and are left to wonder why such an inflexible standard has become the benchmark.

Other accommodations include the use of non-native English speaker judges in order to avoid biases in assessing communicative competence. While Elder and Davies (2006) argue that this may be a form of positive discrimination, a test of EIL needs to be just that; international. Without the involvement of

34

expert users of EIL, how can the instrument hold a valid claim to be a measuring the international use of the language? It is these users who have the language awareness, the sociolinguistic sensitivity, and the experience; they know what is required to use English successfully in a wide variety of contexts.

Canagarajah (2006a) has argued in favour of the use of two raters from different English-speaking communities, one assessing strict linguistic criteria and the other focusing on a holistic, global grading of communicative competence. The latter allows a measure of intelligibility and fluency that can rise above grammatical correctness, which makes sense since EIL users make allowances for such errors, and value meaning over the adherence to native speaker norms. Rather than adhering to strict narrow norms, a candidate should possess the ability to *shuttle* between speech communities (Canagarajah, 2006a) in a test of EIL.

Jenkins' (2000, 2002, 2006c) tenets of mutual intelligibility and her challenging existing pronunciation standards (i.e. Received Pronunciation and General American) are hallmarks of a new approach needed with regard to EIL. Elder and Davis (2006) agree, stating "Strict adherence to native speaker norms of correctness are arguably unreasonable…given that successful communication does not depend on them…" (p. 288). Elder and Davis call for a relaxation of the standard norms, and propose four creative approaches to EIL which depend on how the participants and their interaction are defined. While their suggestions for assessment are noteworthy, they are also complex and leave researchers with no clear idea how to proceed.

2.4   Language policy in ASEAN and the EU

Over half of the test takers in this research are from ASEAN countries. Currently, the ten members of ASEAN include Brunei, Burma, Cambodia, Indonesia, Laos, Malaysia, the Philippines, Singapore, Thailand, and Vietnam. Kirkpatrick (2010) remarks that although ASEAN members have considerable diversity amongst members, the founding Bangkok Declaration of 1967 makes *no mention* of any working language of the association. Instead, English has in fact been adopted as the working language of ASEAN – not through lengthy

debate, but rather through the Malay concepts of *musyawarah* (dialog) and *muafakat* (consensus) (Curley & Thomas, 2012).

Such a consensus likely originated from the historical role colonialism played in the region, creating the modern-day situation for English in ASEAN shown in Table 2.3. While the official role of English in the Philippines and Singapore is well established, English is 'recognized' in two additional countries for historical and institutional (i.e. higher education, law, diplomacy, trade) reasons.

Table 2.3: English language status within ASEAN countries

Foreign language	Official language	Recognized language
Myanmar	Philippines	Brunei
Cambodia	Singapore	Malaysia
Indonesia		
Laos		
Thailand		
Vietnam		

The role of world Englishes can be seen in the differing negotiation styles within ASEAN. While Thambipillai's (1992) research in this area does not comment on whether or not official language status in a member nation created an advantage in negotiations, he did make the following general observations on the speaking style of negotiators:

1. Singaporeans tend to be more articulate, with facts and well-substantiated arguments that support their claims.
2. Philippine negotiators are less specific with their demands.
3. Malaysians less open and more rigid in negotiations.

It is unfortunate that more recent research on negotiation style is unavailable, as it would be interesting to observe any changes or development over the decades. One might expect that the English language policy of ASEAN has created a level-playing field where all member states can communicate effectively in an ideal EIL environment. However, with such a diverse international organization, differences in English varieties occasionally became

problematic. One unintended outcome of making English a working language of ASEAN was the regional differences in style and grammar. Ong Keng Yong, an ambassador-at-large from the Singaporean Foreign Ministry once recalled:

> There is a multiplicity of grammatical and spelling preferences originating from the particular usage of English in each member state. At the Charter drafting sessions, such differences delayed a quick consensus on wording and many formulations had to be put in square brackets (to be revisited later)
> (Yong, 2009, p. 113).

Kirkpatrick (2010) is surprised that in his extensive research into the role of English in ASEAN, not one of nearly two hundred chapters of ASEAN readers discusses language issues. This point highlights the uncontroversial decision to use English as the working language and the confidence local leaders appear to have with their own local languages. The de facto adoption of English as the working language of ASEAN contrasts greatly with that of the EU, which had considerable debates around the issue of official and working languages (cf. Phillipson, 2003; Wright, 2007).

While the European Union (EU) has just over 20 official languages, only three have the higher status of working languages (English, French, and German). Daily work in the European Commission is carried out in one of the three working languages, which is most often English. Documents discussed at the plenary sessions of the European Parliament are translated into the so-called major working languages of the EU first and the minor languages later. Unfortunately, the delay means that delegates who use the minor EU languages may not be able to properly read these documents and prepare for meetings in advance.

Kirkpatrick (2010) sees the likelihood of ASEAN adopting a EU multilingual style model as highly unlikely, especially given the delays caused by and massive budget needed for translation and interpretation services. Tagliabue (2006) estimates that the EU spends about US $1.3 billion dollars annually to employ over 3,000 staff to translate and interpret the thousands of

pages of treaties and agreements passed by EU institutions. ASEAN nations, on the other hand, appear satisfied to avoid this expense and focus on direct dialog in English as the sole working language.

But in order to achieve higher levels of fluency in English and a more level linguistic playing field in ASEAN dialogs, most member nations have steadily been increasing the amount of English taught in schools over the past decade. Presently, English is being taught as part of the primary curriculums of all ASEAN member nations. In the case of Singapore, English is the medium of instruction, while in the Philippines English is used to teach math and sciences.

2.5   Diplomacy and EIL

The test takers involved with the development of a test of EIL are described in detail in Section 4.1.5, however it is suffice to say that they are representative of the government officials working in ASEAN and the EU. As mid-level officials, English language skills are career skills necessary for daily work in various ministries in their respective home countries. More than half of the test takers in this thesis are from ASEAN member states, while the remainder are from other developing Asian and African countries. These officials work in ministries such as Agriculture, Forestry, and Fisheries, Commerce, Education, Finance, Foreign Affairs, Industry, Justice, Mines & Energy, Post & Telecommunications, Rural Development, and Strategic Planning. Work in these ministries involves EIL to communicate with regional counter-parts in organizations such as ASEAN, the African Union, Asia-Pacific Economic Cooperation, and global counter-parts in the World Health Organization, the World Bank, and of course the UN (to list but a few examples).

The test takers are ideal participants in this research for three reasons: first, they are typical of the working professionals who use EIL in international contexts primarily with others like themselves. They are not trying to reach an unachievable native speaker standard of proficiency and correctness, but rather aim for mutual intelligibility and clarity of message. Second, these government officials are interested in the function of EIL, and its usefulness in getting things done (e.g. reaching agreements, communicating national intent, and accessing world knowledge in English). Finally, government officials are ideal participants

because of the consequences of misunderstanding. These three points are discussed in greater detail below.

First, the notion of mutual intelligibility is linked closely with international intelligibility, and is based on the premise that variation across Englishes is natural. It runs counter to the idea that the native speaker (see section 2.1.6) is the best model for intelligibility, and instead argues that exposure to different varieties of English is the best way to promote mutual intelligibility. The government officials participating in this research already have this exposure through work with regional counter-parts. However, upon their arrival in New Zealand (to improve English language skills) they need time to become familiar with the local accent.

Second, the function of EIL is one of the underlying reason for its global success: the simple fact that English is the official or semi-official language in over 60 countries (and recognized another 20), which taken together represent 41% of the 193 nations currently represented in the UN. Within the EU, the perceived functionality of English remains high. Two-thirds of Europeans consider English a useful additional language to learn, a fact which unfortunately comes at a high cost to the popularity of the German and French languages (European Commission, 2001, p. 85; 2012, p. 69). Nonetheless, government officials are well aware of the utility of English language skills when operating at the international level, and as demonstrated by the test takers in this research, are willing to invest heavily in developing their English proficiency.

For countries where English is historically a foreign language (e.g. China, Japan, Russia), English serves both a passive and an active function. While English may not be widely spoken in several countries it serves a passive function as window on the outside world. Scholars, scientists, doctors, engineers, and other professionals use English to access the most current information in their field of expertise, if even only online. In an active sense, the function of English is used as a platform for communicating messages to the English-speaking world. This may be in the form of English protest signs for the world's media or in the form of a national English-language news website. Most countries in the world have an online English news source, allowing one to

acquire different perspectives on regional and international news (helping to broaden mutual understanding).

Third, the consequences of misunderstanding between nations may range from minor perceived insults to major hostile or military actions. Rhetorical speech, implied meanings, misperceptions of intent, or even mistranslations can have tragic consequences. The following are two examples of such consequences, one involving a mistranslation from one language to another, and the second example a case of one side misinterpreting the intentions of the other.

First we will quickly look at a case of a catastrophic translation error dating back to WWII. As the Allies nations concluded the war in Europe in the spring of 1945, attention shifted towards Japan. Pressure mounted for Japan to unconditionally surrender at the Potsdam Conference to avoid a costly land invasion of Japan by the Allies (led by the US). The Japanese government said nothing while they considered their options. When reporters pressed Prime Minister Kantaro Suzuki for an answer in Tokyo, he eventually replied "mokusatsu". In what linguists have called the world's most tragic translation (Kawai, 1950), the Allied leaders of World War II misinterpreted Japan's response to the call for surrender. Instead of translating the Prime Minister's remark to mean "no comment", the term was instead mistranslated as "not worthy of comment; held in silent contempt." International news agencies at the time widely reported that the Allies demand for unconditional surrender was met with the response "not worthy of comment", angering US officials and leading to the decision to drop the atomic bomb on Hiroshima on August 6th, 1945.

While this is an extreme example of a miscommunication gone horribly wrong, there are other cases in inter-government dialog whereby misinterpretation of intent led to disastrous outcomes. In the lead up to the 1990 Gulf War, US Ambassador Glaspie met Iraqi President Saddam Hussein in Baghdad on July 25th 1990 to determine his intentions and reasons for massing Iraqi troops near the Kuwaiti borders. Official transcripts show that President Hussein shared some strategic goals and Iraqi claims, after which he asked for the US opinion. The US ambassador responded with:

We [Americans] have no opinion on your Arab – Arab conflicts, such as your dispute with Kuwait. Secretary [of State James] Baker has directed me to emphasize the instruction, first given to Iraq in the 1960's, that the Kuwait issue is not associated with America. ("Excerpts from Iraqi document on meeting with U.S. Envoy," 1990)

Just the previous day, US State Department spokeswoman Margret Tutwiler was asked by reporters if the US was committed to defend Kuwait, to which she replied ''We do not have any defense treaties with Kuwait, and there are no special defense or security commitments to Kuwait.''(Sciolino & Gordon, 1990). Kuwaiti officials regarded a vague response from spokeswoman Tutwiler about the US supporting individual and collective self-defense in the gulf as too weak to deter an Iraqi invasion.

Of course this miscommunication of intent is far more complex than discussed here, and by no means the single cause for the outbreak of the 1990 Persian Gulf War. However it does highlight the need for intergovernmental communications to be carefully. Whether in a foreign language or shared language, the consequences of miscommunication can be disastrous, especially in times of heightened tension when major national interests are involved.

2.6   Making a distinction from existing work

This thesis aims to apply principles of language testing to the assessment of EIL, drawing knowledge from the fields of pragmatics, intelligibility, strategic language use, and speech acts. Rather than defining EIL as a traditional language code, EIL should be regarded as a *function*, or a set of resources one can employ to enhance international communication. EIL itself is a tool, a function or platform for negotiating a myriad of global tasks, from chatting online to solving the world's problems (climate change, global health epidemics, security alliances).

The following chapters are not concerned with work on a stabilized standard form or phonological core (Jenkins, 2000), which seeks to codify a minimum standard form of intelligibility. Jenkins' efforts to promote a minimum pronunciation standard did break new ground, but have attracted criticism

ranging from lowering pronunciation standards to promoting linguistic phonological imperialism. Rather than attempt to set any kind of standard, the aim herein is to employ testing accommodations (discussed in section 2.3.2.2) into a test of EIL abilities.

Likewise, a distinction can be made from corpus work defining EIL with the Vienna-Oxford Corpus of English (VOICE) project (Seidlhofer, 2001, 2004) and work on EIL pragmatics (Meierkord, 1996) recording how people from different L1 backgrounds use English. These approaches have also been criticized for simply documenting learner language development rather than building a communicative framework for international contexts. Rather than attempting to codify EIL in the traditional sense of language research, this work avoids the entire codification issue altogether by approaching EIL as a function and a set of skills that can be used to enhance successful international communication.

Another distinction concerns assessment. In the postmodern era, with the number of international communications in English far exceeding the number of native speaker interactions, EIL users make far better judges of what constitutes success. This new paradigm removes the native speaker from the center of traditional competency models, an approach that tended to view the foreigner as an outsider seeking entry into an English speaking country. In considering the following points, it becomes apparent that native speakers standards are no longer necessary:

1. The number of bilingual EIL users far exceeds the number of native speakers
2. EIL users tend to be more experienced with intercultural communication
3. EIL users are more able to anticipate and overcome challenges in communication
4. Standard English norms of correctness are unnecessary for most EIL users
5. Strategies (rather than norms) for successful communication are arguably more important in the postmodern era.

An attempt has been made to follow up on proposals for EIL accommodations made by Elder and Davies (2006), whose work was most influential in conceptualizing the assessment of EIL. They suggest avoiding the typical western topics or lexical items that are likely to be unfamiliar to global users of English, as well as a need for the involvement of EIL users as interlocutors in a speaking test, or in test item evaluation itself, to ensure the appropriateness of the test content. Next, Elder and Davies (2006) recommend penalizing only those errors resulting in miscommunication, placing a reduced emphasis on linguistic code, and assessing performance primarily on task fulfillment. Finally, they argue in favour of assessing a simulated interaction in an international context, and recognizing the potential importance of giving strategic competence (or conversation repair) priory over linguistic accuracy. While these accommodations have become potential principles for assessing EIL, Elder and Davies (2006) arguably do not go far enough in their proposals for assessing what they term *English as a lingua franca*. Their approach has the overwhelming aim to identify EIL as a concrete language code, and distinguish between contexts where English is the first, second, third, or other language of the participants. The approach taken by Elder and Davies ignores the fluidity of EIL and the fact that EIL users (often bilinguals) operate in a variety of contexts with both other EIL users and monolingual speakers. This creates a confusing situation and resists traditional classification and codification efforts. A better approach is to measure specific abilities that improve successful communication in English for a variety of trans-cultural contexts. This is the essence of the core construct of EIL, regardless of whether one is a monolingual native speaker or a bilingual EIL user.

2.7   Chapter summary

The goal of this chapter has been to clarify the key terms and issues surrounding EIL. First the overlapping terms world Englishes, English as a lingua franca, and EIL were explained and contrasted (to the degree that they can be 'pulled apart'). Second, in the postmodern world, the older circular models of English language use have been superseded by the unchecked spread of EIL, which defies the constraints of models and labels. Third, the case was

made that the idealized native speaker concept is "…rich in ambiguity"(Davies, 2003, p. 2) and that if "…there is no satisfactory characterization of the term 'native speaker', then it is foolish to accept the construct of native speaker as a model of competence." (McKay, 2002, p. 31). Likewise, a satisfactory characterization of Standard English is complex at best, while the notion of any ideal accent being associated with intelligibility is also moot. Next, the dramatic shift in the demographics of English language learners has challenged the ownership of English, and the ability of so-called inner circle countries to dictate the future norms of the language. This shift in demographics has resulted in a situation where international and intercultural contact is on the rise, meaning the need for flexibility and accommodation in English language use will also increase.

Rather than defining EIL as a language code, it is being defined as a set of skills for using EIL successfully in international contexts, skills that include an awareness and knowledge of other varieties of Englishes, and be able to access this knowledge readily. Second, EIL users need the ability to understand – and be understood – in many contexts, adapting their language to meet the needs of the interlocutor. This flexibility is common in many EIL users, many of whom can shuttle between their local speech communities and that of the L1 users. Third, given the variety of international contexts for use, a degree of intercultural competence is helpful in understanding a speaker's *intended* meaning rather than the *literal* meaning. The most important characteristics of EIL use are the abilities to negotiate meaning and to repair conversation in the event of breakdown.

Chapter 2 criticizes traditional English language instruction for its orientation towards American or British English and the orientation of preparing students for communication with native speakers and/or entry into an English first language country (Jenkins, 2012). This instruction and assessment does not meet the needs of millions who are more likely to be in communication with a much wider audience than just English native speakers users. Regarding the assessment of EIL, researchers (see Canagarajah, 2006a; Elder & Davies, 2006; House, 2003a) have suggested accommodations to address claims of bias against

so-called non-standard English variety speakers. These accommodations are summarized here:

- Rating scales refer to *expert user* of English rather than native speaker
- The involvement of two raters from different linguistic backgrounds
- The involvement of two candidates from different linguistic backgrounds
- Only errors resulting in miscommunication are penalized
- Mutual intelligibility receives priory over linguistic accuracy

Finally, Chapter 2 argues that while working language policies of ASEAN and the EU (as well as other international organizations) have made English language proficiency a must-have skill, government officials need not have a native-like grasp of the language. Instead, EIL serves a very useful *function* for the participants in this research as a means of getting things done.

The following chapter intends to fill the gap between these accommodations and the development and operationalization of a test of EIL abilities by outlining four theoretical constructs argued to be essential for successful international communication. Chapter 3 will identify these four constructs, and aim to provide a theoretical foundation for the fluid and dynamic nature of EIL in an attempt to measure abilities above and beyond basic language proficiency.

PAGE LEFT BLANK

Chapter 3: A conceptual model of EIL competence 3

The previous chapter outlined the key terms and issues in the EIL debate, without going into detail about what abilities contribute to successful use of EIL in international contexts. The role of Chapter 3 is to serve as the theoretical basis for a new test battery of English as an international language, taking into account the extensive and divergent body of knowledge on the components of communicative competence from an EIL perspective.

A top-down argument will be made for the abilities argued necessary for international communication by seeking to identify the characteristics of successful international communication, what abilities enhance it, and how it might be assessed. Two of the constructs deal with intelligibility: *receptive* intelligibility (listening), and *productive* intelligibility (speaking). The third construct addresses the intercultural communication component of EIL use, and the overlapping themes emerging from the intercultural competence literature. The final construct proposed to be a part of an 'EIL skillset' is strategic competence, or the ability to repair (possibility even enhancing) conversation.

3.1 Communicative competence

The past 50 years has seen several advances in the concept of *communicative competence* since the term was initially coined by Hymes (1967, 1972) in his critique of Chomsky's (1957, 1965) examination of *linguistic* competence, which Hymes felt was a narrow view of the language code (grammar, vocabulary, phonology, morphology, syntax, and semantics). Communicative competence was defined by Hymes (1967) as grammatical knowledge and social knowledge of how and when to use language appropriately. The following two sections will contend that this notion of *appropriateness* takes on a broader approach when EIL users are included. Hymes (1972) notion of communicative competence as knowledge of "…when to speak, when not, and as to what to talk about with whom, when, where, in what manner" (p. 277) becomes increasingly complex when interacting in international contexts with others from a variety of sociolinguistic backgrounds. This concept of appropriate language use raises the question of appropriate

language for whom? This section will lead into an answer for this question following a review of three earlier models (Bachman, 1990; Canale, 1983a; Celce-Murcia, Dörnyei, & Thurrell, 1995) and one modern approach (Bachman, 1990; Canale, 1983a; Celce-Murcia, Dörnyei, & Thurrell, 1995) to communicative competence.

3.1.1   Early approaches to communicative competence

While there have been several models of communicative competence since the work of Chomsky and Hymes over forty years ago. This section will focus on three of the major models commonly cited in the literature. Canale and Swain's 1980 model of communicative competence, for example, sought the ultimate goal of improving teaching methods and a "…more valid and reliable measurement of second language communication skills" (1980, p. 1). By reviewing the literature, they came to the definition of communicative competence as the *interaction* between grammatical competence (knowledge of grammar rules) and sociolinguistic competence (knowledge of the rules of language use).

Canale and Swain (1980) make a clear distinction between communicative competence and performance, but have no model for such a performance (or demonstration of competences in a real speech event). They suggest assessment needs to capture grammatical and sociolinguistic competence through some type of communicative performance, allowing participants to demonstrate their knowledge in action. Canale's 1983 model (see Figure 3.1) addressed this gap by including performance in the newer model under the term actual communication, meaning "the realization of such knowledge and skill under limiting psychological and environmental conditions…" (Canale, 1983b, p. 5). This separation of 'what you know' and 'what you can do' was further divided, as shown in figure 3.1.

Figure 3.1: Canale's (1983) model of communicative competence
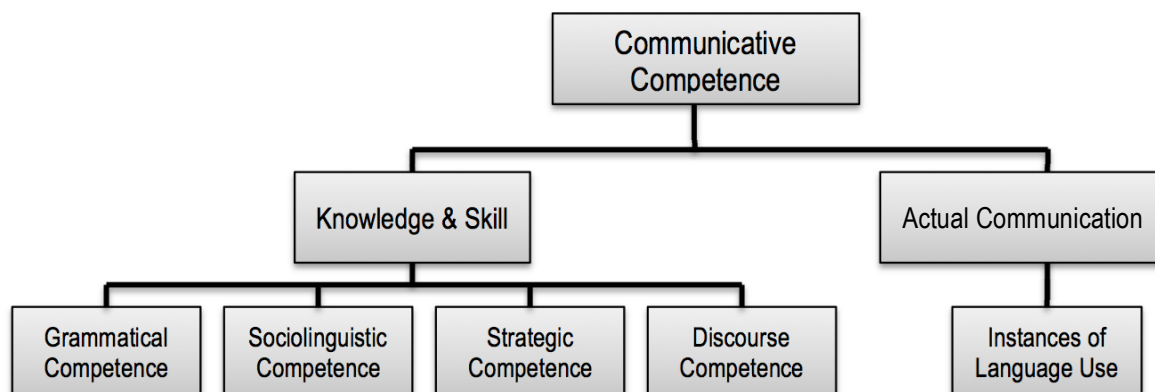
This model of communicative competence distinguishes between knowledge and actual performances, and defines strategic competence as a set of compensatory strategies for conversation repair. However, it appears to place as much importance on grammatical competence (i.e. knowledge of grammar rules) as it does on sociolinguistic competence (knowledge of the rules of language use). In EIL communication, language accuracy gives way to accommodation as speakers negotiate meaning "across linguistic and cultural boundaries" (McKay, 2002, p. 38) and in doing so, employ strategies to reduce ambiguity. In spoken EIL communication, the components of discourse competence, sociolinguistic competence, and strategic competence outweigh grammatical accuracy, since EIL is flexible and dynamic in nature.

Bachman's (1990) model of communicative language ability expands on Canale's (1983a)work by making further distinctions between what constitutes knowledge and what constitutes skill. He places strategic competence at the center of both his 1990 and 1996 model, arguing that *linguistic knowledge*, *knowledge of the world* – or at least society – and *context* are three major influences on the application of strategic competence. Bachman defines linguistic knowledge as a list of competences under two major categories: organizational (linguistic) and pragmatic (using language effectively). Of particular interest are the pragmatic skills of language competence, which include illocutionary competence (functions of language use) and sociolinguistic

competence. These are the types of competencies skilled EIL users require on a daily basis when interacting with others in international contexts.

Celce-Murcia et al. 's (1995) work on communicative competence criticizes earlier work (see Bachman, 1990; Bachman & Palmer, 1996) for conceptualizing communicative competence as a theoretical tool for test development. Moreover, Celce-Murcia et al. (1995) criticize the type of content specifications developed by Bachman on "an ad hoc basis for syllabus design" (Fulcher & Davidson, 2007, p. 47) and aimed to produce a communicative competence model (see Figure 3.2) that could be unpacked and broken down so that the sub-components would serve as a content base in syllabus design. Celce-Murcia et al. (1995) define the five key competencies of communicative competence as socio-cultural, discourse, linguistic, actional, and strategic).

Figure 3.2: Celce-Murcia et al. (1995) model of communicative competence
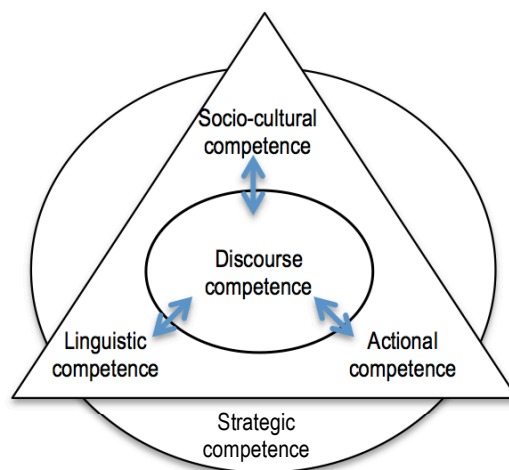
The 1995 Celce-Murcia et al. model and its constituent components are essentially a re-working of earlier models by Canale (1983b) and Bachman and Palmer (1996), giving different names to components or merging/splitting other components to create the model shown above (Figure 3.2). However, some of the sub-components of discourse competence detailed by Celce-Murcia et al. include the skills of substitution/ellipsis, turn-taking and repair of conversation, are directly relevant to the key skills needed for EIL users. Other useful components for EIL users are socio-cultural competence (for adapting to different ways of realizing speech acts, or performative utterances), actional

competence (for applying knowledge to achieve outcomes), and strategic competence (for repairing conversation and maintaining clarity).

Much later, Fulcher and Davidson note that "Writers with a particular interest in the social context of speech and how communication is understood and constructed in a specific context have concentrated on developing the concept of *interactional competence*." (2007, p. 49). This concept was first proposed by Kramsch (1986), who believed that since real-time communication is co-constructed, individuals themselves cannot bear sole responsibility for talk. In other words, what is said in conversation depends largely on what has previously been said, by whom, and for what purpose. Young and He (1998) argue that this kind of interactional competence is not a trait residing in an individual, making the testing of interactional competence a considerable challenge for language testers. With regard to this co-construction in testing, Weir appropriately raises the question: "How to factor this into or out of assessment criteria is yet to be established in a satisfactory manner." (2005, p. 153).

Fulcher (2010, p. 110) however, argues that performance *can* be identified as an individual competency rather than a context-dependent phenomenon, going on to state:

> Sociocultural explanations of constructs confuse the ability of an individual to recognize the contextual constraints and freedoms in communicating using available resources, with the actual communication itself. No one would wish to deny that each instance of communication…arises within a context and is dependent upon the variables at play. But this does not mean that the competence to engage successfully…is also newly generated in each instance of communication.

The traditional approaches are often oriented with a specific target language community in mind, meaning that notions such as pragmatic competence and sociolinguistic competence are anchored in these language communities. What is pragmatically appropriate language use in one language

community may not be the same as in another. Bi- or multilingual EIL users tend to shuttle between these communities where pragmatic norms of language use are less fixed and more co-constructed. In other words, the traditional models of communicative competence are relatively concrete for given language communities, much like the notion of Standard English. What EIL users need is a model that recognizes the diversity of contexts in which they operate, and the flexibility required to negotiate differences in how language is used. The following section acknowledges these abilities, and suggests new approaches for modeling communicative competence in the 21st century.

3.1.2 Recent approaches to communicative competence

Although Nunn suggests that a term like *communicative competence* has outlived its usefulness due to overuse in the literature, he does go on to define competence in broad global terms as "…the total available range of (multi-cultural) abilities, skills, knowledge and experience that can be drawn upon for any particular performance by an individual or group of individuals to address a real-world task or set of tasks." (2011, p. 6).

The theoretical foundation of EIL competency rests on two assumptions: first, the recognition that this thesis is not looking at a fixed language code (grammars, vocabulary, syntax) but rather at the *function* of EIL, or the use of language to complete various tasks in international contexts. This requires a second assumption, which is that a more holistic approach to competence is needed. This is at odds with traditional Chomskian linguistics and its ideal speaker-listener in a homogeneous speech community (Chomsky, 1965); a situation that is increasingly "…difficult to find in multiethnic industrialized urban societies."(Kramsch, 1997, p. 368).

These two assumptions are feasible under Nunn's (2007, 2011) model for communicative competence. Nunn (2011, p. 8) argues that there are five components of international communicative competence that operate in tandem whenever language is used: first, there is a *multiglossic* competence, enabling interlocutors to be sensitive to different identities while intelligibly communicating their own identity. In other words, to be familiar with other varieties and accents of English while being easily understood oneself. Second,

Nunn proposes that strategic competence is an essential element of two-way intercultural communication. Third, pragmatic competence is required to resolve differences in background knowledge, which may be embedded in culture or society. This may need additional training or awareness to develop as a skill. Fourth, intercultural competence for Nunn means the ability to adjust to the uncertainty and unpredictability of multicultural situations. Finally, linguistic competence – in at least one variety of English – is critical for expressing the above skills in terms of language.

  With regard to multiglossic communication and linguistic proficiency, it is arguable that Nunn's components of international communicative competence can be expanded to include intelligibility. International communication is immediately impaired if the speaker and listener lack the familiarity with each other's pronunciation or the linguistic competence (vocabulary, syntax, grammar) to arrange sentences intelligibly. In the event of communication breakdown, the use of strategies is essential to repair conversation and seek clarification if necessary. Finally, international communication goes hand in hand with a basic understanding of different cultures, and how miscommunication can result from these differences. The ability to adjust to multicultural contexts suggests that EIL communication can be made much more effective, making it another useful component of international communicative competence.

3.1.3   Comparing models of communicative competence

  At this stage, several (but by no means all) models of communicative competence from the literature have been reviewed and connected to communicative needs for EIL users. Naturally a degree of overlap can be expected among the various definitions and components (or competences). Table 3.1 summarizes the key components of these models. From this summary (and from filling the gaps in this table), the argument will be made for four large constructs that may serve as key competences for effective EIL communication.

Table 3.1: Comparing models of communicative competence

Researcher	Specific competences						
	Strategic	Discourse	Linguistic / Grammatical	Socio-linguistic	Inter-cultural	Pragmatic	Multi-glossic
(Hymes, 1972)		☑	☑	☑			
(Canale, 1983b)	☑	☑	☑	☑			
(Bachman, 1990)		☑	☑	☑		☑	
(Celce-Murcia et al., 1995)	☑	☑	☑	☑			
(Nunn, 2011)	☑	☑	☑		☑	☑	☑

The noteworthy features of Table 3.1 are the focus on linguistic or grammatical competency as a component of each model. Clearly this aspect is regarded as central to communicative competence by all of the researchers listed in the table summary. However, few models recognize the importance of an intercultural, pragmatic, or multiglossic component, revealing that these components have only recently been recognized as beneficial. The recent additions of intercultural and multiglossic competence may be an unintended byproduct of globalization, where more and more international and intercultural communication is taking place. Over the past half-century, models of communicative competence have developed considerably, perhaps a reflection of the relatively sudden demographic changes discussed in section 2.2.

It is interesting to note the evolution of theories of communicative competence, beginning with Chomsky's (1965) notion of linguistic competence. This model considers the native speaker as an *ideal speaker-listener* with intuitive sense of grammatical correctness. In the postmodern world, such a definition falls short, leading Kramsch (1997) to criticize Chomsky's concept of the ideal native speaker. As the models for communicative competences evolved, one can see the shift away from the native-speaker center (with his/her grammatical mastery) towards a more context-based, performance-oriented, intercultural model more suited to the realities of EIL users.

In other words, there has been a noticeable theoretical reversal away from inflexible concepts of communication standards towards concepts that

recognize the effects of globalization and argue that accommodation, strategic competence, and shuttling between varieties of language are more representative of communicative competence. While language proficiency is still a main feature of these models, increasing importance has been given to the ability to get things done (i.e. functional purposes) in language by demonstrating effective and appropriate skills, such as pragmatic/discourse abilities, intercultural competence, and the ability to repair conversation in the event of breakdown (strategic competence). This evolution of communicative competence models from Chomsky (1965) to Nunn (2007, 2011) is summarized in Figure 3.3.

Figure 3.3: The evolution of communicative competence models

Chomsky (1965)

Linguistic Competence

Hymes (1967)

Linguistic Competence — Sociolinguistic Competence

Canale & Swain (1980)

Grammatical Competence — Strategic Competence — Sociolinguistic Competence

Canale (1983)

Grammatical Competence — Strategic Competence — Sociolinguistic Competence — Discourse Competence

Celce-Murcia et al. (1995)

Linguistic Competence — Strategic Competence — Sociolinguistic Competence — Actional Competence — Discourse Competence

Nunn (2007, 2009)

Linguistic Competence — Strategic Competence — Multiglossic Competence — Intercultural Competence — Pragmatic / Discourse Competence

(adapted from Celce-Murcia, 2007, p. 43)

Figure 3.3 shows Nunn's (2007, 2011) addition of two components to communicative competence models, possibly marking the first time communicative models can truly claim to be international. The two components (intercultural competence and multiglossic competence) signal the shift away from inner circle native speaker-centered norms for communicative competence and towards a more pluralistic and inclusive model that recognizes EIL as a wider language of communication. Within the intercultural competence component, Nunn (2011) proposes that more than just knowledge of other cultures is needed for successful communication; it is the ability to *adjust to unpredictable multicultural situations* that is central to this competence. Multiglossic competence expands on the more commonly known concept of diglossia, whereby a speaker is able to manage two different dialects (or languages) within a single language community. In EIL, multiglossic competence can be argued to mirror Canagarajah's recognition of the ability to "shuttle between varieties of English and different speech communities"(2006a, p. 233), which requires a degree of productive intelligibility (speaking, pronunciation) and receptive intelligibility (familiarity with different varieties and accents of English). The communicative competence model described by Nunn (2007, 2011) is outlined in the following section, and serves as a starting point for conceptualizing an assessment tool for EIL.

3.2   International communicative competence

This section will discuss each aspect of Nunn's (2007, 2011) model for international communicative competence from an EIL perspective. These aspects are strategic, multiglossic, intercultural, pragmatic/discourse, and linguistic competence. Where appropriate, the component of international communicative competence will be connected to examples from the political domain, since the test takers used in this research are government officials. From these aspects a case will be argued for the development of the key constructs for assessing EIL, which will be discussed in greater detail in this chapter from section 3.3.

3.2.1 Strategic competence

Nunn (Nunn, 2011, p. 8) defines strategic competence as "essential two-way components of intercultural communication" whereby competence is measured as the ability to adapt language to international contexts with notions of tolerance and open-mindedness. In the postmodern world, the frequency with which people communicate in multiple English varieties necessitates the ability for those same people to compensate for breakdowns in communication due to English language differences. This *compensation* for communication problems is central to EIL, which no one can claim to be a native speaker of. Therefore knowledge of EIL will never be wholly sufficient and compensation strategies will be needed to bridge the gap to ensure a high degree of understanding. Nunn (2011, pp.8–9) agrees, stating:

> It is a simple but undeniable fact that no one culture and no individual within a culture can demonstrate more than a partial knowledge. This means that all EIL users will need to use their strengths to compensate for inevitable limitations... Compensation for the problems created by partial knowledge is therefore an important skill that all EIL users need.

Problems created by partial knowledge may be intentionally created by what House (House, 2003b) calls *strategic misunderstandings*. Such manipulative tactics may play a role in intergovernmental negotiations where one side is seeking to gain an advantage through deliberately deceitful talk, propaganda, and demagogy. For an example, North Korea is known to mix fake and real agendas in negotiation as a tactic used to achieve hidden objectives (Snyder, 1999; Huh, 2006). Chinese negotiators are also known to employ strategic misunderstandings deliberately to gain the upper hand, usually by wearing down their counterparts with endless issues, employing artificial deadlines, or revisiting previously settled issues (Cain, 2015). While these are extreme cases government officials are unlikely to face, prior knowledge of these tactics and knowing how they can be handled requires a high-level of strategic competence, involving both intercultural awareness and linguistic competence.

Section 3.5 discusses the role of strategic competence as a central component of EIL, and elaborates on how such abilities might be tested as a part of a larger construct called EIL abilities.

3.2.2 Multiglossic competence

To attain proficiency in international communicative competence, EIL users require skills above and beyond traditional linguistic competence. For Canagarajah (2006a) proficiency in EIL means the ability to 'shuttle' between different varieties of English flexibly. McKay (2005) concurs, identifying dialect differentiation and style shifting (e.g. versatility with language use) as useful for this ability to shuttle between varieties. A central component of shuttling between varieties of English and dialect differentiation is intelligibility, which involves the awareness of and familiarity with a broad range of accents in which EIL is spoken. Modern-day English-language educators may or may not be aware that:

A single dialect of that language (Standard American English) fails to equip our students for real-world needs. Not only speakers of the *outer* or *expanding* circle...but those in the *inner* circle of traditional English-speaking communities have to now be proficient in negotiating multiple dialects, registers, discourses...to function effectively in a context of postmodern globalization.
(Canagarajah, 2005, p. xxv original emphasis)

These implications suggest intelligibility – in both productive and receptive forms – is crucial for using EIL in international contexts. Section 3.3 both elaborates on intelligibility (as one of the key constructs of EIL) and suggests a three-level framework that may add clarity to its assessment.

3.2.3 Intercultural competence

Intercultural competence, as defined by Nunn (2011) is much more than simple knowledge of one other culture; it is the ability to adjust to unpredictable multicultural situations. Since EIL communication is always intercultural, Nunn considers intercultural competence to be a key EIL competency. McKay (2002,

2012) sees necessary role for teaching culture in the English language classroom as a means of promoting cross-cultural awareness in order to facilitate intercultural encounters. House (2012) echoes the need to develop intercultural competence in the classroom as a means of raising awareness about the diversity of English language use to avoid misunderstandings. House (2012) also suggests the use of *critical incident* scenarios and role plays (drawn from learner's own experiences) as discussion points for improving EIL communications. Sharifian (2003, 2008, 2009) discusses the role of EIL and *cultural conceptualizations*, a complex-sounding term that simply means "the need to recognize that in international contexts two interlocutors may not share the same system of cultural conceptualizations even though they both use English…" (Sharifian, 2009, p. 246).

While critical cultural incidents can be used to develop intercultural competence (e.g. problem-solving abilities, open-minded attitudes towards other cultures) in the language classroom, the occasional critical incident happens in the real world. For example, in February of 2001, the fast-surfacing USS Navy submarine Greeneville collided with and sank the Japanese fisheries training vessel *Ehime Maru*. Nine crew members (including four students) were killed in the accident which strained international relations between the two close allies and inflamed public opinion on both sides of the Pacific. Lingley (2006) reported that this critical cultural incident raised awareness on the differing cultural norms relating to apologies; on one hand the Japanese families of the victims were demanding a prompt, sincere, and direct apology from the captain of the submarine, while on the other hand, the US government delayed apologizing until after the Navy investigation (and consideration of legal liabilities). Months later an American apology was delivered impersonally through a US envoy in the form of written letters of apology to the families of the victims. The critical cultural incident was essentially that the "Japanese families repeatedly requested a clear and public expression of contrition from [the USS submarine captain] Cmdr. Waddle, given directly to the victims to include an acknowledgement of responsibility and to reflect that he was aware of the grief the families were feeling" (Lingley, 2006, pp. 99–100). However, the Japanese families did not

personally receive an apology from the Cmdr. Waddle until nearly two years after the accident.

The US has dozens of bases and training facilities (and over 50,000 military personnel) in Japan. Critical cultural incidents like the accidental sinking of the Japanese vessel *Ehime Maru* could threaten the military and political relationship between the two countries, especially when inflamed by public opinion and angry politicians.

The potential for serious consequences is high if intercultural competence is lacking at the upper levels of government. This is why organizations such as the ILR (see section 2.3.3.) have developed not only foreign language descriptors, but also five-level skill level descriptions for competence in intercultural communication. These descriptions incorporate both linguistic and extralinguistic elements (e.g. values, beliefs, customs, taboos, etiquette), and are intended to serve as guidelines for use in the US government and may provide a basis for curriculum development, instruction, and assessment (Interagency Language Roundtable, 2012). The ILR regards successful intercultural communication as dependent on characteristics such as interpersonal skills, flexibility, disposition, and the ability to cope with cultural differences. Section 3.4 discusses in greater detail how the intercultural competence construct might be assessed, and which framework might be used to this end.

3.2.4   Pragmatic/discourse competence

Nunn (2007, 2011) identifies pragmatic/discourse competence as the ability to adjust language for given contexts and resolve differences of background knowledge. However, for EIL contexts it needs to be somewhat removed from particular speech communities, given the natural variances in pragmatic norms found internationally. The now dated Canale and Swain (1980, p. 16) communicative approach arguing that "knowledge of what a native speakers is likely to say in a given context…is a crucial component of second-language learners' competence…" runs counter to the notion of EIL as a concept free of native speaker norms or new variety of English. The same problems are raised in section 2.1.7 with regard to Standard English: whose norms are to be applied in measuring pragmatic/discourse competence?

61

Diplomatic protocols have served as the norm for government officials for over a century, outlining how certain activities are to be preformed and how officials are to be appropriately addressed. The pragmatic norms dictated by protocol can simply be described as a set of well-established courtesy rules aimed at maintaining smooth relations with other cultures and countries. While quite formal, these rules might form the basis for an international core of pragmatic norms. However, such a notion would contradict the "…central tenet of pragmatics, namely that social norms regarding language use are open to negotiation" (McKay, 2009, p. 229). Negotiation and fluidity, not norms and standards, are the hallmark of EIL communication.

3.2.5   Linguistic competence

Nunn   (2011, p. 12) sees linguistic competence as unrelated to national origin or first language, noting "There are no native speakers of international English. There are only highly competent speakers without regard to origin." Linguistic competence in at least one variety of English is an essential component of EIL, although being able to shuttle between multilingual communities (Canagarajah, 2006a) of EIL would be desirable.

In this thesis, the test takers are regarded as language *users* rather than language *learners*. As language users, some linguistic competence is assumed, and verified in section 4.1.5 (characteristics of test takers) by means of an IELTS proficiency measure which marks the test takers as intermediate level language users. While linguistic competence is clearly a large component of one's ability to express themselves in English, such measures currently exist in the form of the IELTS, TOEFL, and TOEIC language exams. This thesis aims to capture abilities beyond linguistic measures that enhance and facilitate international communication among people from different linguistic backgrounds.

EIL is prevalent in the working languages of previously mentioned organizations (i.e. ASEAN, the EU, NATO, UN).   Low levels of linguistic competence in the working languages can be problematic, leading to misunderstandings that take time to resolve. While a degree of fluency to communicate among other government officials is a necessary skill, native-like fluency is not needed.

The following sections will examine the theoretical perspective of using intelligibility, strategic competence, and intercultural competence as key skills for successful international communication. For each construct, the literature and supporting frameworks will be outlined as a basis for overseeing the development of these theoretical constructs into operational items in the subsequent chapter.

3.3 Intelligibility

Intelligibility is challenging to define as it means different things to different people. This section will first clear up the complexity surrounding the notion of intelligibility by critically reviewing previous research on intelligibility in different contexts. Next, the three levels of the Smith framework (Smith, 1992; Smith & Nelson, 1985) will be introduced and put forward as a potential model for assessing both receptive (listening) and productive (spoken) intelligibility. The aim of this section is to suggest that when considering the broad contexts in which EIL is used, intelligibility is an international concept and should be judged from an international perspective.

3.3.1 International intelligibility

The literature shows no consensus about what constitutes intelligibility or how to test it (Derwing & Munro, 2005; Jenkins, 2000; Kirkpatrick, 2010). A brief look back at earlier approaches used to capture intelligibility reflects the various definitions and attempts to measure it. Catford (1950) measured intelligibility in terms of effectiveness of communication, which was defined in vague terms of "appropriate response to purpose in speaking". However, of interest to EIL users is Catford's argument that there was a *threshold of intelligibility* related to one's familiarity with another language or variety of a language, such that "more familiarity lowers one's intelligibility threshold, i.e. makes the speech in question more accessible, reduces resistance, and this allows or evince greater intelligibility." (Nelson, 2008, p. 299)

In the decades following Catford's 1950 work, a variety of approaches were used to capture the concept of intelligibility. For example, Lane (1963) measured intelligibility by counting the total number of words listeners transcribed correctly. Later, Brodkey (1972) considered that accurate

paraphrases reflected good intelligibility, while Fayer & Krasinski (1987) used a Likert scale to directly rate intelligibility. Munro (1995) measured intelligibility by means of a dictation where listeners wrote out sentences produced by nonnative speakers and assigned perceived comprehensibility judgments using a 9-point Likert scale.

Despite the variety of approaches towards defining and capturing intelligibility, Smith and Nelson (1985) surveyed over 160 publications on the topic of EIL, intelligibility, and comprehension and summarized the following general consensus from the literature:

1) Variation across Englishes is an indisputable fact.
2) With the increase of EIL users globally, native speakers are neither the sole judge nor the best model of intelligibility.
3) Native speakers of English are not always more intelligible than non-native speakers
4) Intelligibility is a two-way street requiring effort from both speaker and listener.
5) Exposure to different varieties of English enhances intelligibility.
6) Expectations are important; one is likely to find intelligible speakers whom are expected to be intelligible.

<div align="right">(as summarized in Nelson, 2011, p. 42)</div>

This list is not exhaustive, but it does highlight many of the key findings regarding intelligibility. The first two points recognize geographic and demographic realities of the postmodern world. The third point draws from Smith's (1979) work that 'shocked' the English-language teaching world (Nelson, 2011, p. 39) by challenging the native speaker model used in classrooms around the world at the time. The fourth and fifth points from the 1979 Smith and Nelson survey appear logical and not controversial. The final point, however, seems to suggest that preconceived notions about intelligibility override what is actually heard by a listener. It suggests that, for example, if someone believes the pronunciation of Japanese tourists to be unintelligible, then every Japanese

person met thereafter is likely to be unintelligible. This is counter-intuitive, and runs counter to the fourth point on this list.

Intelligibility is a central part of EIL communication because immediate problems arise when pronunciation makes mutual intelligibility problematic. Pronunciation is "possibly the greatest single barrier to successful communication" (Jenkins, 2000, p. 83). Nelson (2011) agrees, defining intelligibility as features of phonetics and phonology required for language recognition. In spoken interaction, conversational participants naturally recognize pronunciation – especially that which differs from our own – as a challenge (see Smith & Nelson, 1985; Smith & Rafiqzad, 1979). In some cases, problems stemming from pronunciation have less to do with features of the language itself, and more to do with social attitudes surrounding certain linguistic features (McKay, 2002).

Intelligibility is connected to the above models of communicative competence (see Table 3.1) in the following ways: first, in a linguistic sense, a minimal threshold of grammar, phonology, and vocabulary must be maintained in order for one to be understood. This is especially true when considering the skills of listening and speaking, areas that can be the most problematic in communication (as opposed to reading and writing). Second, on the discourse level, intelligibility is connected to sentence cohesion and coherence, which also must be maintained. Some of the hallmark skills of successful EIL communications include taking advantage of substitution/ellipsis, collaboration, and backchanneling. Most importantly, EIL users need to be increasingly multiglossic in the postmodern world, as interactions with different varieties and accents of the English-speaking world continue to rise.

That challenge of mutual intelligibility among different varieties and accents has created concerns about international intelligibility that have tended to be more academic than humanistic. Bamgbose laments the academic preoccupation with decontextualized comparisons of English varieties, suggesting that "The point is often missed that it is *people*, not language codes, that understand one another, and people use the varieties they speak for specific functions" (1998, p. 11 original emphasis). In other words, it is the *functional*

purpose of EIL that drives global consumption of English rather than any kind of grandiose ideas about the simplicity, syntax or sound of the English language.

EIL users who work together to communicate successfully demonstrate this functionality by mutually negotiating meaning, regardless of culture, nationality, or language skills. In the era before affordable commercial air travel, contact between peoples of different language backgrounds was more limited. However, in the postmodern era, international intelligibility has arguably become less of a problem, due to the massive amount of global communication taking place in English, as well as in formal language education. The idea that English will fragment into a number of unintelligible languages (like Latin) is a reflection of the wide spread nature of EIL, not a threat to intelligibility. Crystal makes a valid point that for quite some time now, people have been predicting forecasting the ruin of English, citing scholars from previous centuries (Crystal, 2003). Widdowson (1994, p. 385) has a more positive outlook:

> As soon as you accept that English serves the communicative and communal needs of different communities, it follows logically that it must be diverse...it does *not follow logically* [emphasis added], however, that the [English] language will disperse into mutually unintelligible varieties. For it will naturally stabilize into a standard form to the extent required to meet the needs of the communities concerned.

The literature on intelligibility shows a wide variety of approaches. Some researchers focus on the degree of *accentedness*, while others take a broader approach. Table 3.2 summarizes and compares some of the major conceptualizations of intelligibility in recent literature. What this table shows is that intelligibility is either regarded in a *broad sense* (i.e. recognizing an expression or basic literal meaning) or regarded in a *narrow sense* (i.e. accentedness, phonological proficiency, pronunciation differences). Most immediate communication problems in EIL tend to concern both listening and speaking, meaning that EIL users need the dual skills of understanding a wide

variety of accents as well as producing intelligible speech that can be widely understood.

Table 3.2: A summary of research into intelligibility

Researcher	Conceptualization of intelligibility
Bamgbose (1998, p. 11)	"…recognizing an expression, knowing its meaning, and knowing what that meaning signifies in the sociocultural context".
Van der Walt (2000, pp. 140–141)	"…comprehensibility is used to indicate the communication and apprehension of meaning in the broadest sense…and intelligibility only where it refers to the restrictive meaning imposed on it by previous studies".
Munro et al. (2006, p. 112)	Used the constructs intelligibility, comprehensibility, and *accentedness*. Intelligibility is defined as "the extent to which a speaker's utterance is actually understood".
James (2013, p. 212)	Intelligibility can be defined as "the accessibility of the basic, literal meaning, the prepositional content encoded in the utterance".
Hayes-Harb et al. (2008, p. 669)	"In the present study…*phonological proficiency* was operationalized as *accentedness*.
Rajagopalan (2010, p. 468)	Raises the issue of intelligibility by rejecting "blatant cases of unilateral claims of authority to pontificate on intelligibility" and argues forcefully that no variety of English is more intelligible than another.
Jenkins (2000)	Proposed a *Lingua Franca Core* for L2 users to promote wider phonological intelligibility and argues that *pronunciation differences* present the greatest threat to intelligibility.

(As summarized in Nelson, 2011)

The studies summarized in Table 3.2 are by no means a comprehensive list of research into intelligibility. But when taken together, they demonstrate the wide range of conceptualizations of intelligibility and how it can be assessed. This raises a number of methodological issues, in particular the point that native speaker inner-circle criterion for intelligibility runs counter to the notion of EIL. Furthermore, the approach of *foreign accentedness* or comparing various L2 speaker's comprehension of various other accents is out of place in the postmodern era.

It can also be argued that some of the results from intelligibility studies are inconclusive. For example, Matsuura, Chiba, and Fujieda (1999) examined familiarity with accents and comprehensibility among Japanese university students. While these students achieved the highest intelligibility scores with an Irish speaker, they turned around and rated that same Irish speaker second to last on perceived comprehensibility. Other indeterminate results include Munro

et al.'s (2006) conclusion that there was weak evidence to suggest shared L2 accents lead to better comprehension.

Assessments of intelligibility based on a neutral framework (that does not favour any variety of English pronunciation ahead of another) and accommodates both the broad and narrow approaches to intelligibility are summarized in Table 3.2. Moreover, the assessment of intelligibility "should be based on the judgment of *both native and non-native speakers*" (Smith & Bisazza, 1982, p. 259 emphasis added). The following section discusses the Smith framework, which meets the objectives of being variety-neutral, multi-level framework that can be used by any EIL user.

3.3.2   The Smith framework

Intelligibility will be conceptualized in this thesis using Smith's framework (see Smith, 1992; Smith & Nelson, 1985), which is based on his seminal research to address the issue of intelligibility in cross-cultural communication. Smith and Rafiqzad's original 1979 empirical study of 1,386 participants from 11 countries who listened to audiotaped English recordings of nine readers from Hong Kong, India, Japan, Korea, Malaysia, Nepal, Philippines, Sri Lanka, and the U.S. Smith's operational definition for intelligibility was limited to the "capacity for understanding a word or words when spoken/read in the context of a sentence being spoken/read at natural speed" (Smith & Rafiqzad, 1979, p. 371).

Smith and Rafiqzad's 1979 study "…was the first to shock the professional English-language-teaching world with a strong assertion *against* the worldwide supremacy of the pre-eminence of the 'native-speaker'…" and revealed that in *this* study, the native speaker was among the least intelligible of the nine speakers. Further challenging the existing beliefs of native speaker model, Smith and Rafiqzad concluded that "since native speaker phonology doesn't appear to be more intelligible than non-native phonology, there seems to be no reason to insist that the performance target in the English classroom be a native speaker" (1979, p. 380). Further research by Smith and Bisazza (1982) into international intelligibility also challenged another assumption: that non-native students can comprehend fluent non-natives if they comprehend native speakers. Instead,

Smith and Bisazza concluded that it is *active exposure* to different varieties of English that shape a key component of comprehension. This conclusion was reinforced thirteen years later by Smith's (1992) research.

These conclusions make sense on a number of levels. First, the notion that awareness and exposure to a variety of English lead to greater levels of comprehension is intuitive. Second, familiarity with one variety of English does not lead automatically to better comprehension in another variety. Finally, the nuances and subtleties in phonological variations between the varieties and accents of English are often too difficult for all but the most skilled linguists to detect. Speech communities consist of increasingly mobile people who live in different countries, making it harder to tell where someone is from by their accent. In other words, intelligibility is less about nationality (or variety of English), less about whether English was learned as a second or foreign language, and more about active exposure to a medley of English accents and varieties.

3.3.2.1 Three degrees of intelligibility

One of the most significant advances in research into intelligibility was Smith and Nelson's (1985; 1992) division of the concept into three categories: intelligibility (word/utterance recognition), comprehensibility, and interpretability. Smith's conceptualization of intelligibility (hereafter called the Smith framework) allows a more practical discussion and examination of the forms and functions of EIL, and has formed the theoretical basis for other research into intelligibility (see Jenkins, 2000; Munro et al., 2006). The Smith framework sets no standard for intelligibility, no benchmark, and no prescriptive approach that aims to correct that which is outside of a defined standard. Instead it places the responsibility with the listener/speaker.

Figure 3.4: Smith's three levels of intelligibility

and

tends to be more standardized and formal. Misunderstandings in writing and reading are far less likely to have immediate effects, less likely to offend or upset people, and less likely to create misunderstandings. People reading are free to consult a dictionary, or if they disagree with the text, they can simply put away the offending document and forget about it. When writing, people have time to shape their thoughts and ideas carefully, consult a dictionary, draft and edit to ensure the utmost clarity.

The following three sections discuss the Smith framework in greater detail, defining each level of intelligibility and connecting it to EIL contexts.

3.3.2.2   *Word/utterance recognition*

At the base of the Smith framework is 'intelligibility', a less than perfect term, which can be easily confused with the overall broader construct of intelligibility. While Smith uses intelligibility for the most basic level of understanding utterances, this thesis will henceforth use the term *word/utterance recognition* to avoid the obvious confusion between small 'i' and big 'I' intelligibility. Word/utterance recognition is simply the phonological understanding of individual words or other sentence-level element of an utterance Smith suggests that simply asking one to repeat the utterance or write it down as a dictation is an assessment of word/utterance recognition. This is naturally affected by both background noise and is directly related to the speakers' pronunciation and listeners' familiarity with the accent or the speakers' idiosyncrasy (Nelson, 2011). Consider the following example of word/utterance recognition problems caused in this case by a noisy environment:

A: Her family name is Vogeler.

B: Is that spelled V-O-G-E-L?

A: No. Vogeler. V-O-G-E-L-E-R.

B: Oh, Vogeler.

(Y. Kachru & Smith, 2008, p. 61)

The above request for clarification does not involve any word meanings, but rather background noise interfering with the recognition of the family name.

71

Other problems that might interfere with word/utterance recognition are caused by errors in pronunciation, such as the case with very young children or second language learners. Anyone who has ever learned an additional language will most likely have struggled with problematic pronunciation, especially in tonal languages such as Chinese, Vietnamese, or Thai. Likewise, English learners from those three countries often struggle with English pronunciation, and with word stress in particular. Such problems with pronunciation are directly related to a lack of familiarity with the new phonetic systems. An open-mind often helps overcome word/utterance recognition problems because "As one learns to expect differences, one then develops an attitude for understanding varieties different from one's own" (Y. Kachru & Smith, 2008, p. 60).

3.3.2.3 Comprehensibility

Occupying the central rank of the Smith framework is comprehensibility. Smith defined comprehensibility as understanding word/utterance *meaning,* or the recognition of a meaning attached to words uttered in a given context (Y. Kachru & Smith, 2008; Smith, 1992). Nelson (2011) augmented this definition to include the capacity to assign meanings to what one hears and reads. In Smith's (1992)research into intelligibility, he assessed comprehensibility by asking participants to either paraphrase a given statement or answer questions about a given passage. The latter format is quite common on language proficiency tests and exams whereby test takers listen to a conversation and demonstrate comprehensibility by answering questions about it.

Nelson provides an example of comprehensibility by means of a brief dialog from the 1992 film *Lethal Weapon 3* to demonstrate the difference between word/utterance recognition and comprehensibility. In this scene, the characters Murtaugh and Riggs are sitting on a front porch, and Murtaugh is encouraging his departing teenage son to behave. Afterward, it is clear the two older police officers (Murtaugh and Riggs) do not comprehend what means by the slang expression *word* (i.e. acknowledging agreement).

Murtaugh: Hey, Nick…be good.
Nick: Word.

Murtaugh: Hey, word, Nick.

Nick: Word, Riggs.

Riggs: Word, Rog.

Murtaugh: What are we talkin' about?

Riggs: Word – four letters, starts with "w," "d" on the end, "or" in the

middle – word.

Murtaugh: Oh yeah, that word.

Although one can understand and recognize each and every lexical item, this passage has low comprehensibility, since at first reading not much sense can be made of it. This raises the issue of comparing word/utterance recognition and comprehensibility, which can be clarified as follows: in a given passage, if the Smith framework asked test takers to fill in the missing words (cloze activity) it would be a measure of word/utterance recognition. However, if test takers were asked to paraphrase or answer questions about the same passage, the measure would be of comprehensibility. Naturally, the cloze activity is less demanding than comprehensibility. The most challenging aspect of the Smith framework is interpretability, which is described in the following section.

3.3.2.4   Interpretability

This level of intelligibility delves deeper into logical semantics, background knowledge, sensitivity to the listener, and context. The Smith framework defines interpretability as the recognition of an utterance's purpose, or the *meaning behind the words* (Smith, 1992). It can be complex separating and unpacking comprehensibility from interpretability, as the dynamic context of the situation often dictates meaning (as does co-construction). The simplest example might be heard from a colleague, who upon walking into the office declares, "Wow, it sure is hot in here today!" While a listener understanding at a comprehensibility level would concur that it is indeed hot today, a listener understanding at the interpretability level would pick up on the speaker's *intended meaning*, that is to open a window or turn down the thermostat. Another simple example of interpretability is used when answering a phone call. If a caller asks if someone in your office is there, it is interpretability that allows

one to understand that this is not an attendance check, but rather that the caller wishes to speak with said office mate.

While such examples are routine and straightforward contexts, other cases of interpretability are much more complex examples of intended meaning. This may include veiled threats, bribe offers, innuendos, or a sexual solicitation, and are more difficult to convey as social beings. Pinker (2011) suggests that language has two functions in these cases: first, to convey dubious content such as a bribe, command, or proposition. Second, language needs to negotiate a relationship at two levels by first using wording that signals the safest relationship to the listened while simultaneously counting on the listener to read between the lines to entertain a proposition that might be incompatible with that relationship. In relationships where neither side is dominant – such as on a first date – Pinker argues that indirect language is used to avoid socially awkward moments. In international contexts with EIL users the uncertainty of the relationship status (particularly between government officials and business people), can be stressful – that is, until business cards are exchanged and one knows the professional title of their interlocutor, which in turn helps determine first how to address that person and second how directly one might speak to their interlocutor. Unfortunately, there is not an exact correspondence in the world of diplomacy between diplomatic rankings, necessitating the need for indirect language and hedging. In countries where corruption is an endemic problem, indirect language plays a central role in requesting bribes, making threats, and solicitations.

Due to its complexity, the Smith framework has placed interpretability at the highest level of intelligibility, and rightly so. The skills required for deciphering indirect messages are all high-end abilities acquired by adult language users with a high degree of proficiency. For non-native speakers of English, strategic competence skills (section 3.4) are needed to clarify utterances to ensure a high threshold of intelligibility. But since intelligibility is a two-way street, it requires both receptive and productive (speaking) approaches.

The following two sections form the theoretical background for adapting the Smith framework into both a receptive (listening) and productive (speaking) format for assessment of EIL. The goal of the Smith approach is to both

overcome the challenges faced by EIL users who do not follow the standard British or American norms or communicative styles used by major English language proficiency exams, and propose a different type of assessment that aims to capture skills above and beyond language proficiency.

3.3.3   Assessing receptive intelligibility

The conceptualization of the three levels of intelligibility for receptive intelligibility (listening) follows in the steps of the original work of Smith (1992). While Smith used a cloze activity for measuring word/utterance recognition, a phonemic distinction and a short cloze activity (based on a contextualized familiar topic for the participants) will be used in this research. Phonemic discrimination can be used as a means of measuring the degree to which test takers can tell the difference between similar but different sounding words in English, words such as 'ban', 'van', 'pan', and 'dan'. Like Smith, a variety of fluent English speakers from different countries will provide the voice samples for these brief recordings.

Second, comprehensibility can be conceptualized in much the same way as Smith. Test takers can listen to a passage delivered by both native and non-native speakers on a topic similar to what test takers might hear at an international conference. The test takers would then have to answer comprehension questions as a measure of their understanding of word/utterance meaning. Again, this is similar to what is currently done in most English language listening tests, with the exception being that the majority of the speakers will be non-native speakers.

Research into interpretability, however, is complicated by the fact that knowledge of cultural values and frameworks must first be understood. The recognition of the intended meaning of an utterance is dependent upon several variables, such as context, first language influence, communication style, and even familiarity with other English varieties. Any assessment tool of interpretability needs to carefully construct the context and situation in order for intended meanings to carry through to the listener. However, it is possible to draft a number of short nuanced, contextualized conversations carrying indirect meaning. Consider the following example:

The mayor is talking with a senator about an important conference speech on tropical storms and disasters. The mayor wants a search and rescue center built in his city, not on the coast (as was suggested in the speech).

Mayor: *I have an election in two months, and I need to create jobs in my city.*

Senator: *Let me worry about the rescue center. You focus on winning the election for our party.*

In this brief example, it is possible to interpret the senator's remarks to mean that he will exercise influence to ensure that the search and rescue center is built in the mayor's city. This level of understanding goes above and beyond what is actually said, making interpretability one of the greatest intelligibility problems, according to Smith (1992) and McKay (2002). In the case of the participants (see section 4.1.3), managing sensitive topics and information is a critical skill, but also a very difficult skill to acquire, since the subtle nuances of any language rely on a number of variables, such as language proficiency.

3.3.4   Assessing productive intelligibility

Productive intelligibility simply refers to the speaker, who must pronounce and articulate word/utterances in order to be understood by a range of fluent EIL users from different linguistic backgrounds. The primary goal of productive intelligibility is to be understood by a wide range of EIL users, rather than by (typically monolingual) native speakers. This multiuser approach is similar to the testing accommodations suggested by Canagarajah (2006a), who recommended two raters for speaking tasks: one native speaker and one rater from the test takers' own language community. While this restriction is highly impractical, it is possible to have fluent raters from different linguistic communities and assess communicative competence holistically.

The assessment of productive intelligibility using the Smith framework requires three levels of intelligibility: word/utterance recognition, comprehensibility, and interpretability. First, word/utterance recognition can be measured by having participants record themselves reading from a list of

similar sounding words (e.g. ship, sheep, cheap, chip) and assessing the degree to which each word is intelligible and distinguishable from the other words on the list. Jenkins (2000) *Phonology of English as an international language* recommends EIL users focus on segmentals (consonants, long and short vowel distinction) as well as nuclear stress. While nuclear stress is useful for determining the focus of meaning via emphasis, it does not pose an immediate challenge for intelligibility. Of more importance are the frequent problems EIL users have concerning long/short vowel distinction, as well as problematic consonants that impede intelligibility to various first language background users.

Second, the assessment of comprehensibility can be realized by having the participants record a short monologue on a familiar topic. The focus of this assessment is not necessarily on linguistic accuracy, but rather overall intelligibility in terms of fluency and ease of understanding. Scoring rubrics should reflect this notion, and prioritize fluency over accuracy. This notion is supported by Canagarajah (2006a, p. 239), who in his calls for accommodations for EIL users, proposes "The possibility of having someone assess the interaction holistically allows the test to evaluate the candidate on intelligibility and fluency that can transcend grammatical or phonological correctness." The question raised again here concerns whose concept of phonological correctness to follow, to which the solution can be a team of multiple raters from different linguistic backgrounds. Averaging the scores of the various raters would reduce assessment bias and allow for a broader approach to pronunciation than traditionally has been allowed by major language testing companies. Bilingual EIL users arguably are disadvantaged by monolingual native speaker examiners with limited exposure to various accents and might regard unfamiliar accents unfavourably. What makes the type of assessment proposed in this thesis unique is that fluent EIL users from different linguistic backgrounds should serve as judges of intelligibility.

Finally, interpretability (defined as the ability to speak indirectly and appropriately) can be tested by first describing an awkward or uncomfortable context, such as being asked to fire staff (or being evasive). Interpretability can be elicited by requesting an appropriate oral response to these particular contexts

and circumstances. This language skill meets the needs of the test takers in this thesis, since government officials (especially those in the diplomatic domain) occasionally need this ability to operate in the subtleties of the English language.

With all three levels of intelligibility outlined above, what is essential is to maintain mutual intelligibility in oral production throughout the assessment. This can be accomplished by involving fluent raters from different linguistic backgrounds – what Jenkins (2006c, p. 49) calls "a pluricentric approach" – to listen and assess the degree to which the test taker maintains mutual intelligibility (Canagarajah, 2006; Elder & Davies, 2006). This will naturally require criteria to be outlined in scoring rubrics (see Appendix E). The raters from various backgrounds (Appendix D) help ensure that a degree of fairness is accorded the test takers by removing the native speaker as the sole judge of intelligibility. This approach is arguably better from an EIL perspective, placing EIL users themselves as judges of international intelligibility rather than a (typically monolingual) native speaker as a judge.

In addition to international intelligibility among EIL users is the notion of intercultural competence. This multi-faceted, multi-dimensional factor should be accounted for as a component of international communicative competence due to its very nature. The following section introduces some of the key elements of intercultural competence as well as a possible framework for its assessment.

3.4   Intercultural competence

Chapter 2 argued that the trend of global economics and cultural capital has become intertwined in a *transcultural flow* (Appadurai, 1996) of people and goods across porous borders. With this international intercultural flow comes individual, national, and cultural identity which permeates EIL. Knapp agrees, proposing that "that culture-specific knowledge from the native background of a speaker subconsciously seeps into almost every use of EIL" (2015, p. 177). This may largely be the result of EFL teaching materials, which focus on native speaking countries and are ripe with Western topics such as Christmas, Halloween, gender equality, and Western norms of language use.

The teaching and learning of culture as part of EIL is written about extensively (see Canagarajah, 2006a; Kumaravadivelu, 2012; McKay, 2002). EIL users may have a shared language, but not necessarily a shared culture with their interlocutor, which may lead to misunderstandings. To avoid (or at least minimize) these intercultural misunderstandings in communication, international corporations and institutions have invested more time and energy educating and training employees to be aware of intercultural differences and how to be effective in overcoming these differences. This challenge highlights the argument that successful international communication requires knowledge and skills beyond language proficiency.

On the pedagogical front, Kramsch laments this intercultural gap, observing, "After years of communicative euphoria, some language teachers are becoming dissatisfied with purely functional uses of language. Some are pleading to supplement the traditional acquisition of 'communication skills' with some intellectually legitimate, humanistically oriented, cultural 'content'" (1995, p. 85). This might prove challenging, for which other cultural content might teachers focus on, other than the student's own culture? How might such decisions on cultural content be made?

Just as culture is understood to be an intertwined element of any language, the same can be said for EIL. However, unlike most languages, EIL is not tied to any specific culture (although the heavy Western cultural influence is difficult to deny). EIL has many cultures just as it has many users, for each speaker of English brings his or her culture and identity to the conversation. While it is impossible to learn details about every single culture in the world, it is possible to develop a set of characteristics or attributes that enable EIL users to operate flexibly and successfully in most intercultural contexts. In the language classroom, such characteristics and attributes can be developed with the teaching of common cultural norms from major population centers (such as India, China, Brazil, Russia, Nigeria) in order to raise learner awareness of other norms of communication.

However, intercultural competence is also incredibly complex, involving overlapping and multi-faceted research areas such as anthropology, education, psychology, linguistics and cultural studies. As a result, both the concepts of

culture and intercultural communication have created a "wide range of perspectives and emphases and a copious, rather unwieldy amount of knowledge about it" (Matsuo, 2012, p. 348). Moreover, there is not even a consensus on the terminology surrounding intercultural competence (Deardorff, 2006). This is serious problem because how intercultural competence is conceptualized and operationalized depends on how one defines it for research purposes. Given the vast amount of literature on culture and intercultural competence, it is impossible to deal with but a few of the relevant approaches to intercultural competence prior to discussing the approach to be used for assessment in this study.

3.4.1   Conceptualizations of intercultural competence

The literature surrounding intercultural competence appears to revolve around two central concepts: awareness (or knowledge) and the strategic application of that knowledge in the appropriate situation (i.e. performance). Salo-Lee (2006, p. 81) uses the term *competence* broadly, defining it as

> …an ability and skills to do something. It means preparedness, both in terms of motivation and knowledge. Competence also means sufficiency: good enough, yet able and willing to learn. (…) 'intercultural competence' is used broadly to refer to intercultural awareness, knowledge and skills, i.e. both competence and performance.

Therefore it is not only knowledge that is important, but also the *application* of this knowledge. It is this knowing *and* doing that facilitates and enhances EIL communications (as well as minimizing intercultural misunderstandings) and repairing them if they do occur.

Addressing intercultural competence from a European perspective began with the Council of Europe, which in 1989 established the Common European Framework of Reference for Languages (CEFR) to set up systems of validation of language ability. The CEFR, working with teachers and scholars from member nations, identified the key components of intercultural competence as

knowledge, skills, and attitudes (Byram, Gribkova, & Starkey, 2002). This definition is an extension and modification of Byram's (1995, 1997) work on the *five savoirs* (or know-how) approach, which was introduced as a theoretical model for foreign language classroom applications. Framed as learning objectives, the five savoirs are summarized in Table 3.3.

Table 3.3: Byram's (1997) Savoirs of intercultural communicative competence

Savoirs	Skills	Definition
Savoirs	Knowledge	Awareness of social processes both in one's country and the interlocutor's country; awareness of how others are likely to perceive you.
Savoir comprendre	Interpreting and relating	The ability to interpret a document or event from another culture, explain it, and relate it to documents or events from one's own culture.
Savoir apprendre / faire	Discovery / interaction	The ability to learn about a culture and its practices; the ability to apply knowledge, attitudes, and skills in real-time interaction.
Savoir s'engager	Critical cultural awareness	The ability to critically evaluate perspectives, practices, and products of both one's own culture and that of others.
Savoir être	Intercultural attitudes	Showing curiosity and openness; willing to examine one's own beliefs from a different perspective.

Byram (1995, 1997) and the CEFR also underscore the two central intercultural concepts of *knowledge* and the *application* of that knowledge as a demonstration of intercultural competence. Key traits such as 'curiosity', 'openness', 'attitudes', and 'awareness' are again used as benchmarks describing the basic foundation of intercultural competence. Byram is one of few scholars to operationalize intercultural competence in foreign language classrooms (Belz, 2007; Bredella, 2000) and his notion of an *intercultural speaker* – rather than a native speaker – as an ideal model introduced geopolitical correctness to the intercultural competence literature.

There are others who also take an international approach to intercultural competence. Sharifian (2013) examines intercultural competence from an EIL perspective, coining the term *metacultural* competence to describe people who show familiarity and flexibility dealing with various systems of cultural

conceptualizations. For example, loan words such as 'family', 'home', or 'friend' may have different or deeper meanings in other cultures. In some varieties of Aboriginal English, 'mum' can be expanded to include aunts and grandmothers (Sharifian, 2009). In another example, Sharifian argues that EIL users might access their first language system of cultural conceptualizations, such as the Persian *shekasteh-nafsi* (modesty) in response to compliments demonstrated in this dialog between a Persian student and an Australian academic:

Lecturer: I hear you've won a prestigious award. Congratulations! This is
              fantastic.
Student: Thanks so much. I haven't done anything. It the result of your
              effort and your knowledge. I owe it all to you.
Lecturer: Oh, no!!! Don't be ridiculous. It's all your work.

                                                    (Sharifian, 2009, p. 245)

        In the above exchange, both the student and lecturer appear to be lacking intercultural competence. The Persian student is applying his own cultural schema in a western academic context, and the Australian lecturer fails to recognize Persian modesty and reply appropriately. Although this kind of cultural misunderstanding resulted in little more than social awkwardness, incidents of high-stakes cultural incidents between heads of state and their ministers can have serious consequences.

3.4.2   Considerations for assessing intercultural competence

        The plethora of assessment tools for measuring intercultural competence is as varied as definitions for intercultural competence. Fantini (2009) lists forty-four assessment tools, ranging from the Assessment of Intercultural Competence to the Cross-Cultural Sensitivity Scale to the Intercultural Development Inventory. There are all manner of placement tests, attitude tests, diagnostic tests, and proficiency tests, all of which Sercu (2004) suggests cannot be assessed objectively. Examiners are often hard pressed to fairly evaluate candidates performance since culture is often subjectively experienced and construed (Atkinson, 1999; Sercu, 2002). Sercu's (2004) research into

systematic test development for intercultural competence identifies three major characteristics useful for any high-quality assessment of intercultural competence:

1. Validity: constructs must be clearly defined and representative of the target-language domain.
2. Authenticity: a multi cultural setting should be indicative of real-life contexts, as opposed to classroom settings.
3. Reliability: inter-rater calibration is not easy, but using a qualitative assessment rating scale fits well for this kind of assessment.

Deardorff (2006) argues that intercultural competence measures should have multiple perspectives. The learner can indicate to what degree they have been effective in an intercultural setting, but it is only another person who can determine the appropriateness of behavior and communication in the interaction. Unfortunately there is no silver bullet solution for an assessment tool. Given the complexity of this concept, it would be challenging (if not impossible) for any single tool to measure an individual's intercultural competence. An underlying problem remains measuring intercultural competence to what degree? According to whom? A multiple perspective approach is ideal, involving both direct evidence (portfolios, critical reflections, observed performances) and indirect evidence (surveys and questionnaire inventories).

However, in contexts where a measure of intercultural competence does not have the luxury of time for lengthy, multiple perspective approaches, assessment methods must give way to practicality. In using a critical cultural incidents to craft contexts that aim to elicit specific cultural knowledge and attributes (see Fantini, 2009) it is possible and practical to take a 'snapshot' of this knowledge and abilities at a certain point in time.   Of course, any such measure should be useful for its intended purpose and therefore must be subjected to a rigorous examination of its qualities, which is discussed in section 4.3.

When considering the varied approaches to conceptualizing and assessing intercultural competence, one can see a need for a unified approach that encompasses its main components into a neutral framework that can be applied to EIL interactions. For example, educators and researchers in the *Society for Intercultural Education, Training, and Research,* made efforts to identify the most common labels used in the intercultural literature. However, after several years of annual meetings, no clear consensus or single term emerged (Fantini, 2009). As a result, a holistic approach outlined by Fantini (2009) offers up a list of attributes that can be used in the development of an assessment tool of intercultural competence in the following section.

3.4.3   The Fantini Framework

Fantini (2009, p. 459) defines intercultural competence "as complex abilities that are required to perform effectively and appropriately when interacting with others who are linguistically and culturally different from oneself". Fantini breaks down this definition further by explaining how *effective* can be thought of one's own outside view of their performance in a target language-culture, while *appropriate* can be thought of as how the target culture perceives such intercultural performance. For example, a government official might regard a meeting with a hosting foreign counterpart as having gone quite well (from his/her point of view), when in fact the official counterpart may have been upset by something said or done by the visiting official. The ability to conduct effective and appropriate intercultural communication is "…a vital competence in our contemporary world, especially … for specialists involved in mediating between people (diplomats, teachers, consultants…)" (Dervin, 2010, p. 156).

Fantini's (2009) research into identifying the attributes of intercultural competence is based both on his (2006) work and an extensive search of the intercultural competence literature in order to find the commonalities of the construct. These commonalities gave rise to several attributes (in addition to three interrelated areas and four dimensions of intercultural competence that forms a framework for a holistic approach to the complex issue of intercultural competence. This holistic approach is summarized in Table 3.4.

Table 3.4: Fantini's (2009) intercultural competency dimensions

Four Dimensions	Three areas (or abilities)	Attributes
• Knowledge • Positive attitudes • Skills • Awareness	• Establish and maintain relationships • Communicate clearly • Cooperate in order to accomplish tasks of mutual interest	Flexibility, humour, patience, openness, interest, curiosity, empathy, ability to tolerate ambiguity and suspend judgment.

       This theoretical development of the intercultural component of EIL will focus on the attributes "… most commonly cited in the literature (and substantiated by the research)…*flexibility, humor, patience, openness, interest, curiosity, empathy, tolerance for ambiguity*, and *suspending judgments*." (Fantini, 2009, p. 459, emphasis added). These attributes align logically with the literature and core principles of EIL. For example, accommodation skills (Giles & Coupland, 1991; Giles, Coupland, & Coupland, 1991) are used by speakers to cooperate in negotiating meaning, demonstrating empathy, flexibility, and patience. 'Shuttling' between varieties of English (Canagarajah, 2006a) also suggests flexibility. The notion of *transcultural flows* (Appadurai, 1996) and the way that cultures change in fashioning new identities (using global Englishes) is symptomatic of curiosity and interest in other cultures, as well as openness. These concepts are discussed in section 2.3, and form some of the basic underlying principles of communication in EIL. It follows that a complementary approach be used in assessing intercultural competence.

       Government officials operating regionally and internationally need to successfully navigate these transcultural flows and shuttle between varieties of English in the working language of international organizations like ASEAN, NATO, and the EU. The attributes listed in the Fantini (2009) framework are examples of the interpersonal skills sought by embassies and departments of Foreign Affairs worldwide. Moreover, these attributes correspond with Interagency Language Roundtable descriptions of intercultural competence, which is characterized as "…generally dependent on interpersonal skills, disposition, social flexibility, tolerance for ambiguity, and ability to cope with cultural dissonance" (Interagency Language Roundtable, 2012). The challenging

question remaining is how best to elicit these attributes with a practical assessment tool as part of a test battery.

3.4.4   The critical cultural incident approach

These attributes can form the basis for the assessment of intercultural competence in this EIL test battery using the *critical cultural incident* approach, which is defined by Apedaile and Schill (2008, p. 7) as "brief descriptions of situations in which a misunderstanding, problem, or conflict arises as a result of the cultural differences of the interacting parties, or a problem of cross-cultural adaptation and communication." Critical cultural incidents are commonly used on cultural competency training (Sue & Sue, 1990) and can involve critical analysis, interpretation, and the development of strategies for action. Earlier versions of critical cultural incidents were called culture assimilators, which Fiedler et al. (1971) described as "a programmed learning technique that is designed to expose members of one culture to some of the basic concepts, attitudes, role perceptions, customs and values of another culture". To a large extend, cultural assimilators use critical incident scenarios as essential components. Hurn (2011) lists a few of the typical intercultural situations (all of which are directly relevant to the international domain) in which many government officials operate:

- Expatriates interacting with a host nation
- Misinterpreting and mishandling intercultural situations
- Completing tasks or missions assigned to the expat

Critical cultural incidents are often used to train all kinds of employees posted overseas in foreign cultures. Hurn (2011, p.202) provides the following example of an intercultural situation perhaps better suited to a business scenario, but one can see applications to other professions as well.

You are in Beijing as the sales manager of your firm. You are accompanied by your assistant. You are trying to sell malt whiskey to the Chinese. Negotiations have been going on for three days and

86

you feel confident of a satisfactory outcome. You therefore decide to fly back to London on the Wednesday and you tell your assistant to stay to put the final touches to the plans, reminding him of the need to have a draft formal agreement prepared on the Friday. You tell him that any minor details can be clarified later. Comment on this scenario.

While this critical incident appears to be targeting a westerner (e.g. fly back to London), the scenario does tap into Fantini's (2009) framework attributes such as flexibility and patience. However, the format of this incident is instructional, inviting the participants to comment and discuss the situation. Presumably, participants are to display (or not) their knowledge of other cultures and general intercultural competence abilities. From an assessment point of view, it is possible to operationalize these critical cultural incidents as test items that are representative of the real-world tasks and challenges that government officials face on the international stage. From a practicality point of view, the definition and operationalization would need clarity and simplicity, two things that befuddle the assessment of something as vast and deep as intercultural competence. Section 4.1.6.3 will outline the construct definition of intercultural competence and provide examples of potential items that are representative of the kinds of real-world problems government officials face regionally and internationally.

Presently, this chapter has discussed only two of the many possible variables affecting EIL communication: intelligibility (both listening and speaking) and intercultural communication. In the event that misunderstanding actually occurs, EIL users must know what to do next. The next section discusses possible linguistic courses of action that can be followed in order to clarify and repair broken communication.

3.5   Strategic competence

The fourth component of the abilities that are argued to be useful for successful international communication is *strategic competence*. Strategic

competence is one component of the theoretical framework of communicative competence developed by Canale and Swain (1980) and later refined by Canale (1983a). The initial three components of communicative competence are:

1. Grammatical competence: Knowledge of sentence structure (grammar, lexis, syntax, semantics, phonology).
2. Sociolinguistic competence: ability to use language appropriately in given contexts.
3. Discourse competence: ability to combine grammatical forms and meanings for use in different genres (e.g. oral, business letter, essay).

In 1983, Canale added a fourth component (strategic competence) to this model, and defined it as verbal and non-verbal communication strategies used to compensate for breakdowns in communication or to enhance the effectiveness of communication. Others have defined strategic competence in the literature along similar lines, as can be seen by the summary in Table 3.6.

Table 3.5: Definitions of strategic competence

Researcher	Definition
(Canale & Swain, 1980)	Verbal and non-verbal communication strategies that compensate for breakdowns in communication
(Canale, 1983a)	Canale expanded the definition to include strategies that enhance communicative effectiveness (such as changing speed or pitch of delivery for rhetorical effect).
(Bachman, 1990)	Ability to determine communicative goals and resources available in order to execute a message with intent.
(Celce-Murcia et al., 1995)	Communication strategies such as avoidance, circumlocution, stalling, conversation repair, seeking help/clarification, and comprehension checks.
Alptekin (2002)	the ability to cope in communicative situation and to keep the communicative channel open.
(Fulcher & Davidson, 2007)	Knowledge for how to overcome problems when faced with communication difficulties.

The central themes emerging from the literature show that strategic competence involves certain skills and abilities for maintaining and repairing communication problems. In other words, it is simply the development of compensatory communication strategies to manage imperfect knowledge of

language. While this notion may have been conceived from the notion of a language learner, it is easy to see how this construct can apply to native speakers of a language as well. Daily oral communication is rife with misunderstandings, non-understandings, unintelligible utterances, and incomplete expressions. It is through routine clarifications and requests for repetition that meaning is maintained in communication.

Developing strategic competence has already been highlighted as an important aspect of communicative competence by Bachman (1990). Smith (1983, p. 8) agrees, emphasizing that "...the responsibility for effective communication is shared by both the speaker and the listener...The listener must make an effort to understand". With ordinary oral communication, meaning is negotiated and co-constructed in a joint effort to achieve understanding (House, 2003a; Jenkins, Cogo, & Dewey, 2011; Seidlhofer, 2004). It is this joint effort that underlies much of EIL communication, and is facilitated by communication strategies and "the collaborative nature of communication and linguistic negotiation" (Canagarajah, 2007a, p. 928) that are both encompassed by strategic competence. Given the diversity of English varieties around the world, strategic competence is also needed to "negotiate diverse varieties to facilitate communication" (Canagarajah, 2006a, p. 233), even among first language users of English. Again, in cases where speakers and listeners are unfamiliar with the regional connecting linguistic, socio-cultural and actional competence, suggesting that it performs some sort of unifying function that propels these three aforementioned competences. Celce-Murcia et al. (1995, p. 28) have divided strategic competence into five main parts described here:

1. Avoidance or reduction strategies
   - message replacement
   - topic avoidance
   - message abandonment
2. Achievement or compensatory strategies
   - Circumlocution: y'know, that thing you open wine bottles with.
   - Approximation: Fish instead of carp or tuna.
   - All-purpose words: stuff, thingy

- Non-linguistic means: gesturing, pointing

- Restructuring: the bus was very…there were a lot of people on it

- Word-coinage: ex. *Vegetarianist*

- Literal translation from first language

- Foreignizing: pronounced with distinct foreign accent

- Code switching to another language

- Retrieval: bro… bron… bronze

3. Stalling time-gaining strategies

- fillers, hesitation devices and gambits: well, actually…where was I?

- self and other repetition

4. Self-Monitoring Strategies

- self-initiated repair: I mean…

- self-rephrasing: this is for students…pupils… at school.

5. Interactional strategies

o Appeals for help (direct and indirect)

o Meaning Negotiation Strategies

- Repetition requests: pardon? Can you say that again?

- Clarification requests: what do you mean by…?

- Confirmation requests: did you say…?

- Verbal and non-verbal expressions of non-understanding

- Interpretative summary (e.g. You mean…? So what you are saying is…?

o Responses (repetition, rephrasing, expansion, reduction, confirmation, repair)

o Comprehension checks

The above list of communication strategies illustrates the wide variety of options available for repairing, clarifying, and enhancing communication. These strategies are essential for EIL users in the event of unintelligibility, misunderstanding, and non-understanding. EIL users regularly must navigate and negotiate communication between each other, requiring "…the capacity to negotiate diverse varieties to facilitate communication." (Canagarajah, 2006a, p.

233). Strategic competence is not about how much knowledge or language proficiency one has, but rather about the versatility with which one completes tasks using EIL. It concerns the sum of all resources available (e.g. linguistic, intercultural, pragmatic, discourse, multiglossic awareness) and the ability to apply these skills effectively and appropriately for a successful outcome in international communicative contexts. However, eliciting and capturing evidence of these skills is no easy task. Assessment of strategic competence must be realistic and pragmatic itself, in order for such a test to be useful. These issues are raised and addressed in the following section.

3.5.1   Assessing strategic competence

Any useful assessment tool for strategic competence should be defined with clarity. Canale considers a through framework description "…a necessary step in constructing a model since the specification of how various sets of knowledge and skills interact and develop (a model) can only be as strong as the specification of these various competencies (a framework)." (1983a, p. 12). Therefore the definition of what constitutes strategic competence must be as detailed as possible. Drawing from Celce-Murcia et al. (1995), strategic competence is conceptualized in this thesis as a *linguistic* measure of the ability to repair communication. Of the communication strategies described in section 3.4, the following four strategies will be used:

1. Clarification requests: *What do you mean by…?*
2. Requests for repetition: *Pardon? Can you say that again?*
3. Confirmation request: *Did you say…?*
4. Direct appeals for help: *What do you call/mean…?*

These four communication strategies have been chosen from Celce-Murcia et al.'s (1995) *interactional* strategies category, which is indicative of the cooperative aspect of strategic competence. It is this cooperation, the negotiation of meaning, the 'shuttling' back-and-forth, and the accommodation of different speech norms that is of interest in this thesis. Moreover, these

interactional strategies are equally simple and effective means by which EIL users can resolve communication problems caused by a lack of intelligibility, unfamiliar expressions, and limited vocabulary. While it is acknowledged that strategic competence is a much bigger theoretical construct than addressed in this thesis, the limitation of communication strategies to the interactional category enables the practical elicitation and measurement of strategic competence from an assessment point of view. The operationalization of these communication strategies is discussed in Chapter 4.

3.6 Chapter summary

This chapter begins with the comparison and contrast of the traditional and more recent models of communicative competence, arguing that in the case of EIL, flexibility and accommodation are central components of a theoretical EIL foundation rather than the fixed second language acquisition notions of communicative competence. The expansion and evolution of communicative competence shows its increasing complexity and the shift away from Chomsky's (1957, 1965) notion of the native ideal speaker towards Nunn's (2007, 2011) notion of international communicative competence. In the conceptualization of EIL, connections are made between the components of Nunn's model and importance of these skills for government officials, who must navigate the intricacies of EIL even though it is often not their first language.

Intelligibility is argued to be an initial challenge among EIL users unfamiliar with other varieties and accents in English. Using the Smith framework, intelligibility was described and divided into three levels: word/utterance recognition, comprehensibility, and interpretability. In adapting the Smith framework, a productive and receptive approach was given to each of these levels, emphasizing the immediate nature of communication problems in EIL speaking and listening.

From the rich field of intercultural competence, focus was placed on the central concepts of intercultural awareness (or knowledge) and the strategic application of that knowledge in appropriate and effective ways to resolve communication problems. Nine attributes from the Fantini (2009) framework were highlighted and connected to the characteristics often found in

intercultural EIL communication. While there are a wide variety of assessment tools measuring different aspects of intercultural competence, a 'snapshot' of knowledge can be useful for determining the degree of awareness of different cultures and effective ways of negotiating cultural pitfalls.

Strategic competence remains a key component of communicative competence, and even more so now that a shift has occurred in the theoretical models. More emphasis has been placed on the skills and abilities necessary to clarify or even enhance international communication, which is essential for maintaining a threshold of understanding among EIL users. The interactive approach outlined by Celce-Murcia (1995) is considered an appropriate match for the ongoing negotiation and co-construction of meaning in international communication. The assessment of strategic competence may follows a simple approach, but it is argued here that this is necessary as but one part of a larger test of EIL abilities.

This chapter has reviewed and borrowed from a wide number of theoretical concepts and frameworks in an attempt to form a foundation for the recognition, elicitation, and measurement of EIL abilities. While there are other variables that certainly affect international communication, this thesis aims to identify and elicit four of these constructs: receptive intelligibility, productive intelligibility, intercultural competence, and strategic competence. The design and operationalization of these theoretical constructs into testable items will be discussed in the following chapter.

PAGE LEFT BLANK

Chapter 4: Design Statement and Operationalization 4

This chapter is divided into three major sections: the design statement for a test of EIL abilities, its operationalization, and a plan for validating the test. Following Bachman and Palmer's (1996) framework for test development, the design statement covers the purpose of the test, a description of the target language domain, characteristics of the test takers, and a definition of the constructs to be tested. As a part of the design statement, two plans are included: first, to evaluate the qualities and usefulness of the test and second, a plan to allocate and manage resources (people, time) in the administration of this test. This operationalization section describes the test structure, task specifications, and implementation of each of the four constructs into the online EIL test battery. A table of the test battery summarizes the four constructs, their theoretical definition, and scoring to show the overall structure of the test. The final section of Chapter 4 outlines a detailed plan for validating the test battery according to Bachman and Palmer's (1996) six qualities of test usefulness.

4.1   Design statement

The first stage of test development is the detailed assembly of test components that help ensure that performance on the test tasks match the desired language use as closely as possible. Any design statement should be a flexible product, to be revisited throughout the various stages of test development. These stages of the EIL test battery are outlined following the Bachman and Palmer framework as described in the introduction to this chapter.

4.1.1   Purpose of test

The purpose of this test is to measure the abilities argued necessary to successfully use English in international contexts. Explicit references defining these abilities will be made, as well as the inferences that can be drawn from the test results. These test results constitute score-based inferences that can be used for making decisions about the test takers, described in section 4.1.5. Traditionally, such decisions are used for placing test takers in a specific class or grade, or for assigning a grade. Although this EIL test battery is a low-stakes

proficiency test with arguably no consequences from the test results, future refined versions of the test could potentially be used to screen candidates for diplomatic or international work where English is the common language. In such a case, the test might become a high-stakes test (requiring extensive validation) rather than exploratory testing research.

4.1.2   Testing mandate

Traditionally, the word mandate indicates an official order from an authoritative body to carry out some course of action. A *testing mandate* is used to describe the purpose or reason for a test, and whether or not it is internal or external to the testing domain (Davidson & Lynch, 2002). A classroom teacher usually initiates an internal mandate by giving a low-stakes test, such as a quiz or end of mid-term test. Internal mandates are often *formative*, and thus intended to be a part of the teaching and learning process. External mandates, on the other hand, emanate from outside the local context usually from people who don't know the teachers or learners. They are usually designed to measure proficiency, and are *summative* (i.e. at the end of a course).

The EIL test battery stems from an external mandate, and is based on the theoretical argument that there is much more at work in international communication than traditional language skills of vocabulary, grammar, and fluency. The test is designed to measure proficiency in theoretical constructs called EIL abilities. It is an external mandate aimed at answering the call from some in the academic community (Canagarajah, 2006a; Davies, 2009; Elder & Davies, 2006; Jenkins, 2006c; Smith & Nelson, 1985) to re-examine current Standard English proficiency exams in light of shifting demographics (see section 2.3) of English-language users.

4.1.3   Who is this test for?

The EIL test battery is a low-stakes assessment designed for mid-level government officials, who constitute the specific domain of language use in this thesis. In particular, the assessment aims to capture the potential contexts and language use these officials might encounter in the workplace or other international situations such as meeting/guiding foreign guests, international

conferences, social occasions. These government officials (see section 4.1.5) are intermediate-level EIL users who use English on a regular basis but lack a higher level of proficiency and fluency, which is increasingly required for intergovernmental communications.

4.1.4   Description of domain and task types

Bachman and Palmer (1996) define the target language use domain "…as a set of specific language use tasks that the test taker is likely to encounter…and to which we want our inferences about language ability to generalize." (p. 44). There are two types of target language use domains: language classroom and real-life. Of greater interest is the real-life language use domain of the test takers who use EIL both in the workplace and internationally to communicate with other EIL users from different linguistic backgrounds. Specifically, the focus will be on intelligibility (speaking and listening), strategic competence (conversation repair), and intercultural competence. Although these government employees handle a large amount of written work in English (reports, information, meeting minutes), this written format does not present the kind of immediate comprehension challenges as does listening and speaking.

In order to uncover the task types in the target-language use domain, a needs analysis was conducted by two means: an online survey of the test takers and classroom observations (see Appendices B and C). The online survey was very useful in helping to develop tasks that matched the situations for which test takers were currently using EIL in the workplace. Classroom observations, on the other hand, were aimed at uncovering potential communication breakdowns and their resolution. However, since the classroom setting was designed to facilitate communication there were few observed cases of communication breakdown.

4.1.5   Characteristics of test takers

The test takers are learners in the ELTO (English Language Training for Officials) program, in partnership with the Victoria University of Wellington and the New Zealand Ministry of Foreign Affairs and Trade. The ELTO program is

designed to help improve the English language skills of government officials and enable better communication with foreign counterparts in various government ministries.

In the final group of test takers (*N*=65), 58% were male and 42% female, with ages ranging from 20 to 55. Just over half (51%) of these government officials are in their thirties. The test takers are all working professionals holding mid-level positions in their respective governments in various African and Asian nations (as well as Peru). The participants are also bilingual and multi-lingual individuals, as revealed by the survey (see Appendix B). In addition to English, the Asians indicated the ability to speak some European languages (French, Portuguese, German) and a number of regional languages such as Lao, Vietnamese, Thai, Khmer, and Cambodian. The African participants are familiar with other international languages such as English, French, Arabic and Swahili, as well as a long list of assorted local languages.

From an English-language proficiency perspective, the average IELTS test score for the group was band 5.5. According to the IELTS website, this competency level places the participants between being a modest and a competent language user, with skills ranging between partial and effective command of the language, despite some inaccuracies and misunderstandings. Communication skills at this level range from basic to fairly complex use of language.

Table 4.1: Nationalities of test takers

Number of test takers	Nation
11	Vietnam
1	Benin
1	Burundi
14	Cambodia
5	East Timor
1	Gabon
1	Ivory Coast
11	Laos
1	Madagascar
12	Myanmar
5	Peru
1	São Tomé
1	Togo
N = 65	Total: 13 countries

4.1.6   Construct definitions

This section will make explicit the nature of the four EIL abilities described in Chapter 3, resulting in a theoretical definition of the four constructs. It is these definitions that will form the basis for the operationalization of test tasks, which are based on a model of language ability, rather than a more concrete model such as a classroom syllabus. It is in the operationalization stage that Bachman and Palmer urge the test developer "...to make a conscious and deliberate choice to specify particular components of the ability or abilities to be measured in a way that is appropriate to a particular testing situation."

4.1.6.1   Three degrees of receptive intelligibility

Intelligibility was uncovered and discussed early on as a likely construct of EIL abilities.   Classroom observations (see Appendix C) quickly revealed that of immediate concern to the ELTO participants was that of unfamiliar accents. This was apparent particularly in joint Asian-African interclass activities where non-understandings were common. Informal interviews with participants following these activities indicated that unfamiliar accents presented the biggest challenge for mutual intelligibility. McKay argues that it is "...interpretability [understanding a speaker's intentions] that causes the greatest problems in the use of EIL for cross-cultural communication since interpretability entails questions of culture and context." (2002, pp. 52–53). However, of more immediate concern is the unfamiliarity with different varieties and accents of EIL, since intelligibility involves the cognitive processing of words and utterances into literal meaning, which precedes understanding of a speaker's intent. In other words, if the message cannot be understood, issues of context and culture cannot even begin to be negotiated.

As discussed in Chapter 3, the construct of receptive intelligibility is divided into three degrees, according the Smith (1992) framework: word/utterance recognition, comprehensibility, and interpretability. This conceptualization allows a more practical discussion and examination of the forms and functions of EIL, and has formed the theoretical basis for other research into intelligibility (see Jenkins, 2000; Munro et al., 2006).

Word/utterance recognition is a sub-construct of the most basic level of intelligibility. It is defined as the recognition of isolated words spoken by a wide variety of EIL users with a range of different accents. This is a phonetic distinction definition, and aims to test the ability to discern similar sounding words both in and out of context.

Comprehensibility was defined by Smith (1992) as understanding word/utterance *meaning.* This sub-construct is similar to the 'comprehension' construct in existing English language proficiency tests, with the distinction being that the main speaker is a multilingual non-native speaker. The comprehensibility sub-construct will be defined in the same way as Smith, and aim to measure listening comprehension skills following a short speech on a topic that is frequent in intergovernmental conferences.

Interpretability is identified in the Smith framework (1992) as the highest level in the cline of intelligibility, defined as the understanding of the meaning behind words, or their illocutionary force, or the intended meaning). This definition will be maintained and applied to contexts familiar to the test taker (such as an international conference).

4.1.6.2   *The three degrees of productive intelligibility*

As argued in section 3.2.4, productive intelligibility will be focusing on speaking (and not writing) for the following reasons: first, EIL is primarily a spoken language and the literature focuses on this productive skill. While receptive intelligibility aims to capture listening skills, productive intelligibility conceptualizes intelligibility in terms of the *speaker*, who must produce articulate word/utterances in order to be understood by a variety of listeners. Productive intelligibility is the natural match for receptive intelligibility, since every spoken utterance requires both a listener and a speaker to receive and produce communication. The Smith (1992) framework of intelligibility has been adapted to include a speaking and listening approach to account for the fact that most EIL communication is interpersonal and interactional rather than the stable fixed code of reading and writing. Although government officials do deal

with a lot of English documents at work, they rarely cause communication challenges in the same sense as spoken language.

Therefore, the three sub-constructs of intelligibility are defined in the following way: first, word/utterance distinction is defined as the ability to distinguish similar-sounding English words intelligibly and articulately. In this sub-construct, assessment is looking at vowel length (e.g. chip vs. cheap), digraph distinction (e.g. ship vs. chip), and initial consonants (face vs. base vs. pace). The rational draws first from Jenkin's (2000) lingua franca core, which regards the following pronunciation qualities as important for clear intelligibility:

a) A distinction between long and short vowels
b) The phonetic realization of consonants (e.g. /f/ and /v/)
c) The pronunciation of /r/ as a rhotic retroflex approximate (e.g. the clear pronunciation of /r/ as in 'ca_r_ pa_r_k' rather than 'cah pahk'.

The lingua franca core goes into much greater detail, such as focusing on consonant clusters, diphthong length, and nuclear stress. However, due to the format of the test items (i.e. word distinction) it was not possible to incorporate all of the lingua franca core pronunciation features.

The second rationale for selecting vowel length, digraph distinction, and initial consonant distinction as the basis for the word/utterance sub-construct stems from semi-structured interviews with an international group of seven PhD researchers working at Victoria University of Wellington. As bilingual EIL users themselves, these researchers were also experienced language educators, familiar with the pronunciation difficulties of English language students from their respective homelands. Thus, from their knowledge and experience teaching English in East and Southeast Asia, Africa, North America, and Europe seven researchers were asked which sounds in English were difficult for most of their English language learners in their speech community. Results from informal interviews with this international group of seven English language educators combined with the online data regarding problematic phonetic challenges facing EIL users resulted in the information shown in Table 4.2.

101

Table 4.2: Phonemic discrimination categories for receptive intelligibility

Phoneme	Example
*th	there/dare, three/tree, thin/tin
v/b/f	van/ban, fat/vat
l/r	pray/play, late/rate
Long/short vowel pairs	Ship/sheep, fell/feel, not/note

*no distinction is made between the voiced dental fricative /ð/ (as in <u>th</u>is) and the voiceless dental fricative / θ / (as in <u>th</u>ing) as both sounds do not lead to a loss of intelligibility (Deterding & Kirkpatrick, 2006; Jenkins, 2000, p. 137).

This is by no means a comprehensive list of phonetic challenges facing EIL users, but it is a list of salient features resulting from semi-structured interviews with teaching professionals from a number of regions. This pattern of four phonetic challenges will be repeated using different words and different EIL voices.

Spoken comprehensibility is defined as the degree to which the test taker's spoken English is fluent and pronunciation is readily understood by EIL raters from different linguistic backgrounds (see Appendix D). No special topical knowledge is required, as the test takers will be speaking for about a minute on familiar themes and assessed by means of a holistic rubric on intelligibility.

Interpretability was a sub-construct which did not adapt easily into the Smith (1992) framework. However, by extrapolating the passive ability of understanding intended meaning into the productive category, interpretability as a productive skill can be theoretically defined as the ability to speak indirectly in the following situations:

a) An awkward situational context (e.g. being asked to fire a staff member).
b) Using avoidance strategies (e.g. not revealing private information)
c) Handling rejection in such a way as to not upset the speaker

Instead of listening comprehension, test takers will be asked to produce word/utterances that convey meaning indirectly. Government officials

(especially those in the diplomatic domain) arguably face their greatest challenges when dealing in the subtleties of language at this level.

4.1.6.3   Intercultural competence

Intercultural competence in this thesis follows the (2009) Fantini framework, which regards intercultural competence "...as complex abilities that are required to perform effectively and appropriately when interacting with others who are linguistically and culturally different from oneself" (Fantini, 2009, p. 459). The theoretical definition of intercultural competence used in the EIL test battery is more specifically defined as the set of attributes "... most commonly cited in the literature (and substantiated by the research)...flexibility, humor, patience, openness, interest, curiosity, empathy, tolerance for ambiguity, and suspending judgments." (Fantini, 2009, p. 459). Critical cultural incidents will be used to elicit these sub-constructs from the test takers in the form of descriptive contexts outlining an intercultural incident, which must be resolved by test takers. Section 4.2.7 will provide examples of a test item for intercultural competence.

Special topical knowledge is required for intercultural competence, as test takers need specific cultural knowledge in order to answer correctly. As a proficiency test, the EIL test battery cannot point to a syllabus where this information is acquired, but generally speaking, the greater the intercultural awareness of the test takers, the better the results should be on this test construct.

4.1.6.4   Strategic competence

The fourth theoretical component of EIL abilities to enhance international communication is strategic competence. Drawing from the literature in Chapter 3, the construct definition for strategic competence will be a measure of the ability to repair broken conversation, as exemplified by clarification requests, requests for repetition, confirmation requests, and direct appeals for help. This knowledge of how to overcome problems in communication is realized as actual communication, or instances of language use as shown in the following examples:

1. clarification requests: *What do you mean by...?*
2. requests for repetition: *Pardon? Can you say that again?*
3. confirmation request: *Did you say...?*
4. direct appeals for help: *What do you call...? What do you mean...?*

4.1.7 Summary of constructs and design statement

This summary encapsulates two major sections of the test development process – defining the theoretical constructs and the design statement – into two tables (4.2 and 4.3 respectively). This summary enables one to compare and contrast the various components and constructs of the EIL test battery in an easy to follow manner. First, the definitions of the four constructs are specified in Table 4.3. They are derived from the literature and theory discussed in Chapter 3, and operationalized in section 4.2, which outlines how the constructs will be tested. Bachman and Palmer (1996, p. 116) remind us that these construct definitions serve three purposes: to provide a basis for using test scores for their intended purpose, to guide test development efforts, and to support the construct validity for interpreting test scores.

Table 4.3: Summary of construct definitions

Construct	Framework	Theoretical construct definition
Receptive intelligibility	Smith (1985; 1992)	1. Word/utterance recognition: the recognition of isolated words spoken by different accents. 2.Comprehensibility: understanding word/utterance *meaning.* 3. Interpretability: understanding the meaning behind words.
Productive intelligibility	Smith (1985; 1992)	1. Word/utterance recognition: the spoken ability to distinguish similar-sounding words in terms of articulation and intelligibility. 2.Comprehensibility: the degree to which spoken English is fluent and pronunciation is readily understood by EIL raters. 3. Interpretability: the ability to speak indirectly in an awkward situational context.
Intercultural competence	Fantini (2009)	Characteristics of flexibility, humor, patience, openness, interest, curiosity, empathy, tolerance for ambiguity, and suspending judgments.
Strategic competence	Celce-Murcia et al. (1995)	The ability to repair broken conversation by means of clarification requests, requests for repetition, confirmation requests, and direct appeals for help.

Next, the design statement of the test is summarized in Table 4.3. Design statements are used to aid in the planning for the required amount of time, staff, and money necessary to conduct a test efficiently. They also highlight the degree of practicality of the test, which is often a paramount concern to test developers. The plan for managing resources, along with the test purpose, description of the target language domain, definition of the constructs, and plan for evaluating the usefulness of the EIL test battery is summarized in Table 4.4.

Table 4.4: Design statement for the EIL test battery

1. Test Purpose				
Inferences	About the abilities of government officials to use English effectively in international contexts, such as at international conferences and intercultural workplaces.			
Decisions	Low-stakes theoretical test			
Impact	No impact on test takers academic work			
2. Description of target language domain	Real-life tasks based on survey of participants' English language use at work. Government officials using EIL for international communication with counter-parts discussing matters of mutual concern. Abilities include familiarity with other accents, clear pronunciation, intercultural competence, and managing communication breakdowns.			
Test Constructs	1. Receptive intelligibility 2. Productive intelligibility 3. Strategic competence 4. Intercultural competence			
Setting				
Participants	Mid-level government officials for whom English is not a first language.			
Time of task	Test takers complete online test at home.			
Input	Receptive Intelligibility	Productive Intelligibility	Strategic Competence	Intercultural Competence
Channel	Aural and visual	Spoken	Aural and visual	Visual
Form	Language	Language	Language	Language
Length	Short: 40 items	Short: 16 items	Short: 10 items	Short: 10 items
Type	Items	Items and prompt	Items	Items
Speededness	Unspeeded	Unspeeded	Unspeeded	Unspeeded
Language characteristics	Variable	Variable	Variable	Variable
Expected Response				
Channel	Aural, visual	Spoken	Aural, visual	Visual
Form	Language	Language	Language	Language
Length	Long: 40 items	Short: 16 items	Short: 10 items	Short: 10 items
Type		Oral production	Prompt and task	Prompt and task

Speededness	Unspeeded	Unspeeded	Unspeeded	Unspeeded
Language Characteristics	Identifying a variety of English accents in isolation and in context.	Functional, figurative, and pragmatic usage of language.	Functional language use.	Cultural references and topical knowledge.

3. Description of test takers

A. Personal characteristics: Aged 25-55, career government officials, multilingual non-native speakers of English from developing Asian and African countries. Population: volunteers from ELTO program

B. Topical knowledge: mostly non-topical, except for the intercultural competence section.

C. Profiles of language knowledge: Intermediate language users.

D. Possible affective responses to taking the test: As volunteers, test takers are demonstrating interest in the test battery. Test results have no effect on their course grades.

4. Definition of the constructs

The test constructs are based on the theoretical definitions of EIL as discussed in Chapter 3. The test is a proficiency measure of the abilities argued to be necessary for successful international communication in English. This knowledge is not a part of a syllabus and topical knowledge is not part of this test.

A. Knowledge of grammatical forms and meanings

B. Topical knowledge

C. Strategic competence

5. Plans for evaluating test usefulness

A. Reliability	Provide clear scoring rubrics for raters. Provide Cronbach's alpha for internal reliability. Provide inter-rater reliability calculations using Krippendorf's alpha.
B. Construct validity	Collect variety of evidence (e.g. convergence, discriminant validity, content validity) to support inferences from test results.
C. Authenticity	Survey and observations of target language domain and topics.
D. Interactiveness	Test tasks engage limited topical knowledge of test takers with regard to intercultural competence.
E. Impact	Limited; this is a theoretical test of EIL abilities with no academic consequences. Unintended consequences are acknowledged to exist.
F. Practicality	Until the test is operationalized and online, its development will be time-consuming. Once online, administration and data collection will be very efficient.

6. Allocating Resources

A. Human resources (hours)	Design	Operationalization	Administration	Scoring	Analysis
Test coordinator	250	100	20	25	50
Reviewers	5	-	-	-	-
Human raters	10	-	-	10	-

B. Space resources	Computer lab bookings.
C. Equipment	Computers, minimal papers.
D. Costs	Human raters require compensation for their time and skills.
E. Test development timeline	Specify key dates for the development of the EIL test battery.

4.2   Operationalization

For something to be operationalized, it must first be put into measureable terms (Davidson, Hudson, & Lynch, 1985). This, the second major part of this chapter, outlines the online operationalization of the EIL test battery, reporting the procedures used to develop the theory into something measureable. This section will discuss the design of the online assessment tool, its various technical aspects, and how they combine to form the online test battery for EIL abilities.

4.2.1   Instrument design: An overview

The EIL test battery was developed on Qualtrics, a popular survey-based research platform used by academics and business people for online data collection. Online tests have a number of advantages over the traditional paper and pencil tests, as summarized immediately below. The online platform option was chosen because it allows researchers to:

1. email a link to the test anywhere in the world
2. embed audio clips for a listening component
3. take advantage of accurate automatic scoring
4. exported test results for statistical analysis by SPSS
5. download the test into Microsoft Word or Excel
6. randomize the distractors to avoid sequencing effects
7. add timers to measure the time taken for each test section

These advantages far outweigh those of paper and pencil testing, especially with regard to point number one above. In order to reach a broad spectrum of test takers, it was necessary to find a means that enabled easy test administration overseas. This gain in practicality did however result in some sacrifices in authenticity that would arise from personal interactions.

Qualtrics also features advantages for the test takers, such as a *save and continue* feature which means the test does not need be completed in one sitting in a fixed location. This likely contributed towards more complete data sets in the end. Moreover, test takers received immediate scoring at the conclusion of the test, after which all of the test items and correct responses are displayed.
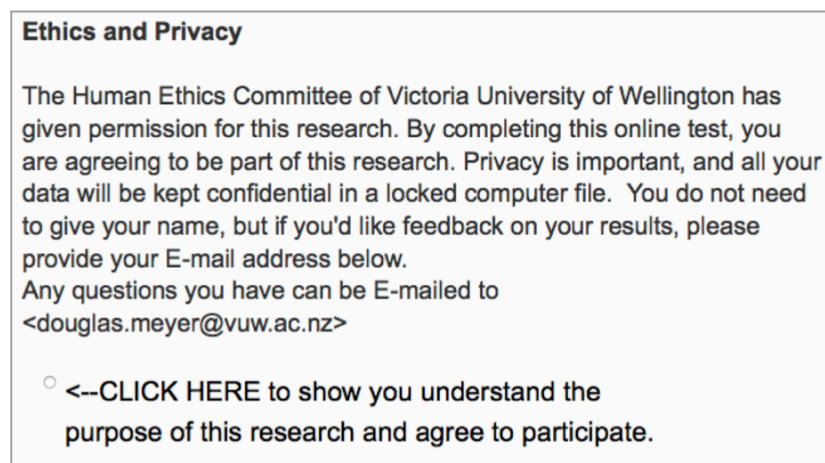
4.2.2 Technical aspects of the online test battery

The following section will describe the technical aspects of the EIL test battery that help to explain the methodology behind its development. This includes how audio files were embedded into the test, the time allotment for test takers, and how the survey was dispersed to test takers in various countries.

4.2.2.1 Ethics, privacy, and security

Consent to participate in this research was granted by all of the test takers, who could not proceed past the introductory page and begin the test battery unless they agreed to the ethics and privacy statement (Figure 4.1).

Figure 4.1: Consent form

> **Ethics and Privacy**
>
> The Human Ethics Committee of Victoria University of Wellington has given permission for this research. By completing this online test, you are agreeing to be part of this research. Privacy is important, and all your data will be kept confidential in a locked computer file. You do not need to give your name, but if you'd like feedback on your results, please provide your E-mail address below.
> Any questions you have can be E-mailed to
> <douglas.meyer@vuw.ac.nz>
>
> ○ <--CLICK HERE to show you understand the
>   purpose of this research and agree to participate.

Storing all test data on computers that require a log in and password to access addressed privacy concerns. Additionally, raters were asked to sign a confidentiality agreement whereby they agreed to respect test taker privacy by protecting test data in password-locked private computers. Test security was not a great concern, since the EIL test battery is a low-stakes theoretical research test. Since results from this test would have no impact on participants' coursework in the ELTO program or future careers, there is no conceivable reason for cheating on a test of this nature.

4.2.2.2 Audio files

Qualtrics allows users to insert audio and video media clips up to 16 MB per question, easily facilitating the insertion of audio clips into the EIL test battery. However, since Qualtrics does not host video or audio clips directly, it was necessary to create an account on an audio file sharing site called SoundCloud (SoundCloud, n.d.). This site hosted the numerous audio clips that were used for the listening section of the test. Embed code links were used from SoundCloud and successfully inserted into Qualtrics for each listening item.

Figure 4.2: Qualtrics audio clip on SoundCloud
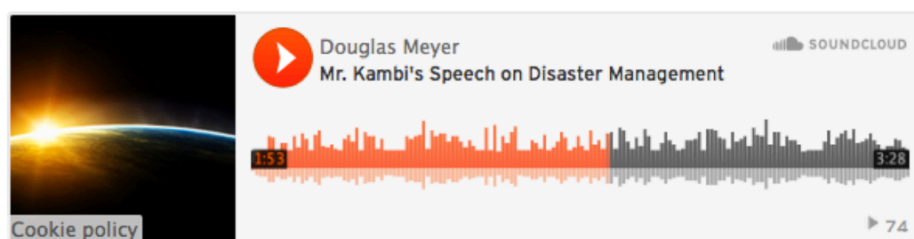
This format proved to be intuitive, and few test takers in the pilot versions of the test battery reported confusion over what was required of them. The play and pause buttons were widely understood and allowed test takers to replay sections of the listening more than once.

4.2.2.3 Timer

To prevent test takers from artificially inflating their scores by listening to audio clips repeatedly, a Qualtrics timer was installed on the receptive intelligibility sections of the test. Once the predetermined time for each section elapsed, test takers were moved onto the next section of the test battery automatically. This timer was removed from the settings when it was discovered to be preventing test takers from completing the first sections of the test battery due to slow Internet speeds in their respective countries.

4.2.2.4   Randomization

Randomization is frequently used in computer-based testing in order to overcome any biases that result from the order in which items are presented. Qualtrics allows responses to be randomized, thereby avoiding sequencing effects (Roever, 2013) thus reducing the ability for test takers to cooperate or cheat during a test. Though the EIL test battery is a low-stakes test, the possible responses for all of the multiple-choice items were randomized as a matter of good test practice.

4.2.2.5   Saving and continuing the test

Qualtrics features a *save and continue* option that permits test takers to quit the test and later resume from the section they most recently completed by placing a *cookie* (a small recognition file specifically for the user) in the participant's computer. At the end of each completed test section, Qualtrics automatically saves the data entered to that point. Since the EIL test battery can easily take over an hour to complete, this feature was especially convenient for test takers busy with either the ELTO coursework or their working duties in their respective ministries.

4.2.2.6   Survey dispersal

The EIL test battery took advantage of the Qualtrics mailer, a built-in system for managing the dispersal of a survey or test to large email lists. The mailer sends a link to the EIL test battery, and as test takers click on the link, names, email addresses, test dates, and start/finish times are recorded. In order to ensure more complete data sets, reminder messages can be sent out to those who start but do not complete the test. Moreover, as each section of the test is completed, data is automatically saved, allowing test takers to return to finish the test on any computer.

The Qualtrics mailer was used to build email lists of past and present ELTO classes, most of who were former program participants. Next, the EIL test battery link was sent to 13 ELTO classes for a grand total of 411 potential test takers, over 120 of whom volunteered to take the test.

4.2.2.7 Data reporting

Final test results from the EIL test battery were initially downloaded from Qualtrics to a Microsoft Office Excel spreadsheet. The resulting in raw data contained extraneous biographic information, such as names, nationalities, languages spoken, times, dates, and email addresses. This data was copied and prepared for scoring by removing this non-critical data, generating a clean data set of test takers and responses to items. These responses where then converted to score values, culminating in a final scored Excel file which was then uploaded into SPSS (Statistical Package for the Social Sciences) for data analysis. The data analysis and validation of the test results are covered in Chapter 5.

4.2.3 Vocabulary control

To limit any language proficiency bias in the EIL test battery, efforts were made to limit the vocabulary load of the texts so as to allow test takers to focus on the constructs being elicited rather than unknown words in the texts. Therefore the EIL test battery text (Appendix A) was loaded into a vocabulary RANGE (Heatley, Coxhead, & Nation, 2002) program using the BNC COCA 25000 corpus) to determine the extent of the vocabulary load on the test takers.

The EIL test battery vocabulary profile (see Table 4.5) showed that 96.45% of the tokens in the text occur in the first 3,000 word families in the BNC lists. Note that the first 1,000 words cover nearly 83% of the lexical items in the text. This figure suggests that the text of the EIL test battery consists mostly of high frequency words in English occurring in the English language.

Table 4.5: Vocabulary profile of the EIL test battery

Frequency Level	Word Families	Raw Tokens (%)	Cumulative Tokens (%)
Top 1,000 words	527	82.85	82.85
Top 2,000 words	208	9.43	92.28
Top 3,000 words	106	4.17	96.45
Top 4,000 words	31	1.73	98.18
Top 5,000 words	13	0.30	98.48

According to Milton and Meara's (2003) research into matching vocabulary size with the Common European Framework of Reference for

Languages (CEFR) levels, test takers with a vocabulary size score of 2,750 to 3,750 ranked within the CEFR levels of B1 and B2. While linking CEFR levels and IELTS band scores can be challenging and complex, several internal and external research projects (summarized in Taylor, 2004), approximates CEFR levels of B1 and B2 to an IELTS band score of 4 to 6.5. The EIL test takers recorded an overall average band score of 5.45, placing their vocabulary knowledge well within the lexical demands of over 96% of the test.

There are just 35 words from the EIL test battery which are above the most frequent 5,000 words. Examples of words outside the top 5,000 words include browser, euros, queue, razor, idioms, fasting, podium, lice, turban, Ramadan, and backhoe. These words were not cut from the text since they were directly relevant to the context of the question.

4.2.4   Test scoring

This section will describe the process of scoring for the EIL test battery. Scoring can be split into two categories: automatic and human rater. Automatic scoring refers to the programmed scoring of items within the online test battery, while human rater scoring was used only for the productive intelligibility (speaking) section of the test. These two scores are added at the end of the test for an overall score of EIL abilities.

These EIL abilities are captured by the four constructs: receptive intelligibility, productive intelligibility, strategic competence, and intercultural competence. Although the number of items for each construct varies considerably (from 10 to 40 items), the score weighting was balanced equally between each construct. The underlying argument here is that no single construct is valued more than the other, and that these combined abilities work together for successful international communication. Together the constructs are scored as four equal components in the test battery, weighted at 60 points each for a grand total of 240 points.

4.2.4.1   Automatic scoring

Scoring for the receptive intelligibility, intercultural competence, and strategic competence construct items was completed automatically within the

online platform. Qualtrics allows scoring for a matrix table, multiple-choice items, and text entry items. In the final EIL test battery, all but five items were multiple-choice items. Assigned score values for correct answers were essentially a test maker decision and programmed into the test battery accordingly. Table 4.10 summarizes how points were designated for each component of the test battery.

Upon completion of the EIL test battery, Qualtrics provided test takers with an online test score out of 180 points as well as an overview of each item and the correct answers. The productive intelligibility (speaking) score was calculated separately by human raters, and added to a test report form that was sent to participating test takers.

4.2.4.2 Human Raters

The assessment of the productive intelligibility required the involvement of human raters to evaluate the spoken component of the test battery. Three PhD students from the School of Linguistics and Applied Language Studies were enlisted, in addition to the researcher, to serve as raters. The raters (described in greater detail in Appendix D) were selected for the following reasons: first, they are fluent English language users, experienced in using English in international contexts. Second, they have all taught English for several years and therefore experience assessing spoken language. Third, these raters have an understanding of intercultural communication, having lived and worked in at least two or more countries. Fourth, they are well-educated adult users of EIL. Fifth, they are either bilingual or multilingual, and arguably have a deeper appreciation of intercultural and international communication. Finally, these raters were selected as representatives of EIL users from their region, which includes Europe, Africa, Asia, and North America. Section 4.2.5.3 explains the details of the scoring system for productive intelligibility.

4.2.5 Receptive intelligibility
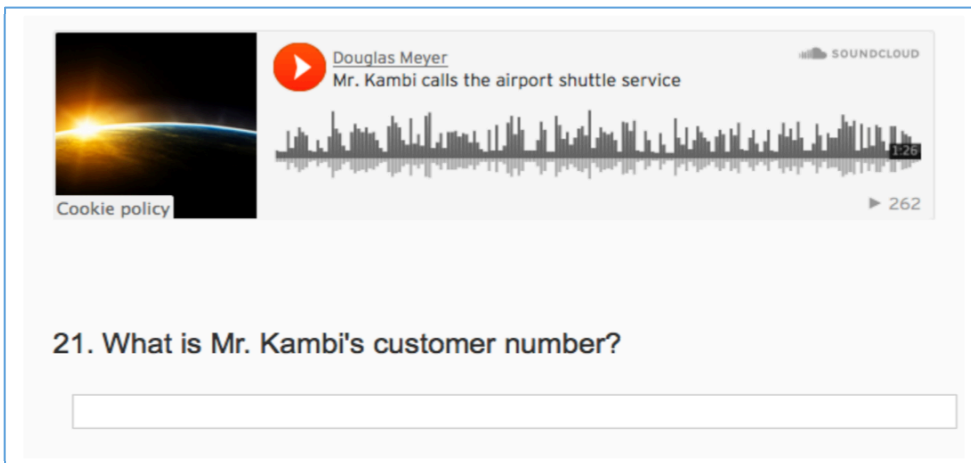
The receptive intelligibility section is essentially a listening test consisting of several items aimed at eliciting awareness of various EIL accents on three increasingly challenging levels: word/utterance recognition, comprehensibility,

and interpretability. The following are examples of these test items, and how they were operationalized and scored.

4.2.5.1   Examples of receptive intelligibility test items

Word/utterance recognition was measured in both isolated and contextualized terms. For example, in distinguishing between [s], [sh], and [th] phonemes, test takers would listen to an EIL user utter one of the following words in isolation: bash, bass, pass, path. Test takers only needed to click on the word they heard, choosing from the list of four similar words (see Figure 4.4). Just 5 of the 25 items for receptive intelligibility were given a context in the form of a very short conversation, after which items elicited dates, numbers, and time. The data entry field and accompanying audio clip are shown here in Figure 4.3.

Figure 4.3: An example of a contextualized receptive intelligibility item

Five comprehensibility items followed a three-minute long speech by a fluent EIL user from Kenya talking about disaster management. Test takers were allowed one minute to read the questions before listening, and had to answer five multiple-choice items.

Example: What is the alarming trend Mr. Kambi is talking about?
  1. The lack of disaster management.
  2. Unprecedented damage caused by tropical storms.
  3. There are more and more disastrous storms.
  4. Countries in the region holding a disaster management conference.

The operationalization of interpretability (defined as the understanding of indirect English) saw the creation of multiple-choice items with matching audio clips. Test takers had to listen to ten brief dialogs containing indirect language use and then identify the second speaker's intended meaning. The context was given as a convention center social event, where several diplomats and guests were discussing the rather rude and blunt speech delivered by "Mr. Kambi" earlier in the day on disaster management.

Example: At a dinner table, two diplomats are talking about Mr. Kambi's speech. Listen to the audio clip above. What is speaker 2's intended meaning?

[audio only]
  Speaker 1: *That was quite a speech, yes? How is your chicken?*
  Speaker 2: *It's as tough and overdone as Mr. Kambi himself.*

  1. He likes his chicken well done.
  2. He doesn't like either the meal or the speech.
  3. He thinks it is important to prepare a better dinner for the conference.
  4. He thinks Mr. Kambi is a tough man.

4.2.5.2   *Operationalization of receptive intelligibility*
While Qualtrics supports the upload of nearly any audio or video file, individual web browsers are less flexible. For improved security and reliability, the audio clips were uploaded to the audio website SoundCloud, which insures the widest compatibility with web browsers. SoundCloud was used to host 32 audio clips on the survey platform Qualtrics. Once embedded, the audio clips

(most of which were less than 10 seconds) were formatted as multiple-choice questions as show in figure 4.4. Test takers were able to listen more than once, and selected responses with a click of the mouse on the correct radio button.

Figure 4.4: An example of a word/utterance recognition online audio clip
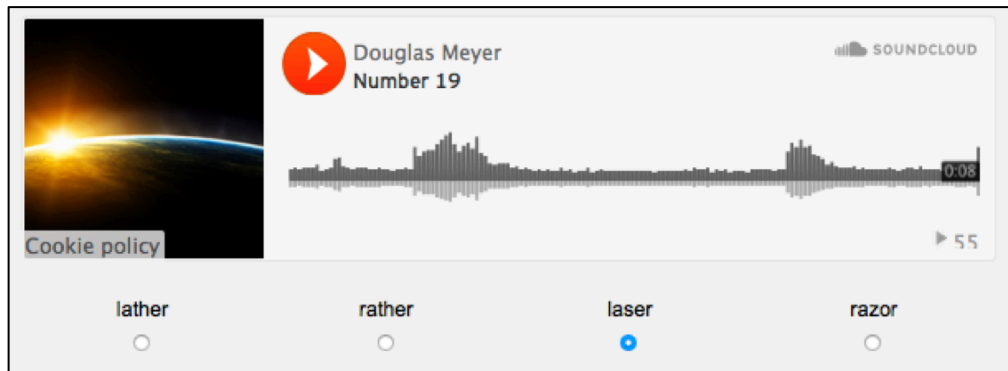
While listening comprehension tests are certainly not new, the approach used in this test is unique. The voice samples provided come from a wide range of voice actors from various linguistic backgrounds, representing the accent-rich field of EIL users. Native-speaker voice samples constitute the minority of these audio clips (25%), which is arguably more representative of global EIL use, if the language demographics provided by Crystal (2003, 2012) and Graddol (2007) are accurate.

4.2.5.3   Scoring receptive intelligibility

The receptive intelligibility component of the EIL test battery was scored automatically by the Qualtrics testing platform as 40 dichotomous items. The first sub-construct (word/utterance recognition) was split into two parts. The first part is as shown in Figure 4.4 above. The second part was a brief listening comprehension section whereby test takers were required to enter dates, times, and numbers into a text field for scoring. The structure of the receptive intelligibility construct is outlined in Table 4.6.

Table 4.6: Receptive intelligibility test structure and scoring

Sub-construct	No. of items	Scoring	Maximum raw score
Word/utterance recognition	25	20 items at ½ point each +5 items at 1 point each	15
Comprehensibility	5	3 points each	15
Interpretability	10	3 points each	30
			Total: 60 points

The receptive intelligibility construct has a far greater number of items that the other constructs, resulting in the need to balance this section score with the other sections of the battery in mind. To this end, it was decided that each construct would be equally weighted. Since the word/utterance recognition sub-section had far more items (20 in total) than the other sub-sections, it was necessary to score each one as a half-point. The more complex receptive intelligibility items (comprehensibility, interpretation) were worth more points, due to their limited number and increased difficulty. As discussed in section 4.2.4, each construct score sub-total worked out to the even number of 60 points, and this is reflected in the equal weighting.

4.2.6   Productive intelligibility

Since the Qualtrics platform is unable to record test taker voices, productive intelligibility was a separate and optional component of the EIL test battery. However, the instructions for completing the productive intelligibility section were provided online, explaining how test takers could submit a voice sample for assessment by themselves.

Productive intelligibility is essentially a speaking test divided into three levels by adopting the 1992 Smith Framework . Word/utterance recognition is a measure of articulate and distinct pronunciation of single words out of context, while the section on comprehensibility provides context for the test taker to speak for about one minute on a familiar topic. The final sub-construct, interpretability, asks test takers what they might say in certain awkward situations using English indirectly. The following sections describe some

example items, how they were operationalized, and the human rater scoring system involved.

4.2.6.1   *Examples of productive intelligibility items*

First, word/utterance recognition elicited the ability to pronounce similar-sounding English words clearly and distinctly. Using a voice recorder and procedures outlined in section 4.2.5.2, test takers had simply to say the each of the isolated words in order as instructed.

Examples:	1.	path	bass	bath	pass
	2.	face	pace	base	vase
	3.	took	tug	dug	dog [...]

A total of ten of the above items containing four similar-sounding words each were spoken and recorded by the test taker, which is estimated to take less than a minute. At this point, test takers should pause their recording and prepare for the comprehensibility section.

Second, the comprehensibility sub-construct was measured by giving test takers a list of topics to talk about for about one minute. Topics included familiar themes, such as their country, job, culture, or past experience with a foreign V.I.P.

Example:

Talk about your country. Where is it? What is your hometown like? What do you like (or not like) about your country? How is the weather there this time of year? What is most interesting about your country?

These questions served as more of a guideline to provide direction for test takers to elicit a sufficiently long voice sample for assessment. Scoring (see section 4.2.5.3.) did not take into account the accuracy of the responses to each of these questions, but rather the overall intelligibility of the English produced.

Third, the interpretability sub-construct asked test takers to use English indirectly and appropriately in a specified context. For each detailed context, a brief response was requested, preventing run-on explanations or diversions from the test task.

Example:

> You are manager of your department with 24 people working under you, some of who are your personal friends. Last month, the head of the ministry conducted a work performance review of all the departments. You were informed that one of your staff (and a close friend) must be fired for poor work performance. You call him into your office to tell him the bad news.

Objective: As the department manager, it is part of your job to fire staff. What do you say to this staff member?

He walks in and says, "Hey, what's up? What did you want to talk about?"

You say " . . . . "

After reading the above context, test takers should resume their recording and provide a brief 2-3-sentence response for the given situation.

4.2.6.2 *Operationalization of productive intelligibility*

Test takers were asked to use their computer or smart phone recording functions to record their responses to the three sub-sections of intelligibility. Following instructions onscreen, test takers recorded their introduction by opening with:

"Hello. My name is … from the country…. I'll begin now." After each sub-section of productive intelligibility, test takers should pause the recording until they were ready to speak for the next section. By pausing the recording between each section, test takers created a single audio file, which is then emailed to the researcher for assessment.

4.2.6.3   *Scoring of productive intelligibility*

As discussed in section 4.2.3.2, human raters were used to assess the three sub-components of productive intelligibility. This necessitated the development of three different scoring rubrics, which can be found in Appendix E. Following two rounds of calibration and discussion, the voice samples provided for productive intelligibility were evaluated and recorded for each test taker. The sum of the scores for each rater was averaged to create a final comprehensive score for productive intelligibility. The test structure and specifications of productive intelligibility are outlined in Table 4.7.

Table 4.7: Productive intelligibility test structure

Sub-construct	No. of items	Scoring	Maximum raw score
Word/utterance recognition	10	2 points each	20
Comprehensibility	1	20 points total	20
Interpretability	5	4 points each	20
			Total: 60 points

The decision was made to balance each component of the Smith Framework equally, resulting in a final score for 60 points for productive intelligibility. Each set of similar-sounding words is scored for pronunciation and distinction, with 0 being poor, 1 moderate, and 2 acceptable – as judged by the four raters from Africa, Asia, North America, and Europe. The same raters also assess the comprehensibility voice sample, making judgments on fluency, pronunciation, and ease of understanding. Accuracy errors were not penalized, unless they impaired intelligibility, answering some calls for accommodations (discussed in section 2.3.2.2) whereby communicative competence is prioritized over accuracy. The interpretability section assessed five items in terms of the appropriate use of indirect language and degree of success in achieving the task objective in given contexts. Raters scored test takers on a scale of 0 to 4 according to the scoring rubric found in Appendix E.

Finally, the scores from each of the four raters were averaged to minimize any regional bias and provide as fair an assessment of intelligibility as possible. Moreover, the raters made comments for each sub-section, noting any salient points with regard to pronunciation, fluency, or appropriate language use. This

feedback can be made available to the test takers in the form of a score sheet or test report form.

4.2.7   Intercultural competence

Intercultural competence was defined using the Fantini (2009) framework (see sections 3.4.2 and 3.4.3) and takes the approach of using intercultural incidents to elicit cultural knowledge as a component of EIL abilities. It is argued that government officials need not only *knowledge* of other cultures, but also the ability to take *action* in the event of a critical cultural incident. It is believed that the attributes listed in the Fantini framework are best suited for eliciting the kinds of cultural knowledge and interpersonal attributes ideal for EIL interactions.

The intercultural competence section of the test battery is measured by ten critical intercultural incidents from around the world. Upon reading each critical incident, the test taker was required to first recognize the cultural incident described. Next, they were asked which is the best way to show intercultural competence from four possible solutions, only one of which was most interculturally appropriate (see section 4.2.7.3 on how correct answers were determined). This approach is the most practical in terms of assessment as part of an EIL test battery, although some authenticity will be sacrificed.

4.2.7.1   *Examples of intercultural competence items*

Ten critical intercultural incidents were developed to serve as test items for this section. A situation was developed, describing a context in which a critical cultural incident occurs, placing the test taker in the first person role. An example is shown below:

You have just started a new job in an international workplace, and have co-workers from all over the world. One Friday evening, your co-workers invite you to join a large group of them going to a Brazilian restaurant bar near your office. Upon arrival, you see a middle-aged lady (the owner) greeting all the customers with a single kiss greeting on the cheek when customers arrive. The owner now turns to you with open arms.

What could you do to BEST show your intercultural competence?

1. Give the owner a quick hug and pat on the back.
2. Smile politely but keep your distance.
3. Extend your hand for a handshake when she comes near.
4. Learn to relax and accept it, even if it makes you a little uncomfortable.

Since diplomatic work often involves a mix of personal and professional contexts, the items are a mixture of professional, business, and personal domains. The above example highlights a case of flexibility and possibly cultural dissonance (overcoming differences) if the test taker is not from a 'hugging' culture. While other answers are indeed possible in this context, according to the Fantini framework, response number four show the best flexibility and openness. Details of the operationalization are discussed in the following section.

4.2.7.2   Operationalization of intercultural competence

These ten items were entered into Qualtrics as descriptive paragraphs summarizing an intercultural incident. To minimize the vocabulary demand (see section 5.2.5) and reading load, each paragraph was limited in length, averaging 75 words per incident. After reading each incident, test takers had four standard multiple-choice items from which to choose as the best way to show intercultural competence.

4.2.7.3   Scoring of intercultural competence

Qualtrics scored these ten items automatically as dichotomous items as shown in Table 4.8. In the beta test version, scoring was divided into two values: six full points for what was considered the best way to show intercultural competence, and three points for the option which was considered *somewhat correct,* although less than ideal. However, item analysis of this construct showed that the beta test takers were frequently selecting the best possible response. Therefore the decision was made to make the somewhat correct responses more *incorrect*, and to score intercultural competence dichotomously.

Table 4.8: Intercultural competence structure

Sub-construct	No. of items	Scoring	Maximum raw score	
Fantini framework	10	6 points each	60	
			Total:	60

4.2.8 Strategic competence

The operational definition of strategic competence was outlined in section 4.1.6.4 as the ability to repair broken conversation by means of clarification requests, requests for repetition, confirmation requests, and direct appeals for help. To develop this concept into measurable test items, scenarios were drafted in international contexts that involve imaginary EIL users experiencing a communication breakdown. The types of communication breakdown were framed around typical problems faced by EIL users, such as not understanding rapid speech, unfamiliar accents, unfamiliar idiomatic expressions, and forgotten lexical items.

The next step in the development process led to the recordings of ten brief dialogs as a means of adding authenticity to the items. Moreover, the voice actors used for these recordings (and throughout the test battery) were a representative mix of fluent English speakers from a variety of linguistic backgrounds. But in order to make a multiple-choice format work for the information gap type of problem planned, it was decided to add the script of the dialog in addition to the audio. An example can be seen in the following section.

4.2.8.1 Examples of strategic competence items

Each of these items was led by a SoundCloud audio clip and followed by the text of the brief audio clip. Very brief background information was provided to give context to the exchanges that followed.

Example:

A work site manager is showing a visiting project leader the construction site of a new school.

Mrs. T: As you can see, the workers are ahead of schedule.

Mr. C: Why are the workers using shovels instead of a backhoe?

Mrs. T: . . . .

Mr. C: It's a machine with an engine and a very large shovel.

What can Mrs. T say to best clarify this conversation? Make sure your answer fits both the context and Mr. C's response.

1. What do you mean by a backhoe?
2. What I wanted to say is that we will finish early.
3. Can you say that again?
4. I don't understand you.

Test takers were able to both read the script and play the audio clip of the conversation before selecting the most appropriate response to repair the miscommunication.

4.2.8.2 *Operationalization of strategic competence*

Each item of this section had three components:, an audio clip, a matching script, and four multiple-choice responses from which to choose. Test takers had the option of either listening to the audio clip, reading the script (or both) after which they had to select the best response.

4.2.8.3 *Scoring strategic competence items*

Qualtrics automatically scored strategic competence items as dichotomous items worth six points each. Unlike intercultural competence, the responses available were not debatable in terms of which is correct or not.

Table 4.9: Strategic competence structure

Sub-construct	No. of items	Scoring	Maximum raw score
Strategic competence	10	6 points each	60
			Total: 60 points

As with intercultural competence, it was necessary to score the items in such a way as to balance the sub-totals equally. Ten items at six points each was the most practical way to achieve this balance without reducing the practicality of the test battery. The addition of more items and text will add time to the strategic competence component of the test battery, and result in fewer test takers completing the test.

4.2.9   Summary of test battery operationalization

Chapter 4 presented two major stages in the test development process: the design statement and the operationalization of the EIL test battery. The first section answered questions regarding *what* and *who*: the purpose of this test, who the test is for, and the definitions of the theoretical constructs. The second section provided concrete details about *how* these constructs are measured, including some of the technical aspects that help build the foundation of the online test battery. These components of the test battery are summarized in Table 4.10, showing each construct of EIL abilities, brief construct definitions, the scoring, and which sections are parts of the online test.

Table 4.10: Summary of the EIL test battery structure

Construct	Construct definition summary	Online	Items	Scoring
Receptive intelligibility	A listening test measuring familiarity with a wide range of EIL accents.	☑	40 items	60 points (computer scored)
Intercultural competence	Multiple-choice test measuring best means to resolve a critical cultural incident.	☑	10 items	60 points (computer scored)
Strategic competence	Multiple-choice test measuring best means to repair conversation.	☑	10 items	60 points (computer scored)
Productive Intelligibility	A speaking test measuring clarity, pronunciation, and indirect usage.	☒	16 items	60 points (human raters)

4.3   Plan for evaluating usefulness of test battery

The final section of this chapter will present the plan for the overall usefulness of the EIL test battery. Bachman and Palmer (1996) call this essential element of test development "the most important consideration in designing and developing a language test..." (p. 17) and outline six qualities of test usefulness:

construct validity, reliability, authenticity, interactiveness, impact, and practicality. The strengths and weaknesses of these six qualities determine how useful this test is for its intended purpose. Overall test usefulness is a collection of both empirical evidence and persuasive logical arguments that bear on the meaning of test scores. Rather than any single argument or piece of empirical evidence, it is the collection of a *variety of evidence* that supports test use and interpretation of test scores. The following sections summarize the plan to determine the degree to which the EIL test battery is useful for its intended purpose.

4.3.1 Construct validity

In general terms, constructs are the underlying abilities or attributes that a test attempts to measure. Construct validity concerns the appropriateness of the *interpretations* made based on test scores. Construct validity is defined by Bachman and Palmer (1996, p. 21) as "...the extent to which we can interpret a given test score as an indicator of the ability(ies), or construct(s) we want to measure." Messick adds that construct validity is about "...determining the degree to which certain explanatory concepts or constructs account for performance on the test" (1989, p. 16). Messick (1989) argues that these categories of evidence should be merged into a single persuasive claim that argues that a test serves its purpose, a persuasive claim which is backed by empirical evidence and theoretical rationales. Each form of evidence supplements and supports the overall construct validity, not of the test per se, but of the validity of the *inferences* drawn from test scores. Messick (1989) is also noted for introducing the concept of *consequential validity* to the testing world by asking test developers to consider all the potential consequences of testing (e.g. teaching to the test, admission and placement of students), bolstering the claim that a given test does indeed serve the purpose for which it was intended – and nothing else.

4.3.1.1 Plan for construct validity

The plan for the construct validity of the test will follow the Bachman and Palmer (1996) framework by answering key questions concerning for logically evaluating test usefulness. Table 4.11 answers these questions in terms of what is planned to demonstrate construct validity for the EIL test battery.

Table 4.11: Validation of the test construct

Questions for evaluation	Extent to which test quality is satisfied	Explanation
Are the constructs clearly defined?	Satisfied	Theoretical constructs are defined in consultation with supervisor, EIL users, and the literature.
Are the constructs relevant to the purpose of the test?	Satisfied	The test tasks
Do the test tasks reflect the construct definition?	Satisfied	The tasks elicit the theoretical knowledge and skills defined as EIL abilities.
Will the test scores assist with interpretations about the test taker's EIL abilities?	Uncertain	As a theoretical construct, it is uncertain if the test scores are a direct reflection of true EIL abilities.
Are there any variations in test setting that may cause test takers to perform differently?	None	The online test experience should be similar as test takers complete the test at home at their leisure.
Are there any variations in the scoring rubrics that may cause test takers to perform differently?	None	The raters are calibrated and the test takers provide voice samples under their own conditions from home.

4.3.2 Reliability

Reliability is a test quality that aims at consistent and dependable test results. Tests which are very reliable provide consistent conditions for administration on every occasion, provide uniform rubrics for scoring and evaluation, utilize a consistent application of those rubrics by the raters, and lack ambiguous items or tasks (H. D. Brown & Abeywickrama, 2010).

An empirical measurement of internal consistency in Classical Test Theory is the coefficient alpha (Cronbach, 1951). Cronbach's alpha (as the coefficient alpha is commonly known) is defined as the degree of interrelatedness of a set of items within a test (Cortina, 1993; Schmitt, 1996). Cronbach's alpha (or α) is the most frequently used and reported measure of reliability (Fulcher, 2010). Furthermore, Cronbach's alpha is superior to the

split-halves reliability estimate in that it represents an average of all possible split-halves. Finally, Cronbach's alpha lends itself to both dichotomously scored items as well as continuously scored variables. These three points make Cronbach's alpha the best (but not only) estimate of internal consistency for the 60 online test items.

The reliability for the spoken section of the test battery (productive intelligibility) will involve inter-rater reliability, which examines the variability in using human raters. The plan to determine this variability among the four raters will use Krippendorff's alpha, an estimate of reliability "…for judgments made at any level of measurement, any number of observers, with or without missing data."

With regard to consistency in the test setting, some variation will occur as test takers complete the test online at a time and place of their choosing. Background noise, audio quality, Internet speed, and other distractions may affect the results of the test, which is not delivered in a controlled environment. Variations in input involve only one brief five-item sub-section of the test, where the test takers must manually type in their responses. All other forms of input involve clicking a button from a multiple-choice list of possible answers.

4.3.3 Authenticity

Authenticity is a test quality used to demonstrate that the performance on a test corresponds with the real world domain that the test aims to simulate. Bachman and Palmer (1996, p. 23) define authenticity as "the degree of correspondence of the characteristics of a language test task to the features of a TLU [target language use] task". However, due to the need for controlled testing environments, it can sometimes be challenging for tests to be authentic. In general, authentic tests offer tasks that simulate real-world tasks, use natural language, have items which are contextualized as well as isolated, and include topics meaningful to the test takers (H. D. Brown & Abeywickrama, 2010). This connection between test task and the domain of the target language is also part of the test construct validity, and strengthens the claims that can be made from interpreting test scores.

In order to account for authenticity, the EIL test battery should draw from

128

both the classroom observations and the online survey of the target language domain users. The classroom observations may provide some authentic samples of communication breakdown and repair, while the online survey should show the ways in which the test takers used English in their workplace. Together, these bottom-up approaches help to align the test tasks with the target language contexts. Additionally, the intercultural items were crosschecked with people from the cultures concerned in order to improve the authenticity of this section of the test battery. However, as an *online* test battery, the test of EIL will have moderate to low authenticity in terms of real world interaction. A computer-based test is no substitute for live person-to-person interculturally negotiated communication. Faced with limited resources, test developers often have to sacrifice authenticity for practicality.

4.3.4 Interactiveness

Another quality of tests is interactiveness, which Bachman and Palmer (1996, p. 25) define as "…ways in which the test taker's areas of language knowledge…topical knowledge, and affective schemata are engaged by the test task." Bachman and Palmer give the hypothetical example of typists who cannot use English well, but nonetheless have developed the skill to read written documents and type them up accurately. A typing test eliciting this skill might have high authenticity, but low interactiveness, since the typists are not required to understand the contents of the documents. In other words, it is a realistic typing test that does not engage the typists' language ability, topical knowledge, or affective schemata. They likely don't know what they are typing in detail, nor are they engaged deeply in the task.

Bachman and Palmer (1996) note the following interesting points to consider about interactiveness: first, interactiveness is relative to authenticity, in that test tasks can be more or less authentic, more or less interactive, rather than absolute terms. Second, test developers must consider test taker characteristics and make *educated guesses* in initial attempts to make test tasks authentic and interactive. Third, despite best efforts to make tests authentic and interactive, the test takers themselves may process test tasks in different ways that may not be anticipated by test developers. Interactiveness is a test quality linked closely

with construct validity, and a good test developer will seek to optimize the levels of interactiveness in the early stages of test development. Interestingly, Bachman and Palmer (1996) note that some test tasks may be relatively useful (for the intended purpose) despite having low authenticity and interactiveness.

The tasks of the EIL test battery seek to elicit the demonstration of four types of knowledge: receptive intelligibility, productive intelligibility, strategic competence, and intercultural competence. Language ability is assumed. The metacognitive strategies engaged with each component of the test vary. For example, in the intercultural competence section, test takers will first need to recognize that an intercultural incident (see section 4.4.7) has occurred, and then apply their knowledge of culture in order to best rectify the situation. The productive intelligibility section engages metacognitive strategies on a deeper level, as test takers are required to produce indirect polite language in a socially-awkward situation (e.g. being evasive). The strategic competence section also elicits a wide range of metacognitive strategies from the test takers as they aim to choose the best method of conversation repair. However, other sections of the test battery are less engaging. For example, the receptive intelligibility section simply requires the aural recognition of various similar-sounding words, which does not deeply elicit test taker's metacognitive strategies.

4.3.5   Test impact

The impact of a test, or the degree to which it affects both individuals and social/educational systems, is yet another important quality of a test. In extreme cases, tests have been used historically to exclude or detect specific individuals (McNamara, 2000). Fortunately, modern day testing is more considerate of the impact of testing, focuses on ethics and social responsibility as guiding principles to ensure that tests are carefully constructed, reliable, and fair.

As an exploratory test, the EIL test battery is a low-stakes test that is not intended to have an impact on the test takers, since it is voluntary and unrelated to their language course. As such, the test will have no impact upon the ELTO teachers or their language course at the university. That said it is possible some of the intercultural incident items may have the unintended consequence of upsetting test takers. Care should be taken to avoid this by checking the

sensitivity of the items (and test takers) with ELTO teachers who know the test takers best.

4.3.6 Test practicality

Practicality is concerned with the down-to-earth points of test preparation, administration, and analysis. It involves more than just the practical and efficient use of time and money. Practicality is focused primarily on test implementation, and whether or not it is worthwhile to see a test through to full development and operationalization. According to Brown and Abeywickrama (2010), a practical test has clear directions for administration, efficiently utilizes human resources, and can be completed within the time constraints. Brown and Abeywickrama further contend that practical tests take into consideration the effort involved for test design and scoring, as well as any budgetary limits one must operate within. Bachman and Palmer (1996) regard practicality as a relationship between the demands of a test and existing resources, concluding that if the test demands do not exceed limited resources, the test is practical and further test development is warranted.

The practicality of the EIL test battery can be planned for in three categories: time, material resources, and human resources. The primary demand on any limited resources will be time. While the administration of the test is highly practical and allows immediate data collection and scoring for the online portion, Roever (2013) notes that online tests "…despite their advantages once operational, the initial implementation of computer-based testing systems tends to be…time-consuming" (p. 217). The second drain on resources will be for the compensation and time necessary to train and calibrate the human raters who listen and assess voice samples. Skilled assessors from a variety of linguistic backgrounds must be recruited to this end. The final component of practicality is material resources. As an online test, most of the material resource responsibility falls to the test takers, who must have access to a computer, Internet connection, and ideally a set of headphones. These resources are commonly available in most (if not all) homes and offices worldwide. However, it should be noted that stable and reliable access to the Internet varies in developing countries, so effort should be made to limit the broadband speed

131

necessary to access the online test, especially with regard to data-heavy images or audio files.

4.3.7   Considering the assessment use argument

More recently, Bachman (2005) and Bachman and Palmer (2010) developed a validity framework called the *assessment use argument* to further justify test use. The assessment use argument can be used to bolster claims of validating tests by providing detail and structure to the consideration of how the EIL test battery might be used and the potential impact of the test battery. This process facilitates communication amongst stakeholders (such as faculty, administrators) with little experience in educational assessment (Johnson & Riazi, 2016) by providing more information about test utilization and impact, equitability, and consideration of test consequences.

The assessment use argument is based on the validity work of Kane (1992, 1994) and Kane, Crooks, and Cohen (1999). Kane's approach was used in the investigation of high-stakes international English-language assessment tools such as the TOEFL (e.g. Chapelle, Enright, & Jamieson, 2010, 2004, 2008), IELTS (e.g. Ebadi, Zamani, & Sharifi, 2014), and the Pearson Test of English: Academic (e.g. Wang, Choi, Schmidgall, & Bachman, 2012). The AUA also supports Messick's call for "…an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores" (1989, p. 13). In other words, the AUA is a theoretical framework for providing a rationale and a set of procedures for justifying the intended use of an assessment tool.

The components of the AUA theoretical framework consist of four claims about a test by addressing the following elements of test planning and design. The first claim concerns test consequences, and asks test developers to foresee potential negative consequences (e.g. anxiety, negative washback) that might affect the stakeholders. The second claim concerns specific decisions made as a result of testing, such as who will be responsible for making decisions based on test results and how the stakeholders might be affected. The third claim from the assessment use argument concerns the interpretations drawn from test scores, and that such interpretations are meaningful, fair, and otherwise reflect

educational and societal values. The final claim concerns the consistency of test results, and asks test developers to ensure the test is scored the same way across different administrations of the test. This is usually a statistical analysis of test results, which generates inter/intra rater reliability estimates.

In addition to the four claims of the AUA, *warrants*, *rebuttals*, and *backing* will be given to each claim in order to shape the validity arguments that support test use. Warrants are essentially statements that elaborate on the test claims. Like a legal argument, rebuttals refer to counterclaims that suggest the warrants have not been carried out (i.e. the test does not do what it claims to do). Finally, backing is provided that either shows evidence to support the warrants (i.e. feedback from test takers, observations, interviews, previous research, statistical analysis) or forces the test developer to acknowledge test weaknesses and restate claims about what the test can or cannot do.

However, while the assessment use argument is a helpful coherent framework for the test developer in both assessment justification and test development (Bachman & Palmer, 2010), the assessment use argument is not necessarily a suitable framework for the EIL test battery. First, the EIL test battery is an exploratory assessment tool under development, and as such, any claims that can be made about the test are quite limited at this stage. Test takers will be volunteers in this research, and therefore test results will have no *consequences* (claim 1) on their coursework at the university. Second, no *decisions* will be made as a result of interpreting test scores (claim 2). In other words, the EIL test battery results will not help with decisions about test taker placement in courses, passing grades, or hiring decisions. A third reason the assessment use argument is not suitable for the EIL test battery (at this stage) has to do with the *interpretations* made from test scores. According to Bachman and Palmer (2010), interpretations about an ability should be meaningful to a syllabus or target language domain, impartial to all test takers, generalizable, relevant for decisions to be made, and provide sufficient information for decisions to be made. Such interpretations are better suited to a test that is not an exploratory assessment tool but rather a more formal testing situation such as in an educational institution.

Chapter 3 outlined a conceptual model for EIL abilities and a framework for assessing it. But as a conceptual model of international communicative competence, it lacks the ability to make the kind of concrete claims and rationale used in the assessment use argument approach. This research is interested in the *qualities* of this exploratory assessment tool, and the degree to which it is useful at its present stage. Therefore in terms of determining a plan for test usefulness, the Bachman and Palmer (1996) six test qualities approach is more suitable for the EIL test battery.

4.4 Chapter summary

This chapter has outlined the design statement and operationalization of the EIL test battery, as well as a plan for determining its overall usefulness. The design statement clearly set out the purpose of the test, and whom it was for. A description of both the test takers and their domain and task types laid the foundation for a definition of the four constructs of EIL abilities. The operationalization covered the instrument design and its numerous technical aspects, such as scoring procedures, timing, and data collection. Examples of the operationalization of each construct demonstrated how the constructs would look as items in an online test, and how the test takers were expected to respond to the format. A plan for validating the test was established using both the Bachman and Palmer (1996) six qualities of test usefulness approach as well as the assessment use argument approach (Bachman, 2005; Bachman & Palmer, 2010).

Chapter 5 will serve as the validation chapter, first discussing the validity of the EIL test battery in terms of the six test qualities (construct validity, reliability, authenticity, interactiveness, test impact, and practicality) followed by the complete assessment use argument in support of test use. This dual approach will build a thorough validity framework for determining how useful this test is for its intended purpose.

Chapter 5: Empirical Validation

<div style="text-align: right; font-size: 3em;">5</div>

While Chapter 4 outlined the design statement and operationalization of the EIL test battery, the goal of this chapter is to examine the results of the EIL test battery using Bachman and Palmer's (1996) six qualities of test usefulness: construct validity, reliability, authenticity, interactiveness, impact, and practicality. The strengths and weaknesses of these six qualities determine how useful this test is for its intended purpose

After reviewing the results of beta test piloting, this chapter presents evidence of test usefulness by addressing the overall construct validity and empirical reliability of the EIL test battery. Next, the authenticity, interactiveness, impact, and practicality of the test is reviewed, followed in each section by a brief discussion of the evidence gathered so far. Finally, an overall summary of test usefulness is given, addressing the overall usefulness of the EIL test battery.

5.1 Pilot testing

An essential part of any test cycle is reviewing a test to ensure the tasks selected can be justified and that the resulting test scores do indeed help to make valid inferences to the constructs developed (Fulcher & Davidson, 2007). The EIL test battery went through both an alpha and a beta process not only to identify poorly performing items, but also because it was a computer-based test delivered online. Links, embedded audio tracks, and download times can become common problems for online tests if not piloted thoroughly.

5.1.1 The alpha pilot test

The main purpose of the alpha test was to ensure the functionality of the online test battery. Four PhD student colleagues from the School of Linguistics and Applied Languages volunteered to complete the alpha version of the EIL test battery in late May 2015. These four alpha test pilot volunteers (originally from Europe, Africa, East Asia, and Southeast Asia) were selected for their experience in teaching language, living abroad, and fluency in English. Moreover, they are well-educated professionals who use English as an international language,

making them well suited to serve as representatives of EIL users. Three of these four PhD students would later become raters for the speaking component of the test battery.

At this stage, the test battery consisted of an introduction for gathering bio-data, a receptive intelligibility (listening) section, an intercultural competence section, and the strategic competence section. It was also uncertain as to how the productive intelligibility (speaking) component might be added to this online test. The speaking component is essentially an offline component that requires test takers to record themselves and email their voice sample (as an audio file) for assessment by four raters.

In the alpha pilot phase, the receptive intelligibility (listening) section had only eight items, and test takers had to distinguish one word out of three possible answers. The intercultural competence and strategic competence sections had ten questions, a number that remained unchanged for the beta and final test versions. The overall focus of the alpha phase of testing was on the technical elements (prompt downloading of audio clips, scoring, timing, entering text responses) performing smoothly for test takers.

5.1.1.1   Summary of changes in the alpha test

The four alpha testers completed a feedback form, which served as an initial trouble-shooting guide for improving the initial online test battery. The alpha testers provided feedback on the quality of the tasks, clarity of wording, and facility level of tasks for each section. In addition, they checked the clarity and general intelligibility of the initial voice samples to be used in future versions of the test. Suggestions from the alpha testers included time allotments for each section, identifying items that were too easy or difficult, and changing the format of input buttons on screen. The experience gained from running the alpha test version correctly (technology-wise) resulted in few major changes for the beta test version because the alpha test generated insufficient data for major changes to be made to the items and tasks themselves.

Based on the results from trial runs and feedback from the alpha version testers, an average time could be calculated for the receptive intelligibility section. The intention behind adding a timer was to prevent the test takers from

136

listening to the audio clips multiple times and artificially inflating scores. The online platform timer was thus programmed to automatically move the test taker on to the next part once the section time elapses. This timer impaired complete data collection later, as discussed in section 7.2.2.

5.1.2   Beta pilot test summary

The main purpose of the alpha version of the EIL test battery was to uncover any *technical* problems resulting from hosting the test online. A single beta round of testing was aimed at uncovering problems with the test items themselves, with the goal of eliminating or changing test questions that were far too easy, hard, or distracting. Eleven test takers from various Southeast Asian countries (Cambodia, Vietnam, Myanmar, Laos) and Peru participated in this beta pilot test. The Beta test takers are volunteers from the English language training for officials program described in section 4.1.5. As a result of the low numbers of test takers for the beta test, a meaningful SPSS data analysis of the results was unrealistic. Instead, a more basic item analysis was performed in order to identify items that were performing poorly. The item analysis consisted of three calculations: the facility index, a distractor analysis, and item discrimination.

5.1.2.1   Facility index

The average facility index indicates the proportion of test takers who answered items correctly. Ideally a facility index of around .50 is ideal (Fulcher, 2010), showing that an item was balanced between being too easy or too hard. Henning (1987) suggested that a facility index range of approximately .30 to .70 is acceptable (p.50) for most given populations of test takers, while Farhady (2012) suggests a range of between .37 and .63 for four-option multiple choice items. However given the small sample size of test takers, these ranges are used as a guideline rather than a steadfast rule.

The average item facility index for each construct is summarized in Table 5.1, showing acceptable values for intercultural competence and strategic competence. However, the 8 items from the receptive intelligibility construct show a facility index of 0.771, indicating that those items were generally too

easy. As a result of this, the decision was made to increase the number of distractors from three to four, thus increasing the difficulty level. Other items in the beta pilot test that were correctly answered by 80% of the test takers were considered too easy, and modified or cut. The same holds true with any items that were answered correctly by less than 20% of test takers. Overall, the facility index of the beta test is acceptable at .604, coming well within the ranges suggested above. Examples of items from the beta test that proved to be too easy or difficult for the test takers are summarized here in Table 5.1.

Table 5.1: Facility indices of poorly performing beta test items

	Item no.	Facility Index	Test task
Test items too difficult	Q2.3	.125	Distinguishing 'bath' from path, bass, and pass
	Q2.27	.133	Listen and enter customer number '797854'
	Q4.7	.103	Comprehension item showing a spending increase
	Q4.21	.103	Comprehension item highlighting indirect language use
Test items too easy	Q2.11	.938	Distinguishing 'alive' from arrive, arise, and alas
	Q2.19	.828	Distinguishing 'packs' from fax, vax, and backs
	Q2.25	.964	Comprehension item: entering a customer number.
	Q7.11	.790	Strategic competence clarification item

5.1.2.2 Discrimination index

The item discrimination calculation assists with one of the many purposes of testing, which is to discriminate between more knowledgeable and less knowledgeable test takers. Each test item should ideally help separate test takers in this regard, meaning that items answered correctly by test takers in the higher scoring group should tend to be answered incorrectly by test takers in the lower scoring group. Although there is no set range, Farhady (2012) suggests that a result of .40 and above is acceptable for item discrimination. The closer the item discrimination value is to +1.0, the more vigorously the test item performs at distinguishing between upper and lower scoring test takers. The final item analysis was hampered in the pilot test by the low number of test takers.

As shown in Table 5.2, the beta pilot test does a poor job of discriminating between higher and lower ability test takers. The mean discrimination indices vary from .060 to .320, with the overall beta test battery averaging a discrimination index of .214, far below the minimum of .40 recommended by Farhady (2012). The beta version of the intercultural competence construct produced the lowest discrimination index of .060, meaning the items failed to separate higher and lower ability test takers to any meaningful degree.

Table 5.2: Example of distractor analysis

Option	Response	Correct response rate
A.	What do you mean by a backhoe? *	91%
B.	What I wanted to say is that we will finish early.	9%
C.	Did you say "back home?"	0%
D.	Which workers are you talking about?	0%

In this case, 91% (or 10 of 11) test takers easily determined that the correct answer was A). Similar results from the distractor analysis were used to identify other foils that were underperforming, such as responses C & D above. Foils that did not strike a better balance were modified either through simple re-wording of the distractor or replacing the distractor in its entirety.

The reasons for the low discrimination indices throughout the test are most certainly due to the low number of beta test takers (n=11) and the low number of test items for each construct. At this stage of test development, it would have been better to seek a bigger number of test takers from outside of the target test taker population, regardless of their proficiency, age, or profession. This would allow for greater statistical analysis of the results and generate more useful data for evolving the test battery.

5.1.2.3 *Distractor analysis*

The EIL test battery is made of primarily of multiple-choice questions, consisting of contextual information and 3-4 response options (one of which is correct). The distractor analysis determines the degree to which certain options were selected as answers. The incorrect responses are called distractors (or

foils), and a common challenge in test development is to create distractors which are neither clearly incorrect, clearly irrelevant, nor problematic.

In the beta test, 27 of 49 items were identified as problematic in the distractor analysis, where the foils did not attract any responses, resulting in several items that were far too easy. As with all multiple-choice testing, the correct answer is present among the foils, and correct answers simply "…require recognition, not active recall or construction of the correct answer as productive items types would. This format facilitates guessing…" (Roever, 2004, p. 293), and in fact the test takers may not truly know what the correct response is. However, in the item example from the strategic competence section, it is clear that the foils lacked the ability to distract neither low ability nor high ability test takers.

Q7.7   A work site manager is showing a visiting project leader the construction site of a new school.

Mrs. T: As you can see, the workers are ahead of schedule.
Mr. C: Why are the workers using shovels instead of a backhoe?
Mrs. T: _____ .
Mr. C: It's a machine with an engine and a very large shovel.

What can Mrs. T say to best clarify this conversation?   Make sure your answer fits both the context and Mr. C's response.

One of the drawbacks with computer-based multiple-choice testing is that such requests for clarification are not normally encountered in writing or text display. In such a format, recognition of correct responses might be artificially easy when compared to actually producing language in spontaneous spoken discourse. In an attempt to accommodate for some degree of authenticity, the strategic competence component of the test included audio clips of the dialog as well as the script.

The beta test results are summarized in Table 5.3, outlining the basic structure of the online beta test, the number of items, facility indices, and discrimination indices for three components of the EIL test battery.

Table 5.3: Online beta test results

	Receptive intelligibility	Intercultural competence	Strategic competence	Overall beta test
Number of items	29	10	10	49
Average item facility index	0.771	0.565	0.682	0.604
Average item discrimination index	0.250	0.060	0.320	0.214

If the beta pilot test had 25 to 35 test takers, more of the problematic items might have been discovered and fixed at this stage, prior to the final version of the test battery. Nonetheless, even the basic classical item analysis tools of item facility and item discrimination indices were useful in achieving the goal of identifying major problems with certain poorly performing items prior to the launch of the final test battery version.

5.1.2.4 *Summary of changes in the beta test*

The beta pilot resulted in a number of changes, the first of which was to increase the number of items in the receptive intelligibility section from 8 to 20. These brief items had equally brief accompanying audio clips of about 8 seconds each, so the additional items did not increase the total test time noticeably. This step also addressed a threat to validity by increasing the coverage of the construct (receptive intelligibility) being studied. In addition, to compensate for overly high facility indices, a fourth option was added to the list of potential answers to each question.

The facility index and discrimination index analysis identified underperforming items throughout the beta test version, leading to minor word changes in 24 items and the removal and replacement of 4 items. In addition, a large number of slight technical adjustments were made to the online test battery, including the timer, font display, scoring, and instructions for each test section.

Although the productive intelligibility (speaking) construct was added as an optional component at the end of the online test, none of the test takers completed this section. This is likely due to time constraints because at this stage of test development, the online test was already taking approximately 80-100 minutes to complete (much longer than the estimate of 1 hour). The

productive intelligibility section explained how test takers could record the voice samples and email their voice samples in as audio files. Even though the speaking section was offline, having it included in the online test battery was a good decision, rather than having it as an orphaned construct, separate from the other test components. The choice to make the speaking component optional was a concession given in to practicality and time constraints.

5.2   Validation of the final EIL test battery

The following sections will examine the usefulness of the EIL test battery by collecting evidence of validity. After the item analysis of the final test, evidence of test validation is presented using Bachman and Palmer's (1996) six qualities of determining how useful a test is for its intended purpose: construct validity, reliability, authenticity, interactiveness, test impact, and test practicality.

5.2.1   Item analysis

The facility indices and discrimination indices for the final EIL test battery are reported in table 5.4. When compared to the item analysis of the beta test (Table 5.3), one can see the decrease in the facility index across the constructs, meaning the final test battery has become slightly more difficult (as intended). The overall facility index of the test battery has dropped from .604 to .501, falling into the middle of Farhady's (2012) recommended facility index range between .37 and .63 (for four-option multiple choice items). This drop is most likely due to the addition of extra foils in the receptive intelligibility section and the rewording of other items initially found to be too easy.

Table 5.4: Facility and discrimination indices for the final test

	Receptive Intelligibility	Intercultural competence	Strategic competence	Overall online test battery
Number of items	40	10	10	60
Average facility index	.511	.453	.510	.501
Average discrimination index	.113	.212	.357	.170

N = 69

With the decreasing facility index came an unexpected drop in the ability of these items to discriminate between higher ability and lower ability test takers. The overall average discrimination index fell from an already low of .214 to .170, resulting in very low discrimination efficiency among the items. Based on Ebel's (1972) guidelines, discrimination indices of 0 to 0.19 are items that poorly discriminate between higher and lower ability test takers (and need revision), while indices ranging between .2 and .29 are acceptable; indices from .3 to .30 are good, and any value > .4 is excellent.

With this in mind, looking at the discrimination indices in Table 5.3 above construct by construct, the items from strategic competence were the most efficient at discriminating, with an index value of .357. This may be due to the more linguistic nature of strategic competence as operationalized within the test battery. Unlike intercultural competence, the correct responses for strategic competence are less vague and more a matter of correct grammar, vocabulary, and perhaps even pragmatics. Intercultural competence had a discrimination index of .212, making those items barely acceptable according to Ebel's guidelines above. However, cultural knowledge, personal choices, or even guessing might have been responsible for answer regarding the best way to show intercultural competence in this test section. The low discrimination index of the receptive intelligibility section are most likely due to the fact that the timer was removed from the online test, enabling test takers to listen to each short audio clip any number of times in order to eventually answer correctly

5.2.2   Construct validity

Three of the four constructs of EIL abilities being assessed by the online test battery (receptive intelligibility, strategic competence, and intercultural

communicative competence) were analyzed by SPSS bivariate Pearson correlation to determine the strength of the relationship between these variables. Looking at Table 5.5, one can observe some moderate degree of inter-relation between these three sub-constructs, accounting for less than a quarter of each other's variance. The strongest relationship exists between strategic competence and intelligibility, whereas a slightly weaker relationship exists between strategic competence and intercultural competence. The weakest relationship is between intercultural competence and intelligibility.

Table 5.5: Correlations between constructs in EIL test battery

Construct		Intelligibility	Intercultural competence	Strategic competence	Total
Intelligibility	Pearson correlation	1	.202	.382**	.629**
	Sig. (2-tailed)		.107	.002	.000
	N	65	65	65	65
Intercultural competence	Pearson correlation	.202	1	.473**	.737**
	Sig. (2-tailed)	.107		.002	.000
	N	65	65	65	65
Strategic competence	Pearson correlation	.382**	.473**	1	.879**
	Sig. (2-tailed)	.002	.000		.000
	N	65	65	65	65
Total	Pearson correlation	.629***	.737**	.879**	1
	Sig. (2-tailed)	.000	.000	.000	
	N	65	65	65	65

**. Correlation is significant at the 0.01 level (2-tailed).

While these three constructs are different, they are based on the principle that successful international communication requires a test taker to be familiar with the wide range of accents in spoken English, being able to be widely understood him/herself, a degree of intercultural competence to enhance interactions, and knowing how to resolve communication breakdowns. It is quite possible that while EIL users may have considerable knowledge about other cultures, the inability to express these ideas may not result in satisfactory intercultural outcomes. Likewise, it is possible an EIL user may understand a variety of spoken English accents, but due to first language influences (or a lack of awareness on the side of the listener) have low intelligibility. The weak correlation between intercultural competence and intelligibility is

understandable, since the former elicits active behaviour in resolving intercultural critical incidents while the later is more of a passive listening skill.

5.2.2.1 *Convergent and discriminant validity*

Convergent and discriminate validity are essentially sub-categories of construct validity, and are used to provide evidence to support the construct validity of the EIL test battery. Theoretically related constructs should show a 'high' correspondence (convergent validity), while constructs that theoretically should not be related should show a 'low' or negative correlation coefficient (discriminant validity). While there is no hard and fast rule for what constitutes 'high' and 'low' in these cases, test items that are related show a higher correlation than test items that are dissimilar. If test items between two different constructs show high inter-correlations, it is likely that they are related to the same construct.

The following section will highlight a few cases showing convergent and discriminate validity and briefly discuss possible reasons for any irregularities. The inter-item correlation matrix of the test items revealed rather random results, with fluctuations and occasional negative values, even within the same construct.

Table 5.6: Inter-item correlation matrix for receptive intelligibility

Item	Q2.7	Q2.10	Q2.13	Q2.14	Q2.16	Q2.18	Q2.19	Q2.21
Q2.7	1.000	**-.101**	.112	.091	.138	-.093	-.045	.071
Q2.10	-.101	1.000	-.071	.066	.087	.131	-.023	.245
Q2.13	.198	.416	1.000	.365	.238	.380	.253	.249
Q2.14	.117	.091	.365	1.000	.503	.649	**.577**	.445
Q2.16	.080	.080	.238	.503	1.000	.482	.459	.277
Q2.18	.201	.037	.380	**.649**	.482	1.000	**.562**	.542
Q2.21	.181	.224	.249	.445	.277	.542	.522	1.000

With results as shown in table 5.6 above, only a few clear cases of convergent and discriminate validity can be seen. Item Q2.18 shows convergence with both items Q2.14 and Q2.19 despite the fact that the phonemes being tested are actually quite different. Item Q2.14 aims to distinguish the [p], [b], [f], and [v] phonemes while item Q2.18 aims to distinguish [th] and [t]

phonemes. The evidence of convergence between items Q2.14 and Q2.19 is much more intuitive, since the identical phonemes [p], [b], [f], and [v] are being tested. Item Q2.18 is also showing some convergence with item Q2.16, a [r] [l] distinction item. Items testing similar constructs should tend to converge, while different constructs tend to diverge, suggesting item Q2.18 is problematic. Table 5.7 shows convergence involving one item (Q2.18), which is clearly different from the other two items.

Table 5.7: Convergence between three items in receptive intelligibility

Item	Possible responses				Construct
Q2.14	pace	base	face*	vase	[p], [b], [f], and [v]
Q2.18	thee*	tree	she	three	[th] and [t]
Q2.19	fax	packs*	vax	backs	[p], [b], [f], and [v]

* Indicates correct response

Discriminant validity can be seen in table 5.6 above between items Q2.7 and Q2.10, with a low correlation between items measuring [ch] and [th] and [t] [d] respectively. This confirms that these items are not necessarily related and do in fact discriminate between the different phonemes tested.

Returning to Table 5.5, the convergent and discriminant validity between all of the constructs can be seen. Convergence can be seen between strategic competence and intercultural competence, correlating at .473, showing a moderate degree of relationship between these two constructs of EIL abilities. To a lesser degree, a similar relationship can be seen between strategic competence and receptive intelligibility, with a correlation of .382. The moderate relationship between intercultural competence and strategic competence is understandable since both constructs involve two separate problem-solving abilities: intercultural and linguistic. The intercultural competence construct involves the recognition and resolution of critical cultural incidents, requiring some knowledge about other cultural norms. The strategic competence construct involves linguistic knowledge of the best way to resolve communication problems, and thus shares a stronger relationship with intercultural competence. The weaker correlation (.202) is between intelligibility (i.e. listening and distinguishing different phonemes and accents)

and intercultural competence. These two constructs naturally shared a weaker relationship since they are the least alike of the three constructs. A more in-depth examination of the relationship between the constructs will follow in the Discussion Chapter.

Discriminate validity was a problematic issue in the intercultural competence section of the EIL test battery. The items were aiming to capture the same construct (i.e. the ability to recognize and solve intercultural problems), however the inter-item correlation matrix (Table 5.8) highlights the numerous low and negative correlations, meaning that these items are not related or do not relate to each other highly. Upon analysis, it was discovered that each of these items was indeed testing a different construct: specific cultural knowledge about different cultural norms. Discriminate validity within the intercultural competence section are shown in Table 5.8 with item Q6.8 and Q6.11 examples:

Q6.8: You are visiting a small Pacific island on vacation. One part of your trip involves a 3-hour ferry trip to another island. When you arrive at the ferry terminal, you see there is no line (or queue) for tickets. Instead, there is a noisy crowd of people around the ticket booth, making you a little uncomfortable. You check your watch and see that the ferry is leaving in 10 minutes.

What could you do to BEST show your intercultural competence?
   a) Wait for the crowd to quiet down and then get a ticket for a later ferry.
   b) Walk into the crowd and make sure you buy your ticket. *
   c) Find another way to get to the other island.
   d) Push your way to the front of the line and get your ticket.

Item Q6.8 above was intended to elicit the ability to overcome personal discomfort and achieve an objective such as getting a ferry ticket. This item is substantially different from item Q6.11 below, which elicits cultural knowledge of weddings in India. Item Q6.11 aims to capture specific knowledge that weddings in India last several days and are quite time-consuming to organize. This item also entails the recognition that the character Mrs. Singh will be

147

extremely busy preparing for the wedding, and thus extending the project deadline is the best face-saving option.

Q6.11: You work in an international workplace for a growing company in India. At a meeting, one of the local co-workers (Mrs. Singh) is asked if she can complete a project by August 1st. She says yes, she understands that it can be finished by then. After that, Mrs. Singh mentions that her daughter will be getting married in July to a man from a nice family.
What could you do to BEST show your intercultural competence?

a) Ask her if she would rather give the project to a co-worker.
b) Assume that everything is fine and the project will be completed on time.
c) Set a new deadline so that it is possible for Mrs. Singh to finish on time.*
d) Remember to buy something nice (such as a coffee maker) as a wedding gift.

The negative correlations between the above two examples and the remaining items can be seen in Table 5.8. As a result of the numerous negative and low correlation values, the Cronbach's alpha value was initially a negative value of -.439, meaning the internal consistency for the intercultural section of the test battery requires substantial revision.

Table 5.8: Intercultural competence inter-item correlation matrix

Item	Q6.2	Q6.3	Q6.4	Q6.5	Q6.6	Q6.8	Q6.9	Q6.10	Q6.11
Q6.2	1.000	.019	.092	-.024	-.035	.105	.000	-.076	-.214
Q6.3	.019	1.000	.008	-.134	-.024	.172	.139	-.016	.011
Q6.4	.092	.008	1.000	.129	.062	-.008	-.109	-.081	.017
Q6.5	-.024	-.134	.129	1.000	.057	-.175	-.015	.079	.096
Q6.6	-.035	-.024	.062	.057	1.000	-.051	-.019	-.008	-.020
Q6.8	.105	.172	-.008	-.175	-.051	1.000	.092	-.169	-.231
Q6.9	.000	.139	-.109	-.015	-.019	.092	1.000	-.262	-.091
Q6.10	-.076	-.016	-.081	.079	-.008	-.169	-.262	1.000	.101
Q6.11	-.214	.011	.017	.096	-.020	-.231	-.091	.101	1.000

5.2.3   Concurrent validity

Concurrent validity is another form (albeit it a weak form) of evidence commonly used to both defend a test as well as predict outcomes in similar validated tests of related constructs. Concurrent validity is shown when a similar test correlates 'well' (.70 and higher) with an existing similar test which has been previously validated.

While the IELTS English language proficiency exam is not exactly a similar test to the EIL test battery, it is a well-known validated test of listening, reading, writing, and speaking. Proficiency in English was an assumed indicator of success in the EIL test battery, since the greater the language ability, the greater the linguistic resources test takers can tap in responding to test items. While the EIL test battery does not attempt to measure language proficiency, language proficiency remains an assumed ability of EIL users.

 To uncover any connection between reported IELTS proficiency band scores and results in the EIL test battery, a Spearman's rank-order correlation determined that there is no statistically significant relationship (r_s = -0.017, p = 0.903). In addition, a one-way ANOVA confirmed that there was no statistically significant difference in the mean final test scores between the four most common IELTS test bands (Bands 5, 5.5, 6, and 6.5) of the test takers, $F(3,48)=0.81$, p=0.970. The ANOVA assumptions were met, with constant variance confirmed using Levene's test (p = 0.273) and normality assessed visually with a Q-Q plot. This analysis means that regardless of test takers' IELTS proficiency bands between 5 and 6.5, their score on the EIL test battery was essentially the same.

It was expected that, to some degree, test takers with a higher language proficiency would score higher on the EIL test battery, mirroring similar work on speech acts and implications (Roever, 2004). The strategic competence and intercultural competence sections of the test involved a heavy reading load, an area where language proficiency might affect outcomes. It appears that the efforts made to control vocabulary (section 5.2.5) were beneficial in minimizing the vocabulary demands of text takers to the degree that language proficiency did not influence final test scores.

5.2.4   Content validity

One way of collecting evidence for content validity is to involve recognized subject matter experts, who determine the relevance of test content compared to the real-world domain of interest. These experts also consider the degree to which test tasks and items are representative of the same domain in an item and task selection process. Messick cautions that a clear distinction be made between content validity as *judgmental evidence* "…rather than evidence in support of inferences to be made from test scores" (1989, p. 17) but goes on to argue that evidence of content relevance and representativeness be supported by *additional* categories of validity (i.e. criterion-related validity, consequential validity, and especially construct validity).

Lawshe (1975) takes an empirical approach towards quantifying content validity using a formula called the content validity ratio. Lawshe proposes that if more than half of subject matter experts consider a test item to be essential to the skill or knowledge of the construct, the item has some content validity. The content validity formula $CVR=(n_e-N/2) \div (N/2)$ yields a value ranging from -1 (unnecessary) to +1 (essential), which aids in the selection of relevant and representative test tasks and items. Content validity is ultimately estimated as the average CVR across the test as a whole.

While this attempt to quantify content validity is noteworthy, it is arguably very impractical for three reasons. First, one needs to assemble a panel of experts. Next, calibration and discussion is required to set levels for how essential, useful, or necessary a test item is. Finally, a robust debate should follow the introduction of each item in order to quantify the level of agreement amongst panelists. This is a very time-consuming process for all involved, and therefore reduces the practicality of a test. Therefore the content validity for the EIL test battery draws from two approaches: the traditional approach of subject matter experts, and drawing evidence for test tasks from a bottom-up study of the target-language domain (see Appendix B and C).

5.2.4.1 Expert input into test task/item development

For research into low-stakes experimental testing, expert opinion was sought from fluent well-educated EIL users and educators from within Victoria University of Wellington. On three occasions, discussions were held with groups of Ph.D. students and language educators to review and debate the relevance and representativeness of several test tasks and items. Each of these item drafts was then taken to the supervisor of this thesis for discussion and revisions.

Content for the intercultural competence items of the test were first drawn from both intercultural websites from the Internet and informal discussions with international Ph.D. students at the university. Draft items were confirmed for cultural content accuracy with university Ph.D. students from the respective cultures in question. In this way, expert opinion was gathered to support the content validity of this test battery.

5.2.4.2 Bottom-up approaches: Online survey and classroom observations

The first bottom-up approach used to ensure the content validity of this test was to gather information from the target-language use domain. This was completed in two ways: first, an online survey (Appendix B) was used as a bottom-up approach to elicit English language use at work directly from the test takers. The information gathered from this survey was used to shape the context for many items in the EIL test battery. Typical tasks described by the participants included work project proposals, writing reports, accompanying foreign guests around the city, meetings, and handling phone calls in English. Topics identified by the survey participants included international conferences, climate change, territorial disputes, and world health issues. Sixty-eight government officials (over two-thirds of who were from Cambodia, Laos, and Vietnam) responded to the online survey on English in the workplace in February of 2015. A summary of the main findings is given in Table 5.9.

Table 5.9: Online survey of test takers

English use at work	*n*	Percentage
Reading emails, letters or reports	47	69%
Speaking with foreign guests	37	54%
Translating documents	37	54%
Speaking in meetings		43%
Speaking with business people	27	40%
Planning projects	26	39%
Comprehension strategies		
Aim for general meaning rather than try to understand every word	43	63%
Clarification	34	50%
Rephrasing	33	49%
Speaking slowly	32	47%

n = 68

This survey did not inquire about the respondents' intercultural competence. At the time of this online survey, the government officials did not identify culture as a noticeable problem in their EIL communications during informal interviews. In retrospect, with intercultural competence identified as a core construct of EIL abilities, the online survey should have attempted to gather data on how the respondents felt about this aspect of international communication. It is possible that the respondents were unaware of any intercultural incidents in their EIL communications, or that they simply let such incidents pass without response. Unlike most people, these government officials may have had intercultural sensitivity training, enabling them to minimize creating intercultural incidents and learn not to respond to such incidents when faced with them. Extra survey items would have identified if this were truly the case or not, and should have been included.

The second bottom-up approach consisted of classroom observations of the test takers while improving their English abilities at the university. The focus of the observations (see Appendix C) was to uncover authentic cases of miscommunication and how such miscommunications were resolved. Nine different classes were observed in interactive activities such as pair work, split readings, and interclass Asian-African meetings in late 2014. However, due to the learner-friendly atmosphere of the classroom, incidents of miscommunication were rare. Nonetheless, numerous cases of EIL usage were

observed in a somewhat authentic setting. A summary of the classroom observations can be seen in Table 5.10.

Table 5.10: Classroom observations

Date	Class	Activity	Observation Summary
26 Sept.	Asian	Pair work	Passive listening
25 Oct.	Asian	Pair work	Passive listening
31 Oct.	African	Split reading	Back-and-forth negotiation for meaning
1 Nov.	Asian	Pair work	Back-and-forth negotiation for meaning
4 Nov.	African	Pair work	Laughing, gesturing, backchanneling
7 Nov.	African	Split reading	Overlapping agreement/disagreement
8 Nov.	Asian	Interclass	More clarifications, giving examples
14 Nov.	Joint class	Interclass	Asians used small gestures, Africans big gestures)
15 Nov.	Joint class	Interclass	Good comprehension, back-and-forth

5.2.5   Threats to validity

Threats to validity are factors that weaken evidence that a test is useful for its purpose by degrading construct validity. Minor common threats to validity include a noisy test centre, small sample sizes of test takers, and entering responses on a keyboard instead of on paper. Two major threats to validity identified by Messick (1989) are construct under-representation and construct irrelevant variance. The former simply means that a test may not have a sufficient number of items to measure a construct accurately, while the latter concerns the "...possible role of irrelevant variables contaminating test performance" (Messick, 1989, p. 19). These two threats will now be discussed, as they are directly relevant to the validity of the EIL test battery.

First, construct under-representation is felt to be an issue for the intercultural competence and strategic competence sections of the test battery, where only 10 items aim to elicit the desired skills. More test items on these constructs are desired, since more test items lead to better overall internal consistency and representativeness (or coverage) of the construct. However, a balance must be sought with practicality, whereby the test is completed in a timely manner by as many test takers as possible. Intercultural competence and

153

strategic competence were text-heavy sections of the test battery, and involved longer periods of time for test takers to read and respond. Thus the decision was made to limit the number of these text-heavy items in exchange for greater practicality and likely a greater number of responses. As with many test development decisions, one test quality is often sacrificed in exchange for another test quality. Of course, construct under-representation deals with more than an insufficient number of test items. Intercultural competence is a complex multi-dimensional construct involving cultural knowledge, motivations for different behaviors, and of context (gender, cultural setting, prior knowledge). At the present stage of development, the various aspects of intercultural competence are not covered sufficiently by the measure used, presenting a serious threat to the validity of this section of the EIL test battery.

Second, construct-irrelevant variance may have been an issue with regard to using an online computer-based testing platform. While Roever (2006) asked test takers to rate familiarity using computers and web-browsers, the EIL test battery assumed the government officials already had computer literacy. In addition to the test platform, the heavy reading load introduced especially by the strategic competence and intercultural competence test sections may have been a form of construct-irrelevant variance. In other words, these sections of the test may have been testing the skill of reading rather than the targeted skills of strategic competence and intercultural competence. Efforts to reduce this are discussed in section 7.3.

Construct-irrelevant variance was clearly an issue with the five brief items in the sub-section of receptive intelligibility where test takers needed to listen and type in the brief responses manually. This produced several responses with slight spelling errors (or typos), extraneous information, or insufficient information, making scoring a time-consuming manual task. While it is reasonable to believe that, as educated government officials, the test takers possessed computer literacy and typing skills, the format of this sub-section did not lend itself to the practicality and convenience normally associated with computer-based testing. If designed strategically, construct irrelevant variance can be addressed by taking advantage of test taker response times in order to

bolster validity claims by finding "...the optimal mix of items...to ensure reliability while maintaining practicality"(Roever, 2013, p. 229).

Another example of construct-irrelevant variance may have been in the form of test bias (also called differential validity). The Educational Testing Service (2007) provides item-writing guidelines against which items and tasks can be reviewed for bias, however the EIL test battery is a low-stakes research test, and it was not possible to assemble and fund the creation of a gender-balanced, culturally diverse review panel which must be trained and meet on a regular basis.

Test bias may have been an issue with the intercultural competence section, which was defined in section 3.3.2 as flexibility, openness, and patience with other cultures. Test takers from traditional or religious societies may have less flexibility with how they handle intercultural incidents, such as a man trying to shake hands with a Muslim woman, or a prompt German being patient with a late client. This test bias is arguably justified because a central underlying skill of using EIL successfully is the flexibility with which all manner of accents, varieties, and cultures are handled.

5.2.6   Consequential validity

Messick (1996, p. 251) defines consequential validity as "evidence and rationales for evaluating the intended and unintended consequences of score interpretation and use in both the short- and long- term...[and] unfairness in test use...". Consequential validity (also known as test impact) is considered to have three central aspects: test bias, washback validity, and effect on society (Weir, 2005). As a theoretical test model, the EIL test battery is expected to have little impact at this stage of development. It does exhibit degrees of test bias with regard to the intercultural competence section, and efforts were made to limit the vocabulary load for proficiency bias. However, washback (the effect of a test on instruction) is arguably negligible, although it is hoped to raise awareness of EIL among the test takers. In the future, as EIL becomes increasingly recognized as something that needs to be addressed in language testing, it is hoped to have a wider effect on society. But this is not likely as long as the native speaking countries hold sway over the major English language proficiency tests.

5.3   Reliability analysis

In order to statistically analyze the final test data for reliability, all of the final test data was downloaded from *Qualtrics* into a .csv-formatted file in Microsoft Excel. This raw data file was stripped of extraneous information and scored. Since the EIL test battery items ranged in score from ½ up to 6 points per item, the Excel file was scored as 1's and 0's to maintain consistency and accuracy. This Excel file was next uploaded into the Statistical Package for the Social Sciences (SPSS) version 23.0 for analysis of reliability.

Internal consistency is usually based on the correlations between different items on the same test, determining to what degree items that claim to measure the same construct result in similar scores. However, as shall be seen in the following section, internal consistency may be adversely affected by a test which is too short in length (Clark & Watson, 1995; Cortina, 1993; Schmidt & Embretson, 2013). As a result, test development requires a continual balancing act between maintaining sufficient internal consistency and overall test practicality.

5.3.1   Reliability for receptive intelligibility

The initial overall Cronbach's alpha for the internal consistency for all three subsections (word recognition, comprehensibility, and interpretability) of the intelligibility construct was estimated by SPSS to be .781. The Cronbach's alpha for all subsections of the test is summarized below in Table 5.11. While high reliability values are always desired, there is no *cut off* value for whether a test is reliable or unreliable (see Lance, Butts, & Michels, 2006). In the early stages of exploratory research Nunnally (1978) suggests that modest reliabilities of 0.70 will suffice (in contrast to 0.80 and higher reliabilities for higher stakes testing and research purposes).

The reliabilities were also estimated for each sub-construct of intelligibility. For the initial word recognition sub-section, with the deletion of the poor-performing items (Q2.10), Cronbach's alpha is re-estimated to be .763. Upon examination, this item was found to be a word spoken a little too quickly, with the speaker too far from the microphone to be clear. Further trimming of

items leads only to marginal increases in reliability. The next step is to examine the internal consistency for each sub-section of intelligibility. This is summarized in Table 5.11, which shows both the original Cronbach's alpha, and the improved value if poor-performing items are deleted from future versions of the EIL test battery.

Table 5.11: Internal consistencies of the sub-sections of intelligibility

Sub-section	Cronbach's alpha	If under-performing items are deleted	Resulting Cronbach's alpha
Word recognition	.751	Q2.10	.763
Word recognition (short answer)	.728	Q2.27	.750
Comprehensibility	.233	Q3.4 & Q3.5	.339
Interpretability	.394	Q4.21 & Q4.17	.453
Overall Construct	.781		

Possible reasons for under-performing items are discussed in the following section. Upon reconsideration, item Q2.10 and Q2.27 were found to be two examples of recordings where the speaker's pace was too fast. Items Q3.4 and Q3.5 (shown below) also proved problematic in that the former item involved somewhat indirect language use, while the latter involves distractors reversing the damage and number of casualties.

Q3.4 In which area does the government most need to invest, according to
    Mr. Kambi?
a) Updating weather stations.
b) Providing gear and training search and rescue personnel.
c) Building search and rescue centers.
d) Providing real-time flood information for citizens.

Q3.5   What does Mr. Kambi think is the largest damage caused by the recent tropical storm?

a) The storm resulted in $829 million in damage and 529 deaths.

b) The largest damage occurred the night before the storm.

c) The storm resulted in 829 deaths and $529 million in damage.

d) Public trust in the government.

Based on experience gained designing and developing this section, the following changes in a future test version are hoped to improve its reliability. First, all audio clips should be recorded in a soundproof room with a high-quality microphone. While all audio was checked for clarity, a high stakes test requires extra caution be taken to ensure sound quality is excellent. Second, I was not satisfied with the results from the short answer section of the word recognition sub-construct. Test takers had to type very brief responses into the response field, which became a form of construct-irrelevant variance (typos, extraneous information) as controlling the computer mouse while reading and listening at the same time became distracting. Moreover, the voice actor for this section spoke a little too fast, especially for item Q2.27.

5.3.2   Intercultural competence

The initial reliability analysis for intercultural competence generated a negative Cronbach's alpha (-.439), meaning that the internal consistency of this section of the test contained a high number of weak and negatively correlated items. This means that there was a problem with the construct as defined and operationalized, as there is greater *within*-subject variability than *between*-subject variability. Essentially, this section of the test performed inconsistently as a measure of intercultural competence, and deleting the weakest items made very marginal improvements in the overall reliability (see Table 5.12).

Table 5.12: Internal consistency of intercultural competence

Cronbach's Alpha	Number of items
-.439	10
-.153	9
-.029	8

Intercultural competence is a deep and abstract concept, and this test attempted to capture these skills with only ten items. Moreover, there are likely different dimensions of intercultural competence at work, dimensions that should have been defined and operationalized more clearly. The approach used to capture intercultural competence in this test was far too holistic, and should have identified at least a few dimensions with this construct, and target questions to elicit dimensions such as cultural knowledge, flexibility, and intercultural comfort level. The inconsistency of the intercultural competence section may be due to the fact that knowledge about one aspect of culture may be completely unrelated to another aspect of culture. For example, a test taker may know that the Sikh turban is not headwear that is simply removed like a baseball cap. But that knowledge is completely different from the knowledge that during the Ramadan period, Muslims eat only after sunset.

The unpacking of intercultural competence into specific dimensions might have increased the internal consistency. However, this would result in several items being added to this section of the test battery, adding to both the reading text load and time required to complete the test. The issue of improving the intercultural competence section of the test battery is discussed in greater detail in section 7.3.1.

5.3.3 Strategic competence

The internal consistency for the strategic competence section of the test is also low, but considerably better than for intercultural competence. With two of the poorest performing items removed, Cronbach's alpha rises from an original value of .366 to .483, a slightly better result at the cost of representativeness. This section of the test battery was much more linguistic in nature, asking test takers for the most appropriate response in order to clarify a communication

breakdown. Language proficiency was certainly a factor in this test section, as with the other text-heavy section of the test (intercultural competence), and the item discrimination index of .357 indicated that the items were acceptable, although ranges of .400 and above are desired (Farhady, 2012).

The strategic competence construct of EIL abilities performed better as a measurement tool due to the more concrete nature of language. In this construct, a more native speaker-like use of language emerges. EIL users would readily be understood saying "I don't know what it is" rather than the more linguistically accurate response of "What do you mean?" when seeking clarification of meaning. Strategic competence as measured in the test battery resembles discourse completion tasks, a popular instrument in speech act research. A comparison and discussion of this instrument in comparison to the strategic competence section of the EIL test battery will follow in the Discussion chapter.

Table 5.13: Internal consistency of strategic competence

Cronbach's Alpha	Number of items
.366	10
.430	9
.483	8

5.3.4 Productive intelligibility

The reliability for the spoken section of the test battery involves inter-rater reliability, which examines the variability in using human raters. In other words, it investigates the degree to which different raters agree amongst themselves on scoring for a given set of test performances. The raters (described in Appendix D) assessed 15 voice samples provided by the test takers. An analysis of the degree to which the human raters agree with each other in the scoring of test takers is a useful measure of reliability for this offline section of the test, due to the variability of using human raters.

With two raters, calculating the degree to which they agree on scoring is a straightforward matter. However, with four raters, a more complex statistical analysis was necessary to determine inter-rater reliability. Krippendorff's alpha

(see section 4.3.2) was estimated via the KALPHA macro, which can be downloaded for free from < http://afhayes.com/spss-sas-and-mplus-macros-and-code.html > (Hayes, n.d.).   The SPSS utility function then estimated Krippendorf's alpha as shown here in Table 5.14.

Table 5.14: Inter-rater reliability for productive intelligibility

Krippendorff's alpha	Number of cases	Number of raters	Number of decisions	Krippendorff's alpha if Rater 4 omitted
0.599	15	4	90	0.746

Huang (1984) suggests that raters can be treated like items in a test, and discarded if they deviate too far from the norm. By following Huang's approach, removing Rater 4 from the calculation will increase Krippendorff's alpha to a more satisfactory value of .746. Rater 4 might have been in need of more training or benefited from a longer discussion about the scoring rubrics.

Landis and Koch (1977) describe benchmark alpha values of .41 to .60 as moderate, and .61 to .80 as substantial indicators of reliability. Krippendorff himself argues that while there are no ideal alpha values for inter-rater reliability, obtaining $\alpha \geqq 0.800$ is a threshold for perfect agreement for high-stakes decisions, and $\alpha \geqq 0.667$ is the lowest conceivable limit for drawing tentative conclusions (Krippendorff, 2004, p. 429). By using Huang's (1984) approach to increase the Krippendorff alpha to .746, the inter-rater reliability is nearing what Krippendorff calls the .800 threshold for perfect agreement, which is more than sufficient for supporting test use and interpretation of scores for this section of the EIL test battery.

5.4   Authenticity

Authenticity is a test quality used to demonstrate that the performance on a test corresponds with the real world domain that the test aims to simulate. Bachman and Palmer (1996, p. 23) define authenticity as "the degree of correspondence of the characteristics of a language test task to the features of a TLU [target language use] task". However, due to the need for controlled testing environments, it can sometimes be challenging for tests to be authentic. In

general, authentic tests offer tasks that simulate real-world tasks, use natural language, have items which are contextualized as well as isolated, and include topics meaningful to the test takers (H. D. Brown & Abeywickrama, 2010). This connection between test task and the domain of the target language is also part of the test construct validity, and strengthens the claims that can be made from interpreting test scores.

5.4.1 Authenticity of intelligibility

The variety of accents used in the test is both realistic and representative of how EIL users interact, tapping the construct-relevant international intelligibility knowledge that this section of the test aimed to capture. The medley of voices and accents from a wide range of fluent EIL users is representative of real-world authentic English language use, unlike the limited inner-circle accents used in the big three language proficiency tests (i.e. IELTS, TOEIC, and TOEFL iBT). The voice actors who provided voice samples for the test battery are fluent English speakers coming from 17 different countries on five continents. Examples from this list include countries such as Canada, China, India, Kenya, New Zealand, Nigeria, South Africa, the UK, the US, and Vietnam.

However, the EIL test battery lacks situational authenticity because of the setting, which ideally should involve face-to-face interactions with other EIL users. This is one of the drawbacks of online testing, noted by Roever (2013, p. 230) who argues that "While comprehension and non-interactional productive abilities are arguably pre-conditions for online production, the test does not allow direct conclusions as to how learners would perform in real-world interactions. This is not necessarily a problem as long as inferences and subsequent decisions only concern non-interactional...abilities."

The productive intelligibility section of the test battery was arguably the least authentic component. As a productive skill, an interactive element would have made this section much more authentic, especially with regard to the use of indirect language. Although a context was provided for each scenario, only five items tested this sub-construct, which was under-represented. Moreover, the lack of authentic interaction made the test taker's response awkward and difficult to evaluate for the raters. The lack of face-to-face interaction can be

regarded as a form of construct-irrelevant variance that may have affected the meaningfulness and accuracy of scoring for this sub-section.

5.4.2   Authenticity of intercultural and strategic competence

The intercultural competence test section described ten critical cultural incidents, and asked test takers what they might do in such situations. The strategic competence section provided test takers with both an audio clip and a script of short dialogs in which there was a communication breakdown, and asked them which response would best repair or resolve the conversation. As with many multiple-choice tests, whether online or paper and pencil, a key problem is the balancing of test authenticity and practicality. These two concepts are often at odds, as the more realistic a measurement tool is, the more time and money are usually required to operationalize it. Measuring intercultural competence and strategic competence by means of an online multiple-choice test is no exception, and it is arguable whether or not such a test is an authentic measure of the constructs in question.

Stakeholders are indeed interested in the real-world performances of test takers. Although stakeholders want to know if a test is useful in predicting language use in social contexts, they should exercise caution when making consequential decisions based on online testing scores. While the topics and contexts were high in meaningfulness and representativeness of the constructs, the online format of the test may not be an accurate measure of real-world abilities, weakening the construct validity of the overall test battery.

5.5   Interactiveness

In section 4.3.4, Bachman and Palmer (1996) identified interactiveness as a test quality where language knowledge, topical knowledge, and affective schema are engaged by test tasks. The EIL test battery has an overall moderate to low degree of engagement with test takers for the following reasons: First, it is an online assessment tool rather than a more authentic person-to-person interactive test task. The reading load coupled with the multiple-choice format of the test is likely to be less engaging than a role playing task, for example. Second, the test takers are volunteers taking a test that is unrelated to

coursework, and thus their degree of engagement with each of the test tasks may be limited.

The first component of interactiveness is language knowledge. The average test taker's language knowledge is rated at the IELTS English language proficiency test band 5.5. This band level falls between a modest and competent language user, according to the IELTS website. Because efforts have been made to limit the vocabulary in the test (see section 5.2.5), the demand on test taker's language knowledge is wide-ranging, but not deep. The strategic competence section of the test battery draws primarily on strategic language use, making this construct more of a language knowledge section than the other constructs.

The second of the Bachman and Palmer definition of interactiveness is topical, or real-world knowledge, which forms the information base upon which one uses language with reference to the world in which we live. Topical knowledge is a challenging subject to address, since EIL users will have a wide variety of experiences and cultural backgrounds from which to regard the rest of the world. While the test battery aims to elicit responses from contexts based on personal experiences of the test takers (as revealed by the survey in Appendix B) topical knowledge coverage will always be an uncertainty.

The final component is affective schemata, or the degree to which a test taker engages with a language task in terms of similar past experiences. The affective schemata can determine to a large part a test taker's affective response to the test task, facilitating or restricting responses to items in various contexts. For example, sensitive test topics on territorial disputes, national pride, or war might result in strong emotions that affect responses either positively or negatively.

This point may have been easily overlooked, and it may explain some of the poor results generated by the EIL test battery. While topics for the test battery were drawn from a survey of test takers' English language use, this test was not a part of their standard coursework or evaluation. The test takers were uncompensated volunteers, and as such, may have not been fully engaged by the test tasks and responded without much thought or consideration.

5.6 Test impact

At the micro-level of test impact, the EIL test battery is expected to have little or no impact on the daily lives of either test takers or classroom instruction. The test battery is a theoretical test of EIL abilities, and inferences based on test scores do not impact on any decisions made regarding the test takers. Moreover, only the researcher and the supervisor, as well as the test takers themselves know the test scores.

However, future validated versions of this test battery may become useful for making decisions regarding candidates who are government officials working overseas or frequently involved in international matters. The current major language proficiency tests are essentially linguistic in nature, and do not cover the wider range of skills needed for successful international communication.

At the macro level, the intended uses and potential consequences of this test are quite limited. The EIL test battery is expected to have little impact on society, in its current state as an exploratory test. However, there is potential for the core concepts (intelligibility, intercultural competence, strategic competence) in this test battery to influence the current major English language proficiency exams in the future, especially with regard to intelligibility. The globalization of English is making it increasingly difficult for major English language proficiency exams to cling to native speaker standards (in listening sections) and it is here I hope this EIL test battery has the greatest impact.

5.7 Test practicality

Test practicality can essentially be summarized as the efficiency with which a test can be administered compared to the available resources necessary for its administration. The practicality of a theoretical test battery of EIL abilities will be discussed in three terms: time demands, materials required, and human resources.

5.7.1 Time demands

The initial time demand concerns test development. From the conceptualization to the operationalization stage, the time required for test design and development was by far the biggest factor affecting practicality. Once the constructs were identified, it was a long process operationalizing them into

practical test items in multiple-choice format. For example, intercultural competence draft items were checked and re-checked for authenticity and correctness by colleagues from the culture in question. Voice recordings and their upload onto an audio support website (SoundCloud.com) took considerable time, as well as formatting of a survey platform into a testing platform.

The second demand on time involves test administration. As with most computer-based testing, it is "...the initial implementation of computer-based testing systems [that] tends to be costly and time-consuming." (Roever, 2013, p. 217). Since efforts made to elicit as much construct-related information as possible within a single sitting of the test, the burden on test takers' time was set at 60 to 75 minutes. In retrospect, this was too ambitious and ultimately unnecessary, since the Qualtrics platform allowed the test takers to quit the test following each section and resume it at a later time. While this feature sought to make the test as practical as possible, it did result in some abandoned tests.

Finally, the time involved in scoring the test battery was minimal. The final test consisted of a grand total of 76 items, most (72%) of which were multiple choice, and effortlessly scored dichotomously by the computer. Five short answers in the word recognition section took time to score manually, as test takers often entered too much information, had spelling errors, or even unnecessary spaces which confounded computer scoring. This is a prime example of construct-irrelevant variance, which means a category of error stemming from test method effects. In this case, the data entry method of asking test takers to type in the short answer responses produced construct-irrelevant variance. As government officials, the test takers are quite familiar with the Roman alphabet keyboard layout since they reported regularly using English at work to write emails and reports. Nonetheless, the cognitive processes of simultaneously typing, thinking, reading, and listening to a second language may have been too much to manage for reliable data entry.

The other time-consuming element of scoring was the involvement of human raters for the productive intelligibility (speaking) section of the test battery. An average of six hours each was required for rubric discussion, calibration, and scoring the 15 voice samples. Smith's (1992) theoretical framework used for intelligibility was sound in reasoning, but unfortunately

complex for scoring purposes. For the small number of test takers who submitted voice samples for this construct of the test, a large amount of time was needed.

5.7.2 Material resources

The lack of a need for material resources for an online test is the greatest advantage of computer-based testing. It was not necessary to book a testing space, hire invigilators, or use traditional testing resources (paper, pencils, answer sheets, etc.). Essentially, test takers only needed a computer, access to the Internet, and perhaps a set of headphones to assist with the listening comprehension section. Test takers were able to complete the test on their own personal computers, which at the time of this thesis are essential for business and government professionals. For the test takers who provided voice samples, the voice recorder commonly available on most, if not all, personal computers was used. Others took advantage of the same feature on their personal Smartphone, and emailed the voice sample to the researcher as an audio file.

Other materials for the online test included *SoundCloud* and *Qualtrics* (see section 4.2), two freely available applications found on the Internet. The former served as an audio-hosting site for the numerous short audio clips that were embedded into the online test. The latter, Qualtrics, served as the platform from which this test battery was launched and sent to test takers. SPSS, for statistical analysis, is also freely available to most researchers.

5.7.3 Human resources

Human resources (in addition to the researcher) were limited to the enlistment of colleagues to serve as raters for the spoken section of the test battery. Occasionally the opinion of the raters was sought when drafting items and creating scenarios, but this was quite limited. Four raters each dedicated six hours of time to the discussion of scoring rubrics, calibration of scoring, and rating 15 voice samples provided by test takers. Again, technology was beneficial, defeating the need for test administrators and clerical support in the administration of this test.

5.8   Summarizing the six test qualities

The six qualities examined in order to determine the usefulness of the EIL test battery are summarized in Table 5.15, followed by a final statement regarding test usefulness. In collecting evidence to support the use of the test and the interpretation of test scores, test validation showed that the construct validity was weak, due to reliability *within* the constructs and *between* the constructs of EIL.   The quality of the test battery must be improved to support test use and interpretation of scores. While supporting validity arguments (content validity, face validity) are made, they are not strong enough to make the test battery a robust measure of the constructs as defined.

The correlations between the EIL sub-constructs (receptive intelligibility, productive intelligibility, strategic competence, and intercultural competence) are related to some degree, but each has some unique variance. Although internal consistencies were quite low, convergent and discriminant validity could be seen, providing moderate support for construct validity.

The authenticity of the topics covered by the test was arguably a good representation of real-world meaningful tasks from the target-language domain. The online format of the test battery weakened authenticity, but this was regarded as a reasonable trade-off for the practicality of capturing large amounts of data from test takers overseas. In the initial design phase, a role-play was considered as an authentic measure of live interaction and negotiation of meaning. However, once again, the demands of time for both the researcher and participants would not have been feasible under the circumstances.

The interactiveness of the test battery examined three characteristics of the test takers: language ability, topical knowledge, and affective schemata. Test takers' language abilities were engaged to a large extent, although efforts were made to limit the vocabulary load and make the test less about proficiency and more about EIL abilities. Topical knowledge was derived from an earlier survey of the test takers' target language use domain. This provided the background context for the test items, which should be familiar for many government officials. The level of affective schemata of the test takers is believed to be low for the following reasons: first, all of the volunteers were uncompensated for

their time on a task unrelated to their coursework, and therefore quite possibly less invested in their responses to test tasks. Second, the online format of the test battery did not likely engage their affective schemata in the same way as a role-play or a speaking task might.

The impact on the test takers was negligible at both the micro and macro levels. Completed outside of classroom time, the results from this test had no impact on their coursework at the university or any future endeavors. Likewise, as a research project, the test is unlikely to have any impact on society as a whole. However, it is hoped that this test will raise awareness of EIL and broader approaches of measuring communicative competence.

An analysis of practicality showed a minimal use of resources, one of the primary benefits of online testing. Test development was time-consuming, but that is acceptable since the testing cycle is ongoing. In terms of gathering data from overseas, the test battery was highly practical, despite some technical problems. The EIL test battery was sufficiently practical as an exploratory research project, but for large scale testing, it should be streamlined.

As an exploratory assessment tool, the EIL test battery produces mixed results, indicative of the early stages of test development. Further development of the test is likely to improve test quality, which requires creating more items which are both reliable and help to discriminate higher and lower ability test takers. Bachman and Palmer's (1996) six qualities of test usefulness are shown in Table 5.14, which summarizes the overall usefulness of the EIL test battery.

Table 5.15: Overall usefulness of EIL test battery

EIL Construct	Quality of Evidence	Reliability	With items deleted	Authenticity	Interactive -ness	Impact	Practicality
Receptive Intelligibility	Good	α = .781	-	Moderate	Moderate	Low	High
Productive Intelligibility	Good	α = .599	α = .746	Low	Moderate	Low	Low
Intercultural Competence	Poor	α = -.439	-.153	Low	Low	Low	High
Strategic Competence	Weak	α = .366	α =.483	Moderate	Low	Low	High

The evidence presented in Table 5.15 shows that the EIL test battery has moderate validity for making low-stakes decisions about test takers abilities in using EIL successfully in international contexts. This study shows that it is possible to test EIL abilities, however the development and validation of the assessment tools requires further testing cycles in order to improve internal consistency and strengthen the evidence supporting overall test usefulness. This would make the test battery more suitable for moderate stakes decisions about test takers' abilities to communicate successfully in English in international contexts.

While Chapter 5 has dealt primarily with the empirical validation of the test results, the following Discussion Chapter will discuss these results in more detail. The major findings from the test will be summarized as well as possible explanations for said results. Chapter 6 will move past the numbers and tables and take a more distant perspective of the EIL test battery as a whole.

Chapter 6: Discussion

6

The Discussion Chapter will first briefly summarize the research. Next, these findings are discussed and connected to the literature where relevant. Third, the major issues arising from the EIL test battery are looked at from the point of view of the researcher. The chapter will draw to a close by examining the potential implications of this research with respect to the English language classroom and English language testing.

6.1 Research summary

EIL is described in this thesis as a series of abilities and strategies that facilitate international communication in English. It is characterized by positive attitudes towards other English varieties and it entails a certain degree of intercultural knowledge about peoples and cultures different than one's own, and the ability to engage in the collaborative nature of international communication. This was specified in the thesis as four constructs of EIL ability: receptive intelligibility, productive intelligibility, intercultural competence and strategic competence.

These four constructs were conceptualized as follows: first, the productive intelligibility construct followed an adapted version of the Smith (1992) framework, and placed emphasis on phonemic distinction, fluency, ease of understanding, and the ability to use English indirectly. This was assessed by four raters from different linguistic backgrounds, which was one answer to the call for changes in testing (Canagarajah, 2006a). As a second measure of intelligibility, receptive intelligibility dealt with the ability to recognize different accents and varieties of fluent English. Next, intercultural competence was measured as a snapshot of knowledge about other cultures, and how best to resolve intercultural problems. Finally, the strategic competence construct served as a linguistic measure of the best way to clarify or repair a breakdown in communication.

These four constructs were operationalized into an online test battery aimed at mid-level government officials from developing Asian and African countries. Sixty-nine volunteers served as test takers in the final version of the

EIL test battery, generating test data that was subjected to Bachman and Palmer's (1996) six qualities of test usefulness. The Empirical Validation Chapter concluded that the EIL test battery, at its present stage of test development, has moderate validity for making low-stakes decisions about test taker's EIL abilities. Further test development is necessary before any substantial inferences can be drawn from the resulting scores of the test takers.

6.2 Threats to the validity of the EIL test battery

As an exploratory assessment tool, the EIL test battery is at a moderate level of test development. A number of problems exist at this stage of development, and these problems must be addressed. The following section will discuss the major threats to the validity of the EIL test battery, including the issues of construct under-representation and the moderate relationship between the EIL abilities. This will be followed by a re-examination of construct validity and a deeper look at the usefulness of the test.

6.2.1 Construct under-representation

The greatest threat to the validity of the EIL test battery is likely construct under-representation, or the insufficient number of items necessary to accurately cover and measure the constructs being assessed. While the receptive intelligibility section has the largest number of items (40) to ensure satisfactory coverage of that construct, the remaining three components of the test battery have too few items to ensure accurate measurement and coverage. The strategic competence and intercultural competence sections of the test battery have just ten items each, and the productive intelligibility section has only 16 items (which include a 4-5 minute voice sample). While there is no ideal number of items to use for testing a construct, a sufficient quantity of reasonable quality items must be found in order to find a balance between practicality and validity in the EIL test battery. This means that a test can perform satisfactorily using relatively few robust test items, which emerge after months of test development.

However, given that the test takers were volunteering their free time while studying on the exceptionally busy ELTO program, care was taken to prioritize practicality and minimize the time spent completing the test. As a

172

result, the measurement of EIL abilities was shallow and somewhat narrow. An assessment battery with better coverage would take considerably more time, and certainly result in fewer volunteers and therefore less data for empirical analysis. This is the trade off every test developer must make, striving to find an appropriate balance between practicality and validity. However, given the complexity of the four constructs outlined herein, far too much focus was placed on making a short, practical test battery rather than the depth required to give the test stronger validity claims.

6.2.2 The relationship between EIL constructs

While the EIL test battery is essentially a test in need of further development, some of the low correlations between the constructs were surprising. By re-examining Table 5.5 and considering the relationship between the constructs, it can be seen from an empirical vantage point that receptive intelligibility, productive intelligibility, intercultural competence, and strategic competence are not as closely related as previously thought. While they are different constructs, placing them together in a single test battery was an attempt to capture a much larger theoretical 'thing' called EIL abilities. The moderate to weak relationship is understandable when one considers that, for example, intercultural competence and receptive intelligibility are two separate abilities (resolving intercultural incidents and understanding different accents). While both skills are arguably related to international communication skills, they do not have a strong relationship. Overall, the low correlation between constructs demonstrates that the EIL test battery needs further development and is symbolic of the complexity of what this thesis is attempting to do: merge four constructs into one test battery. From a theoretical standpoint, this is possible. However, from a practical standpoint, such a merger presents too many validity issues and problems, especially if test scores are used to make real world decisions. While the four aspects of international communication in the EIL test battery are related, they are logically and empirically distinct. This mirrors the distinct abilities of listening, speaking, reading, and writing, which enjoyed satisfactory correlation ranging from .57 to .76, according to research (cited in Powers, 2010, p. 7) into 7,500 TOEIC language proficiency test takers.

173

In other words, the concrete linguistic skills of listening, speaking, reading, and writing share a greater relationship than the four skills identified as EIL constructs. It may be unreasonable to compare exploratory assessment with a thoroughly validated instrument like the TOEIC exam, yet this example does draw attention to the complexity of assembling different constructs into one category in assessment.

6.2.3   Re-examining construct validity

The reliability estimates for the EIL test battery varied greatly, from initial alphas of .781 for receptive intelligibility to -.439 for intercultural competence. While Messick (1989) notes that reliability is an important part of test homogeneity, it is not the only way to understand a test's internal structure. Reliability can be considered as the degree of precision of a test in general (rather than a measured construct) since "...a test can be reliable without its scores supporting certain inferences that a user may draw" (Roever, 2001). The reliability estimates were obtained from test takers with moderate English language skills (IELTS band 5.5) and although the test was targeted at this particular domain, the reliability estimates may have been higher if native speakers were included among the test takers. In other words, language ability may have been a bigger factor in test results than previously believed.

The construct's internal structure, assessed by considering the correlations between the test's sections indicate that the constructs of the test are tapping EIL abilities to a moderate degree, and that these abilities are not a unified set of abilities. While each of the constructs is of course distinct (as outlined in the test format) the relationship between the constructs is not strong or uniform enough to assume that each section taps into the exact same kind of knowledge. This is intuitive given the breadth and depth of EIL abilities, and the assumption that these abilities could be captured in a single test battery was ambitious at best. A future test of EIL abilities would require substantially more items (construct representation) and possibly additional EIL constructs, both of which requires more time and less test practicality.

6.3 Summary of major findings

This section will briefly summarize the major findings that have emerged. As an exploration into assessing EIL, care is taken to avoid making any grand claims, given the moderate results from the empirical validation of the test battery. However, I believe that the findings of this research show:

1. The construct of intercultural competence is not closely related to the more linguistic components of EIL abilities. The four identified constructs of EIL are not as closely related as previously believed.
2. The use of bilingual raters in assessing productive intelligibility produced satisfactory inter-rater reliability estimates.

These major findings are discussed in greater detail in the following sections. While this research does not present specific research questions to be answered, it does connect some of the findings with the research aims discussed in the introductory chapter as well as what other knowledge has been learned since the beginning of the thesis.

6.3.1 Intercultural competence in international communication

The analysis of the EIL test battery shows that intercultural competence, as defined in this thesis, shows the weakest relationship with the other components of international communication. With an initial Cronbach's alpha of -.439, clear problems exist with the conceptualization and operationalization of this section of the test. There are a few possibilities for this: first, the most likely reason is that each item is capturing a different construct within the notion of intercultural competence. For example, while knowledge about Japanese and Indian culture may fall under the general category of intercultural competence, specific cultural knowledge about each domain is in fact *different knowledge*. In other words, just because a test taker is aware of cultural norms in Japan, it does not mean they have similar knowledge about cultural norms in India – the knowledge may help, but it does not guarantee *higher abilities* in the general category of intercultural competence. As such, these two examples are different

175

sub-constructs, which adversely affects the internal consistency of the entire test section.

Another potential cause for the negative internal consistency of intercultural competence is that the test takers struggled and resorted to making guesses at this stage of the test. The heavy reading load that accompanies the description of each intercultural critical incident may have resulted in test taker fatigue, leading them to be less engaged in responding thoughtfully on each item. Likewise, if the correct response is completely unknown, the temptation to answer randomly may have been high. A post-test interview with test takers would have been useful for determining if this was indeed the case.

Moreover, with only ten items, the accurate measurement and content coverage of intercultural competence could not be ensured. This means the test is not eliciting every aspect of the ability in question. As argued in section 7.3.2, intercultural competence is a *multi-dimensional* construct, which should have been reflected in the test development. However, a more in-depth measure of the kinds of intercultural skills sought by this section of the test battery would have resulted in a much longer test, reducing the overall practicality of the EIL test battery.

Care must be taken, however, to avoid the introduction of text-heavy items that describe cultural incidents. Asking test takers to read long passages introduces construct-irrelevant variance in scoring, whereby factors like vocabulary, reading abilities, and other extraneous factors are introduced into the measurement of intercultural competence. Although care was taken to limit these factors (see for example, section 4.2.3 on vocabulary control), such problems can rarely be eliminated completely. A possible work-around would be to introduce the intercultural incidents in the form of short video clips (as described in section 7.3.1). The challenge for test developers is to find a good balance in the trade off between practicality and validity (especially authenticity); as you increase one, the other is inversely affected.

The central challenge facing the role of language and culture in EIL resembles the arguments made in section 2.1.7 concerning Standard English: which cultural norms can be applied to an international language shared by so many? The same challenge was evident early in test development when a test

section on pragmatics was considered for part of the EIL test battery: which pragmatic norms should be applied? The answer is a paradoxically no one's – and everyone's. As an international language, the English language is no longer owned by the British (or Americans) and thus it would be difficult establishing a firm connection between EIL and Western culture. That said, it is also difficult to deny the power of Western culture and its influence on the English language as well as the preference many language learners (and teachers) have for native speaker norms (Norton & Tang, 1997; Timmis, 2002). Inventor and entrepreneur Jay Walker proposes that "English is becoming the language of problem solving – not because America is *pushing* it, but because the world is *pulling* it." (Walker, 2013 emphasis added). With this 'pulling' of English comes a lot of the Western cultural norms that are embedded in the language, which is not "…the insidious hand of Western imperialism, but rather an expected outcome of the inevitable acceleration of globalization..." (Alsagoff, 2012, p. 5). As a result, the pragmatic and cultural norms for using English have complicated international communication, making accurate assessment increasingly difficult. Hu appropriately notes "The challenge for EIL test developers, then, lies in how to devise appropriate language tasks to elicit, in a rigorous manner, those pragmatic strategies that facilitate effective communication and allow valid inferences about the test takers' strategic competence to attain interaction success in EIL" (Hu, 2012, pp. 137–138). Hu does not elaborate on what such pragmatic strategies might entail, perhaps due to the complexity underlying intercultural competence.

Although intercultural competence is a deep and complex pool of knowledge, it can be used to enhance international communication by avoiding pitfalls and social faux pas that occasionally create tension or misunderstanding. Such knowledge can be taught explicitly in order to foster a 'think globally, act locally' identity in people. In other words, in intercultural communications, one should *think globally* in terms of EIL, and *act locally* with listeners to collaborate in the construction of meaning (Sharifian, 2009).

6.3.2 The complexity of EIL constructs

The theoretical concept of EIL is enormous, and cannot be captured by a 60 to 90 minute-long test battery as operationalized in this thesis. Each of the four constructs is much more theoretically complex than initially believed, requiring a proper test to probe these abilities more deeply. This would necessitate both a greater number of items and several more testing cycles in order to produce reliable and valid results that are meaningful in real world contexts. In other words, decision makers would be confident that the test results are reflective of real international communicative abilities. To add a high degree of authenticity to such measures, Canagarajah (2006a) suggests that realistic assessments involve multiple tasks, multiple raters, and multiple candidates. While this was attempted with moderate success in the EIL test battery, numerous threats to test validity remained.

For example, the multidimensionality of intercultural competence should have seen the construct divided into perhaps two or three sub-constructs (e.g. attitudes, cultural knowledge, problem-solving abilities) in order to better account for the depth intercultural competence brings to assessment. Although the demands of practicality limit an ideal assessment using qualitative and quantitative method approach such as interviews, and observations (Deardorff, 2006), several more items for each of these sub-constructs would boost the representativeness and validity claims for this section of the test. Moreover, the critical cultural incident approach used combined two constructs into one, accounting for the mixed results described in Chapter 5. These two combined constructs were the *recognition* – or knowledge – of a critical cultural incident and the behavioral *action* to remedy or minimize that particular incident.

Strategic competence is another example of a complex construct of EIL abilities. Strategic competence includes both verbal and non-verbal strategies, as well as interactional strategies. For EIL users, this interaction and negotiation of meaning is key to maintaining comprehension. Canagarajah (2006a, p. 238) suggests that negotiation skills are a component of strategic competence that "cannot be truly examined in the traditional format of individually answered written tests" and that a more interactive and collaborative format for assessment can be found in a situated performance. While such suggestions

might form the basis for an ideal test, it would not be an effective test in terms of administration, scoring, and time constraints.

In addition, the issue of pragmatics complicated the productive intelligibility (interpretability) section of the test. While this sub-construct aimed to measure the ability to use English indirectly, it does have pragmatic aspects, which were not intended to be part of the EIL test battery. A pragmatic component of EIL abilities was avoided because, like the argument against Standard English, the question of which norms to follow is raised and not easily answered. Thus when the raters were scoring interpretability according to the degree of appropriateness for the given context, they may have done so according to native speaker norms (despite their international backgrounds). The matter of 'appropriateness' or intercultural competence in EIL should not scored against any linguistic or cultural norm, but rather on the functional ability to complete a communicative task.

In terms of a model of international communicative competence, Nunn's (2007, 2011) proposal (see section 3.1.2) best addresses the complexity of international communication. The components of this model fit closely with the four constructs outlined in the EIL battery (with the exception of the linguistic component). However, in order to produce a meaningful operationalization and validation of Nunn's model, substantial in-depth testing would also be required, reducing the overall practicality of such an assessment tool. This raises the issue of test meaningfulness and usefulness, which leads Iyldyz to argue that:

> ...the attempt to validate any framework will be obstructed by problems of isolating and operationalizing the desired construct in test items and dangers of the data reflecting characteristics of a particular learner population. What is measured through taking into consideration generally accepted views on the nature of 'proficiency' should be made clear. Instead of asking whether the test is valid, we should ponder the utility of such a test, which brings us back to the faults of proficiency tests. (2011, p. 37).

Designed as a proficiency test, the EIL test battery was intended as a

179

practical and useful test of abilities above and beyond simple language skills. However, it was discovered that the complex abilities of successful EIL use defy accurate measurement using a short test battery of proficiency in EIL. However, the test may be used for low-stakes decisions or perhaps diagnostic purposes in the English language classroom. The implications section discusses what some of these test uses might be, as well as some of the potential implications of this research.

6.3.3 Reliability of bilingual raters

The decision to use bilingual raters in the speaking section of the EIL test battery (productive intelligibility) produced very satisfactory results. With one rater's results deleted, the inter-rater reliability estimate is .746, showing very good agreement between the different raters from very different regions. Possible reasons for the initial moderate level of agreement (.599) may have been due to the need for greater training and calibration, as well as the need for more discussion about the scoring rubrics (Appendix E).

I suspect that with greater resources and time for training, an even higher of agreement can be reached on what features constitute intelligibility in EIL. Agreement at this high level confirms research by You-Hee Kim (2009) which demonstrates that native and non-native raters are consistent, and that non-native raters are dependable in assessing oral performance. In addition, the use of bilingual raters in this assessment answers the calls for change from Canagarajah (2006a), Elder and Davies (2006), Hu (2012), and Mauranen (2012) to use raters from bilingual backgrounds as part of the assessment procedure of EIL, rather than monolingual native speakers.

6.4 Research implications

The implications stemming from this research fall into two large categories: implications for testing EIL and pedagogical implications. The challenges that face testing EIL are considerable, but this research has made some headway into overcoming these challenges. The role of bilingual EIL users in assessment is also discussed, as well as potential uses for this test battery. As

testing and pedagogy are closely linked, this section concludes with implications for teaching EIL and curriculum considerations.

6.4.1   Implications for testing EIL

While much has been written about the concept of EIL, little has been written to address the testing of EIL. This section discusses the challenges that face test developers, as well as the role of bilinguals as raters in a test of EIL. Potential test uses are also suggested for future versions of the EIL test battery.

6.4.1.1 Problems to overcome

Proposals for making accommodations for testing EIL have been suggested (see Jenkins, 2000; Davies et al., 2003; Canagarajah, 2006a; Elder & Davies, 2006; Hu, 2012) but to little or no avail.   The big three English language proficiency exams (IELTS, TOEIC, TOEFL) have made little headway in adopting an EIL-styled approach, beyond simply adding a small variety of native speaker accents to the listening section of their tests. Nor has headway been made with developing a new test of EIL abilities, incorporating the fixed language code of Standard English with the flexibility and awareness of using EIL effectively.

Reasons for this lack of progress stem from at least three difficulties concerning EIL: its conceptualization, pedagogical applications, and testing concerns. The conceptualization of a fluid dynamic 'thing' such as EIL is no simple task, and the research contained within barely manages to scratch the surface. EIL is not a "tangible product" (Wen, 2012, p. 85) that can be codified, packaged, and sold to language learners, who might later be tested on concrete abilities that enhance international communication. Instead, EIL is fluid, co-constructed, and negotiated by users in each specific context that arises in international communication.

One major implication for the Standard English of the big three proficiency exams is that limiting speaking norms to inner circle varieties poses a threat to claims that their test is valid. This threat is manifested by the notion that language proficiency exams should reflect how English is used in the world today (Canagarajah, 2006, 2007a) rather than reflecting how English is used by

the minority elite in a small handful of nations. While the letter 'I' for 'international' is contained within the acronym of the big three exams, they are international only in the sense that exam centers can be found around the world.

In reflecting on the diverse norms in which English is used in the world today, measuring EIL as an ability to interact with others in a broad range of international contexts and situations creates a construct so varied and flexible that it presents considerable validity issues which are not easily overcome. While some suggest more research is needed to codify EIL (see Elder & Davies, 2006; Lowenberg, 2012), the central problem that thwarts test developers is the flexibility with which EIL users need to operate. Moreover, EIL abilities encompass something much more than linguistic proficiency; it entails intercultural competence, familiarity with other varieties of English, and strategies for repairing/enhancing international communication.

To this end, it is hoped that the recognition of the abilities needed for effective international communication will mark a shift in language assessment from linguistic accuracy towards the negotiation and accommodation skills needed by speakers from different linguistic and cultural backgrounds. Assessment has room to evolve "from language as a *system* to language as a *social practice*, from grammar to pragmatics, from competence to performance."(Canagarajah, 2006a, p. 234 emphasis added) making *versatility* of English use a primary measure of proficiency.

6.4.1.2 The role of bilingual raters

The successful use of bilingual raters in the assessment of productive intelligibility shows that non-native English-speaking raters are reliable judges of spoken English, confirming earlier research by Kim (2009). This suggests that the typically monolingual native-speaking rater (qualified by birthright) is not the only type of acceptable rater of spoken English. In fact, the use of raters from different linguistic backgrounds is not only appropriate for EIL, but it recognizes the fact that English is used more in international contexts by bilingual speakers (Crystal, 2008; Graddol, 2003).

In order to address this fundamental shift in the use of English globally, the assessment (and teaching) of EIL is in need of an epistemic break from its

deference to inner circle norms. It is critical for examiners to be aware of stylistic and grammatical features in non-native varieties of English, and recognize features as *differences* rather than *deficiencies* (Lowenberg, 2012). Bilingual raters are better positioned to make such distinctions, given the breadth of their experiences as fluent EIL users. The raters used bring widely varied English experiences to bear on assessing the oral portion of the EIL test battery, and have done so with satisfactory reliability. Moreover, with this variety of experiences, they are able to arguably be fair in oral assessment, since research (see Carey, Mannell, & Dunn, 2011; Major, Fitzmaurice, Bunta, & Balasubramanian, 2002) shows that accent familiarity leads to higher evaluations in oral proficiency tests. This suggests that using raters or examiners from a variety of linguistic backgrounds may improve the validity of oral assessments, as well as strengthen the inferences that can be drawn from the results.

In addition to the use of bilingual raters, the EIL test battery used fluent voice actors from a wide variety of linguistic backgrounds, adding authenticity and representativeness to the listening section of the test. Section 2.2 contends that due to demographic changes, resulting in a situation where the number of non-native speakers of English outnumber the native speakers by a ratio of 4:1 (Crystal, 2012), meaning that authentic listening tests *should* include non-native speakers, something the big three proficiency tests have been reluctant to change. As a low-stakes assessment tool, the EIL test battery successfully used a variety of fluent bilingual speakers (in addition to native speakers) to produce a listening section with a satisfactory reliability estimate of .78, showing that test takers found the variety of accents manageable.

6.4.1.3 *Potential uses for a test of EIL*

While section 4.1.3 answered the question of whom the EIL test battery is for, this section will answer the question of who might use this test (or a subsequent version), and for what purposes. There are three potential test categories for the EIL test battery: proficiency, selection/recruitment, or diagnostic purposes. First, as a proficiency test, the EIL test battery aims to measure test taker abilities (as defined) and determine in a more general way

one's abilities to use EIL effectively in international contexts and situations. This is different from the high-stakes tests discussed in Section 2.3 in that the inner circle bias is replaced and EIL accommodations – prioritizing meaning over accuracy, accepting multiple varieties of English – are accepted.

Second, the EIL test battery may be used as a selection or recruitment tool for international public or private organizations. This includes public organizations such as the United Nations, the Association of Southeast Asian Nations (ASEAN), or the European Union. Large international private corporations, such as Toyota or Royal Dutch Shell may also utilize the EIL test battery to assist with overseas recruitment and hiring decisions. An increasing number of companies are already using personality tests as a means to vet candidates prior the interview to ensure they meet employer's expectations as well as match up well with the position advertised (Learnvest, 2015). Companies presently have access to over 90 existing intercultural assessment tools of various types (Fantini & Tirmizi, 2006) that can be used to aid decisions about recruitment. The EIL test battery might serve alongside existing measures in this regard, but requires several more testing development cycles to ensure a high degree of validity and reliability to justify its use in assisting with recruitment decisions.

The EIL test battery is more likely to serve as a diagnostic tool for identifying weak areas in candidate's EIL abilities. In this case, the individual section scores of the test battery might serve to highlight specific areas requiring improvement, rather than making judgments based on a total score of EIL abilities. For example, a monolingual English native speaker might score low in productive intelligibility, and may benefit from some awareness raising about speaking English clearly for an international audience. Just such an example is reported in a BCC story about native speakers, who often have trouble making themselves understood. UK-based communication skills and intercultural trainer Chia Suan Chong laments cases where "…you have a boardroom full of people from different countries communicating in English and all understanding each other and then suddenly the American or Brit walks into the room and nobody can understand them"(Morrison, 2016). Here Chong means that native speakers often talk too fast, make jokes, use slang and make references specific

to their own culture. Of course, international intelligibility is an issue for all EIL users, as are the abilities of intercultural competence and strategic competence. Diagnostic testing of these abilities may help identify areas requiring improvement to ensure a high degree of international communicative competence.

6.4.2 Pedagogical implications for teaching EIL

Implications relating to the teaching of EIL sometimes raise more questions than they answer. Which variety of English should teachers follow? How can culture be taught as part of EIL? Will too much variety overwhelm learners? While on one hand educators are increasing their awareness and interest in EIL, on the other hand they are not given an approach, pedagogy, or target model needed to begin implementing change. One thing that some scholars are certain of is that teaching EIL will require "an entirely different set of assumptions than the teaching and learning of any other second or foreign language" (McKay, 2002, p. 1). These assumptions are based on the demographic reality (discussed in section 2.2) that the learners are more likely to communicate with non-native speakers, and therefore significant changes are needed in the minds of teachers and learners as well as specific classroom practices (see Matsuda, 2006; McKay & Bokhorst-Heng, 2008; Sharifian, 2009).

Classroom practices are informed by curriculum. Drawing from EIL literature, Brown (2012, p. 156) compiled an extensive list of considerations for the development of an EIL curriculum, components of which can be used to best fit local contexts. This list includes:

1. Using successful bilinguals as English language and pedagogic models
2. Fostering language and cultural behaviors which help students communicate effectively
3. Promoting intelligibility in the classroom
4. Helping students to feel better about their English learning
5. Promoting awareness of linguistic and cultural differences, and teaching strategies for handling such differences effectively
6. Helping learners to have both a local and a international English voice

7. Including teaching materials and activities based on local and international situations

8. Including models of the different varieties of English worldwide

In extending the discussion beyond curriculum development, other pedagogical implications for the teaching of EIL include a shift away from the communicative language teaching approach, which Jenkins (2006a) sees as biased towards Western communicative styles and norms. EIL-based methodologies are argued to follow what Berman (1994, p. 7) calls a "global thinking, local teaching" approach which is more sensitive to local cultural contexts. Rather than continue with Western-influenced (i.e Anglo-Saxon) methodologies (Sullivan, 2000, p. 118), an EIL approach reinforces the notion that English is a global language used for a wide variety of cross-cultural and international functions.

The implications of this research are aimed primarily at raising awareness and expanding the knowledge base among language education professionals. We are not meeting all of the needs of English language learners by assuming the ultimate goal is enabling them to communicate effectively with native speakers only. Section 2.2 discussed the changing demographic situation, challenging the long-held assumption that all English-language learners are bound for English first-language countries and will interact primarily with native speakers of the language. In the postmodern world, this is no longer the case, and only the recognition of this fact will lead to positive changes in classroom pedagogy.

A second major implication of this research concerns the explicit teaching of EIL abilities to learners following diagnostic testing. The components of the EIL test battery may be useful in identifying which learners have which weak points in international communication. For example, a learner who scores low in the strategic competence section will require specific instruction and practice in order to improve that specific skill. Or a businessperson, prior to an overseas posting, may benefit from diagnostic testing which in turn may reveal weak knowledge of the major cultures in the world. Such diagnostic testing will

inform decisions that need to be made with regard to a) who needs what kind of teaching and b) how it should be taught.

The four constructs used in this thesis (receptive intelligibility, productive intelligibility, intercultural competence, strategic competence) are by no means the only EIL abilities requiring attention in the class. However, these four constructs align well with most of Brown's (2012) EIL curriculum points listed at the top of this section. Moreover, the four constructs are not based on a Western or inner-circle model of communicative or pragmatic norms, and therefore lend themselves to the EIL pedagogical approach discussed so far.

6.5   Chapter summary

Chapter 6 opened with a brief review of the research process from unpacking key terminology, conceptualizing, operationalizing, testing, and empirical validation processes utilized in this study. The major issues included the relationship between the constructs, which were originally argued to possess some unitary value as skills and abilities contributing towards successful communication in EIL. Results showed a moderate to weak relationship at this stage of test development. The results were affected by construct under-representation, which means an insufficient number of items were given to adequately measure each construct. This may also improve with further test development cycles, which were not possible under time constraints of the test takers. Despite the limited claims that can be drawn from the test results, an incredibly complex topic was tackled, albeit in a more holistic manner than standardized tests. It was felt that this holistic approach was necessary to capture the kinds of skills outlined as EIL abilities.

The major issues were identified as threats to the validity of the EIL test battery, the largest of which was construct under-representation. In its current stage of development, the number of items assigned to each construct is low. In addition, the moderate relationship between the constructs suggests that the notion of a unified set of EIL abilities is far from being realized. The constructs identified require further examination and scrutiny in order to strengthen the evidence in support of the use of this test.

The potential implications of this research addressed both testing and teaching. While several challenges need to be overcome, the testing of EIL herein suggested that bilingual raters and voice actors can enhance the validity of a test of international English. Moreover, the EIL test battery does have potential uses as a diagnostic assessment tool aimed at highlighting problem areas (e.g. intelligibility, intercultural competence) that require additional teaching or training.

Chapter 7 will conclude with the theoretical and methodological contributions made by research, as well as a discussion of its the limitations and future directions for further investigation.

Chapter 7: Conclusion 7

The conclusion chapter opens with the theoretical and methodological research contributions made by this thesis. This is followed by a discussion of the limitations of this thesis, such as the size of the test taker population, a number of technical issues, underperforming test items, and other factors that may have affected the outcome of this assessment tool. This section then leads into potential areas for further research and development of the EIL test battery before wrapping up this thesis with concluding remarks.

7.1 Research contributions

The EIL test battery is an assessment of skills argued to be essential for international communications. The test itself is exploratory research aimed at filling a gap in the literature, for although volumes have been written on EIL and English as a lingua franca, much remains to be written on the assessment of the dynamic, adaptive, and accommodative nature of international communication. It is this dynamic nature of EIL that makes it challenging to assess, since traditional language assessment focuses on the *fixed* language code. In its present stage, this research contributes to the three areas discussed in the following section.

7.1.1 Theoretical contributions

As discussed in Chapter 2, the aim was to answer the call for changes in Standard English proficiency exams by making accommodations that reflect modern-day English language use by bi- and multilingual individuals. The case was argued (in section 2.3.2) that existing measures of English language proficiency no longer meet the needs of the increasing numbers of people who use the language in international contexts, which are often free of native speakers (Seidlhofer, 2005). This thesis aimed to fill the gap by moving beyond what a test of EIL might look like (Canagarajah, 2006a; Elder & Davies, 2006; Hu, 2012) towards operationalizing the theoretical principles underlying successful international communication. This thesis has taken the initial steps by developing a test designed with specific EIL users – and their contexts – in mind, as well as producing results that were subjected to empirical validation.

Moreover, this thesis has built upon existing theoretical frameworks to develop the constructs argued to be essential for international communication in spoken English: receptive intelligibility, productive intelligibility, intercultural competence and strategic competence. While these four constructs are not a proposed new model for communicative competence, it does show the potential for a new operationalized approach that aims to capture the abilities that facilitate international communication in a test battery. This advance from a theoretical concept to an operationalized assessment tool is something that, to the knowledge of this researcher, represents a gap in the relevant literature. While it is acknowledged that a number of threats to the validity of the EIL test battery require addressing, and no large claims are made stemming from the test results, the operationalization of this EIL test is nonetheless a step forward, both in terms of building on theory, as well as test development.

7.1.2 Methodological contributions

The evolution of language testing and the need for change with regard to testing EIL has been discussed thoroughly in the literature (see Canagarajah, 2006; Elder & Davies, 2006; Hu, 2012; Jenkins, 2006c). However, given the complexity, fluidity, and uncertainty surrounding EIL, it is unsurprising that test developers have yet to engage with EIL, since EIL lacks a concrete authoritative description. The approach used to overcome this challenge was to focus on the *functions* of EIL and identify the skills and abilities that facilitate its effective use.

These skills and abilities are identifiable and measurable, and will help inform classroom instruction by raising the awareness of EIL for both learners and teachers alike. Suggestions for evolving classroom methodologies stem from the testing and pedagogical implications discussed in section 6.4, and build on existing work showcasing EIL programs and courses around the world (see Matsuda, 2012). The underlying theme presented herein is to encourage educators to shift away from inner-circle native speaker norms and standards of language use and add to this the instruction of non-linguistic skills such as intercultural competence.

A further contribution towards the testing of EIL involves the use of raters from different backgrounds to assess productive intelligibility. This multi-lateral approach answers the call made in section 2.3.2.2 for the involvement of raters from different linguistic backgrounds, a choice which adds face validity and construct validity to this research. Moreover, the high level of agreement of the different raters demonstrates that some EIL users are looking for similar things – such as ease of understanding and some degree of fluency – with regard to intelligibility.

7.2 Limitations of this research

The EIL test battery has only scratched the surface of the assessment of international communicative competence abilities, and obstacles must be overcome for more substantial claims to be made about the test takers and their abilities. The main limitations of this thesis included a relatively small test population, technical problems with the testing platform Qualtrics, and underperforming test items that need further development. A further limitation is that the abilities described in this thesis are by no means the *only* abilities that facilitate international communication. Additional research is needed to investigate what these abilities might be, and how they can be assessed.

7.2.1 Size of test taker population

While the test takers come from thirteen different countries, over half of the test takers were from just three (Cambodia, Laos, Myanmar). This is not sufficiently representative of the wider EIL community, which should include more Africans, Europeans, and test takers from English-speaking countries as well). Although there are nearly 200 countries in the world, gathering data from all or even most of them would be a mammoth task, even with the aid of the Internet. The possibility of opening up this test battery to the World Wide Web was discussed with the supervisor of this research, but decided against since it would be difficult to control exactly who becomes a test taker and what kind of results might be generated. In addition, the final number of test takers had to be culled from a raw grand total of 122 down to a final 69 due to technical problems

(see section 7.2.2) downloading the EIL test battery audio clips. This small final sample size limits the inferences that we can say about the test takers' EIL abilities and the extrapolation of scores into the real world.

7.2.2 Technical issues

The EIL test battery is a computer-based test, and as such, was subject to benefits and pitfalls. Computer-based tests for research purposes can take advantage of freely available commercial survey websites like *Qualtrics* or *SurveyGizmo* (see Itomitsu, 2009). Likewise, a test developer with programming skills can write the code for programming a custom-built computer-based test using JavaScript and HTML (see Roever, 2005). In either case, there are likely to be various technical difficulties that do not happen with standard paper-and-pencil tests.

The Qualtrics platform used for the EIL test battery had several built-in features that both aided and impaired data collection from test takers. For example, a time limit feature set up in the initial wave of data collection (June – July 2015) was intended to prevent test takers from repeatedly listening to audio clips and getting an artificially-high score (and threaten test validity). The timer was set to give test takers a generous amount of time to complete the first section of intelligibility and upon expiration, move test takers on to the next section of the test battery.

However, with the majority of test takers completing the online test in developing African and Asian countries, reliable Internet speed proved to be an unanticipated problem. Despite the fact that the initial 32 audio clips of the intelligibility section were compressed into smaller data files, this data download still presented problems for the test takers. While the timer began with the initial response click, the audio files took much longer to load than was anticipated. Test takers reported in email conversations that the Internet is not very reliable in their (developing) countries, leading to very slow download speeds and intermittent connectivity. The result was that approximately half of the overall group of test takers had to be dropped from the data collection due to incomplete data samples (largely the result of incomplete response sets and inability to log in).

The researcher sent the initial test link to the individual e-mail addresses of the participants, which was quickly discovered to be an inefficient method. Not only was it time-consuming, but also if the participants tried to continue the test battery from a different computer on a different date.   Qualtrics did not 'recognize' the test takers as partially complete testers, and restarted the test. This problem was overcome after two weeks by using the Qualtrics *Mailer*, a built-in E-mail list manager. This allowed the creation and management of test taker class E-mail lists, which could be sent en mass at a predetermined time. The Qualtrics Mailer is capable of tracking test takers by automatically following up on incomplete tests with periodic reminders. Moreover, it ensures that test takers who took a break from the lengthy test were able to continue the test from the last completed section of the test battery.

With good test practice in mind, the test security features of Qualtrics were enabled with the initial release of the test battery. This feature ensured that the link to the test could be sent from the Qualtrics Mailer only in order to prevent test takers from completing the test more than once. However, as a theoretical test of EIL abilities with no connection to their coursework whatsoever, this kind of cheating would be extraordinarily improbable. Following the first week of the EIL test battery release, this security feature was disabled.

Five text entry items from the word/utterance recognition sub-section of the test battery introduced construct-irrelevant variance problems, weakening the overall validity of the test battery. While test takers listened to a brief conversation, they simultaneously read and responded to short direct questions eliciting word/utterance recognition. Just over 50% of the test takers had problems with typos and excess information, making scoring unnecessarily complicated and thus prone to human error. For example, if test takers entered an unnecessary space anywhere in the correct response, the automatic scoring system of Qualtrics would mark it incorrect. The programmed scoring system would need to account for any possible variations in the field entry that were still correct, but different from the exact lettering and spacing of the correct response. These variations had to be manually scored, which defeats the advantage of computer-based testing and introduces human error in scoring.

For these reasons, a future higher-stakes version of the EIL test battery should take full advantage of custom designed online test by writing the computer code manually, as did Roever (2013) using JavaScript. This functionality allows a wider range of data to be collected, such as response times. This option also circumvents problems caused by different web browsers on different types of computers, such as technical problems viewing video/audio clips, and availability of headphones and plug-ins.

7.2.3 Underperforming test items

A second challenge with multiple-choice test on speech acts concerns the distractors; it can be quite difficult to write truly incorrect distractors that are not acceptable to some of the test takers. These problems also plagued Yamashita (1996) and Hudson et al. (1995) in their multiple-choice tests involving speech acts and cross cultural pragmatics. The psycholinguistic variables that make certain distractors more or less attractive is one area for further research, though for the near future, the practicality offered by multiple-choice testing will ensure its popularity.

7.2.4 Other research limitations

The conspicuous absence of native speakers of English from the test data is one avenue not explored. Throughout Chapters 2 and 3 the argument is made that any comparisons between native and non-native speakers is unfair, unnecessary, and contrary to the very spirit of EIL. However, from a research perspective, I am confident that scholars would like to learn about such comparisons, and see to what degree English language proficiency alone accounts for successful international communication. It is possible that a threshold exists, whereby being a native speaker is no longer an advantage in the EIL test battery, and abilities above and beyond proficiency start to emerge. In other words, a threshold point where bi- and multilingual EIL users begin to show the communicative abilities which monolingual native speakers sometimes lack. This includes familiarity with different accents, being understood by a wide variety of English users, and intercultural competence. While the EIL use described in this thesis most certainly includes first language users of English, a

comparison of native and non-native English speakers was not a research objective of this thesis.

Other EIL abilities beyond the four described herein are certainly believed to exist, and to be helpful for facilitating international communication. Pragmatic competence, for example, was explored as a potential construct of EIL. Elements of it emerge in the productive intelligibility section, where test takers are asked to use English indirectly in an awkward social situation. However, the main theoretical stumbling block I encountered with pragmatic competence was similar to the issue of intelligibility: pragmatic competence according to *whom*? As an international language bound by no specific culture, which pragmatic norms can be applied to discourse? Although this question could not be easily answered, I remain interested in some element of pragmatic (or sociolinguistic) competence as a part of a future EIL test battery.

7.3   Future directions for research

This research tackled a demanding and unwieldy topic knowing that the final result would be far from a finished product or a conclusive body of academic work. The abilities and complexities surrounding EIL extend well beyond the four constructs identified and attempted to measure. The topic of assessing EIL is as fascinating as it is complex, "an expected outcome of the inevitable acceleration of globalization that brings with it opportunities (as well as challenges) for change." (Alsagoff, 2012, p. 5). This section addresses the question of *How would you do things differently in the future?* forcing myself to admit that such research would take a considerably different approach, which is outlined in the following sections in terms of the four constructs of EIL.

7.3.1   Fine tuning a measure of intercultural competence

As mentioned in Chapter 3, the measure of intercultural competence in this thesis is a decidedly limited and holistic measure of intercultural awareness and how best to resolve incidents. A more accurate measure of intercultural competence is best measured and developed over time (Deardorff, 2011; Fantini, 2009; Lussier et al., 2007), multidimensional in its approach (Byram, 1997;

195

Fantini, 2009; Sercu, 2004) and best demonstrated in a real-time language performance interaction (Fantini & Tirmizi, 2006; Sercu, 2004).

Future versions of the EIL test battery, a more authentic and valid assessment of intercultural competence should have greater resources for involving a multi-cultural team to create and develop several items for different dimensions of intercultural competence. The construct should separate intercultural knowledge and intercultural problem-solving abilities as two dimensions rather than one. Rather than portraying intercultural incidents as reading passages, a video clip or role-play format might be more successful. Moreover, these skills and abilities should ideally be assessed over an extended period of several months, rather than just a one-time measure.

These steps are certain to strengthen the test quality for this section of the test battery, which presently cedes authenticity and validity for practicality. Video clips replacing text would increase validity by reducing the construct irrelevant variance of lengthy reading passages describing intercultural incidents. Reducing the reading load for test takers will allow time for more items, boosting construct coverage. Following each video of an intercultural incident, test takers would have to identify what went wrong in the critical incident and then provide a solution for solving or minimizing the intercultural problem.

By adding more items in this type of video format, it is almost certain that the internal consistency, coverage and representation, and validity would be noticeably improved. However, a testing platform other than Qualtrics would be necessary for video clips, given the problems reported in section 7.2.2 with compressed audio clips. Moreover, more time and resources would be needed to script, act out, and record these video clips, decreasing practicality. Test developers must carefully weigh the benefits and costs of such trade offs in a never-ending balancing act throughout the testing cycle. In the end, priorities will emerge based on the goals of and constraints on the test developer.

7.3.2   Revisiting productive intelligibility

The present use of the Smith framework is unnecessarily complex in its present adapted form for use in productive intelligibility. Future versions of the

EIL test battery should focus on spoken comprehensibility alone rather than producing similar-sounding words or measuring indirect language use. That is to say, word/utterance distinction and interpretability are less likely to be skills that facilitate international communication in daily EIL discourse. By cutting the high and low end of the Smith framework, two scoring rubrics from the assessment equation, allowing raters to focus on the intelligibility skills that EIL users need to be understood: fluency and ease of understanding. The use of fluent EIL raters from different linguistic backgrounds generated good results in terms of agreement, and this mode of assessment should continue to be investigated further.

7.3.3 Revisiting receptive intelligibility

Again, by placing focus comprehensibility, a future EIL test battery would better meet the needs of EIL users, who are not likely to encounter the kinds of indirect language use as described in the interpretability section whereby test takers are required to decipher indirect, implied meaning rather than clear explicit communication. The word/utterance section can be accounted for within comprehensibility since that construct subsumes word/utterance recognition.

With this focus, more items should be created along the lines of traditional English-language listening tests, with the distinction being that most of the speakers are fluent EIL users themselves. Topics can reflect the wide array of contexts in which English is used in international contexts as well as including natural EIL features such as clarification and negotiating meaning.

7.4 Concluding remarks

This thesis investigated the possibilities for testing English as an international language, a concept that has been discussed in the literature but never tackled. I suggested that the four key constructs of successful international communication are understanding different accents in English, being understood oneself, having a degree of intercultural competence, and when all else fails, the strategic competence to repair communication. The

results were subjected to empirical validation, which showed that these constructs are not as closely related as hoped, complicating the argument that these constructs belong to an overarching category called *EIL abilities*.

The findings suggest that further research is needed into each of these components, as each one is deeper in conceptualization than originally believed. The attempt to capture these four broad constructs in a convenient 1-hour long test battery was naïve on one hand, but arguably daring on the other. Nonetheless, an initial step forward in testing EIL has been taken, and it is hoped that more research in this exciting area continues through the work of other scholars.

I was encouraged to take on this endeavor by my supervisor, who warned me of the political challenges ahead, and the considerable gray zones into which I might wander and never emerge. He had me at the word 'challenges', and so I began a rather long journey that has truly broadened my approach towards the teaching and testing of English. I have had the fortune to meet and talk to several new and exciting people from all over the world, from places and cultures I knew little about. I have emerged at the end of this journey having invented something, an imperfect little tool to be sure, but an invention nonetheless.

References

Alptekin, C. (2002). Towards intercultural communicative competence in ELT. *ELT Journal*, *56*(1), 57–64.

Alsagoff, L. (2012). Another book on EIL? Heralding the need for new ways of thinking, doing, and being. In L. Alsagoff, S. McKay, G. Hu, & W. A. Renandya (Eds.), *Principles and practices for teaching English as an international language.* (pp. 1–6). London: Routledge.

Apedaile, S., & Schill, L. (2008). Critical incidents for intercultural communication: An interactive tool for developing awareness, knowledge, and skills.

Appadurai, A. (1996). *Modernity at large: Cultural dimensions of globalization*. Minneapolis: University of Minnesota Press.

Atkinson, D. (1999). TESOL and culture. *TESOL Quarterly*, *33*(4), 625–654.

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

Bachman, L. F. (1996). *Language testing in practice: Designing and developing useful language tests* (Vol. 1). Oxford: Oxford University Press.

Bachman, L. F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly*, *2*(1), 1–34. https://doi.org/10.1207/s15434311laq0201_1

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.

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

Bamgbose, A. (1998). Torn between the norms: Innovations in world Englishes. *World Englishes*, *17*(1), 1–14. https://doi.org/10.1111/1467-971X.00078

Bauer, L. (2002). *An introduction to international varieties of English*. Edinburgh: Edinburgh University Press.

Belz, J. A. (2007). The development of intercultural communicative competence in telecollaborative partnerships. In R. O'Dowd (Ed.), *Online intercultural exchange: An introduction for foreign language teachers* (pp. 127–166). Clevedon, UK: Multilingual Matters.

Beneke, J. (1991). Englisch als lingua franca oder als Medium interkultureller Kommunikation [English as a lingua franca or as a medium of intercultural communication]. *Grenzenloses Sprachenlernen. Berlin: Cornelsen*, 54–66.

Berman, R. A. (1994). Global thinking, local teaching: Departments, curricula, and culture. *ADFL Bulletin*, *26*(1), 7–11.

Berns, M. (2009). English as lingua franca and English in Europe. *World Englishes*, *28*(2), 192–199.

Bolton, K. (2002). Chinese Englishes: from Canton jargon to global English. *World Englishes*, *21*(2), 181–199.

Bredella, L. (2000). The significance of intercultural understanding in the foreign language classroom. In T. Harden & A. Witte (Eds.), *The notion of intercultural understanding in the context of German as a foreign language* (pp. 145–66). Bern: Peter Lang.

British Council. (2013). The English effect. Retrieved May 6, 2017, from

https://www.britishcouncil.org/sites/default/files/english-effect-report-

v2.pdf

Brodkey, D. (1972). Dictation as a measure of mutual intelligibility: A pilot study.

Language Learning, *22*(2), 203–217.

Brown, H. D., & Abeywickrama, P. (2010). *Language assessment: Principles and*

classroom practices (2nd ed.). New York: Pearson-Longman.

Brown, J. D. (2012). EIL curriculum development. In L. Alsagoff, S. L. McKay, G.

Hu, & W. A. Renandya (Eds.), *Principles and practices for teaching English*

as an international language (pp. 147–167). London: Routledge.

Bruthiaux, P. (2003). Squaring the circles: Issues in modeling English worldwide.

International Journal of Applied Linguistics, *13*(2), 159–178.

https://doi.org/10.1111/1473-4192.00042

Brutt-Griffler, J. (2002). *World English: A study of its development*. Clevedon:

Multilingual Matters.

Byram, M. (1995). Acquiring intercultural competence: A review of learning

theories. In L. Sercu (Ed.), *Intercultural competence* (Vol. 1, pp. 53–70).

Aalborg: Aalborg University Press.

Byram, M. (1997). *Teaching and assessing intercultural communicative*

competence. Clevedon: Multilingual Matters.

Byram, M., Gribkova, B., & Starkey, H. (2002). *Developing the intercultural*

dimension in language teaching: A practical introduction for teachers.

Strasbourg: Council of Europe.

Cain, R. (2015, May 28). Three helpful tips for dealing with vexing Chinese

negotiating tactics. Retrieved May 23, 2017, from

http://www.forbes.com/sites/robcain/2015/05/28/3-helpful-tips-for-

dealing-with-classic-chinese-negotiating-tactics/

Canagarajah, S. (1999). *Resisting linguistic imperialism in English teaching*.

Oxford: Oxford University Press.

Canagarajah, S. (2005). *Reclaiming the local in language policy and practice*.

Mahwah: Routledge.

Canagarajah, S. (2006a). Changing communicative needs, revised assessment

objectives: Testing English as an international language. *Language

Assessment Quarterly*, *3*(3), 229–242.

Canagarajah, S. (2006b). Negotiating the local in English as a lingua franca.

Annual Review of Applied Linguistics, *26*, 197–218.

Canagarajah, S. (2007a). Lingua franca English, multilingual communities, and

language acquisition. *The Modern Language Journal*, *91*(5), 921–937.

Canagarajah, S. (2007b). The ecology of global English. *International Multilingual

Research Journal*, *1*(2), 89–100.

Canale, M. (1983a). From communicative competence to communicative

language pedagogy. In J. C. Richards & R. W. Schmidt (Eds.), *Language and

communication* (pp. 2–27). London: Routledge.

Canale, M. (1983b). Language and communication. In J. C. Richards & R. W.

Schmidt (Eds.), *From communicative competence to communicative

language pedagogy* (pp. 2–27). London: Longman.

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches

to second language teaching and testing. *Applied Linguistics*, *1*(1), 1–47.

Carey, M. D., Mannell, R. H., & Dunn, P. K. (2011). Does a rater's familiarity with a candidate's pronunciation affect the rating in oral proficiency interviews? *Language Testing*, *28*(2), 201–219.

Catford, J. C. (1950). Intelligibility. *ELT Journal*, *5*(1), 7–15.

Celce-Murcia, M. (2007). Rethinking the role of communicative competence in language teaching. In E. A. Soler & M. P. S. Jordà (Eds.), *Intercultural language use and language learning* (pp. 41–57). Dordrecht, The Netherlands: Springer.

Celce-Murcia, M., Dörnyei, Z., & Thurrell, S. (1995). Communicative competence: A pedagogically motivated model with content specifications. *Issues in Applied Linguistics*, *6*(2), 5–35.

Chapelle, C. A., Enright, M. K., & Jamieson, J. (2010). Does an argument-based approach to validity make a difference? *Educational Measurement: Issues and Practice*, *29*(1), 3–13.

Chapelle, C. A., Enright, M. K., & Jamieson, J. M. (2004). Issues in developing a TOEFL validity argument. Presented at the 26th Annual Language Testing Research Colloquium, Temecula, CA.

Chapelle, C. A., Enright, M. K., & Jamieson, J. M. (Eds.). (2008). *Building a validity argument for the test of English as a foreign language*. New York: Routledge.

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, *7*(3), 309.

Cogo, A., & Jenkins, J. (2010). English as a lingua franca in Europe: A mismatch between policy and practice. *European Journal of Language Policy*, *2*(2), 271–293.

Common European framework of reference for languages: Learning, teaching, assessment. (2001). Cambridge: Cambridge University Press.

Cook, V. (1999). Going beyond the native speaker in language teaching. *TESOL Quarterly*, *33*(2), 185–209.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*(1), 98.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334.

Crystal, D. (1997). *English as a Global Language* (1st ed.). Cambridge: Cambridge University Press.

Crystal, D. (2003). *English as a global language* (2nd ed.). Cambridge: Cambridge University Press.

Crystal, D. (2008). Two thousand million? *English Today*, *24*(01), 3–6.

Crystal, D. (2010). *The Cambridge encyclopedia of language* (3rd ed.). Cambridge: Cambridge University Press.

Crystal, D. (2012). A global language. In P. Seargeant & J. Swann (Eds.), *English in the world: History, diversity, change.* (pp. 151–177). New York: Routledge.

Curley, M., & Thomas, N. (2012). *Advancing East Asian regionalism*. London: Routledge.

Davidson, F. (1994). Comments and replies. *World Englishes*, *13*(1), 119–120.

Davidson, F. (2006). World Englishes and test construction. *The Handbook of World Englishes*, 709–717.

Davidson, F., Hudson, T. D., & Lynch, B. K. (1985). Language testing: Operationalization in classroom measurement and L2 research. In M. Celce-Murcia (Ed.), *Beyond basics: Issues and research in TESOL* (pp. 137–152). Rowley, MA: Newbury House Publishers.

Davidson, F., & Lynch, B. K. (2002). *Testcraft: A teacher`s guide to writing and using language test specifications*. New Haven: Yale University Press.

Davies, A. (2003). *The native speaker: Myth and reality*. Clevedon: Multilingual Matters.

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

Davies, A., Hamp-Lyons, L., & Kemp, C. (2003). Whose norms? International proficiency tests in English. *World Englishes*, *22*(4), 571–584.

Deardorff, D. K. (2006). Identification and assessment of intercultural competence as a student outcome of internationalization. *Journal of Studies in International Education*, *10*(3), 241–266.

Deardorff, D. K. (2011). Assessing intercultural competence. *New Directions for Institutional Research*, (149), 65–79.

Dervin, F. (2010). Assessing intercultural competence in language learning and teaching: A critical review of current efforts. In F. Dervin & E. Suomela-Salmi (Eds.), *New approaches to assessment in higher education* (pp. 155–172). Bern: Peter Lang.

Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly*, *39*(3), 379–397.

Deterding, D., & Kirkpatrick, A. (2006). Emerging South-East Asian Englishes and intelligibility. *World Englishes*, *25*(3-4), 391–409.

Dewey, M., & Jenkins, J. (2010). English as a lingua franca in the global context: Interconnectedness, variation and change. In M. Saxena & T. Omoniyi (Eds.), *Contending with globalization in world Englishes* (pp. 72–92). Bristol: Multilingual Matters.

Ebadi, S., Zamani, G., & Sharifi, M. M. (2014). Interpretive argument of the academic reading module of IELTS: A fusion of current research? *International Journal of Language Learning and Applied Linguistics World*, *6*(4), 317–328.

Ebel, R. L. (1972). *Essentials of educational measurement.* (1st ed.). New Jersey: Prentice Hall.

Economist Intelligence Unit. (2012). *Competing across borders: How cultural and communication barriers affect business*. London: British Council. Retrieved from https://www.jku.at/zsp/content/e273302/e273317/Competing_across_borders_ger.pdf

Educational Testing Service. (2007). *International principles for fairness review of assessments*. Princeton, N.J.: Educational Testing Service.

Elder, C., & Davies, A. (2006). Assessing English as a lingua franca. *Annual Review of Applied Linguistics*, *26*, 282–301.

European Commission. (2001). *Eurobarometer: Public opinion in the European Union* (No. 55). Brussels: European Commission.

European Commission. (2012). *Special Eurobarometer 386: Europeans and their languages* (No. 386). Brussels: European Commission. Retrieved from http://ec.europa.eu/commfrontoffice/publicopinion/archives/ebs/ebs_386_en.pdf

Excerpts from Iraqi document on meeting with U.S. Envoy. (1990, September
23). Retrieved May 20, 2017, from
https://msuweb.montclair.edu/~furrg/glaspie.html

Fantini, A. E. (2009). Assessing intercultural competence. *The SAGE Handbook of
Intercultural Competence*, 456–476.

Fantini, A. E., & Tirmizi, A. (2006). Exploring and assessing intercultural
competence. Retrieved from http://www.experiment.org/

Farhady, H. (2012). Principles of language assessment. In C. Coombe, P.
Davidson, B. O'Sullivan, & S. Stoynoff (Eds.), *The Cambridge Guide to
Second Language Assessment* (pp. 37–46). Cambridge: Cambridge
University Press.

Fayer, J. M., & Krasinski, E. (1987). Native and nonnative judgments of
intelligibility and irritation. *Language Learning*, *37*(3), 313–326.

Fiedler, F. E., Mitchell, T., & Triandis, H. C. (1971). The culture assimilator: An
approach to cross-cultural training. *Journal of Applied Psychology*, *55*(2),
95.

Firth, A. (1996). The discursive accomplishment of normality: On 'lingua franca'
English and conversation analysis. *Journal of Pragmatics*, *26*(2), 237–259.
https://doi.org/10.1016/0378-2166(96)00014-8

Firth, A. (2009). The lingua franca factor. *Intercultural Pragmatics*, *6*(2), 147–
170.

Friedrich, P., & Matsuda, A. (2010). When five words are not enough: A
conceptual and terminological discussion of English as a lingua franca.
International Multilingual Research Journal, *4*(1), 20–30.

Fulcher, G. (2010). *Practical language testing*. London: Taylor & Francis Limited.

Fulcher, G., & Davidson, F. (2007). *Language testing and assessment*. New York: Routledge.

Giles, H., & Coupland, N. (1991). *Language: Contexts and consequences.* Milton Keynes: Open University Press.

Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. *Contexts of Accommodation: Developments in Applied Sociolinguistics*, *1*, 1–68.

Gnutzmann, C. (2000). Lingua franca. In M. Byram (Ed.), *Routledge Encyclopedia of Language Teaching and Learning* (pp. 356–359). London: Routledge.

Gnutzmann, C. (2005). 'Standard English' and 'World Standard English': Linguistic and pedagogical considerations. In C. Gnutzmann & F. Intemann (Eds.), *The Globalisation of English and the English Language Classroom* (pp. 107–118). Tübingen: Gunter Narr.

Görlach, M. (1990). The development of standard Englishes. In M. Görlach (Ed.), *Studies in the history of the English language* (pp. 9–64). Heidelberg: Carl Winter.

Graddol, D. (1997). *The future of English?* London: The British Council.

Graddol, D. (2003). The decline of the native speaker. *Translation Today: Trends and Perspectives*, 152–167.

Graddol, D. (2007). *English next: Why global English may mean the end of 'English as a Foreign Language'* (Vol. 62). London: British Council London. Retrieved from http://hi.sses.tc.edu.tw/blog/gallery/4/study03.pdf

Hall, G. (2010). International English language testing: A critical response. *ELT Journal*, *64*(3), 321–328.

Hamilton, J., Lopes, M., McNamara, T., & Sheridan, E. (1993). Rating scales and

native speaker performance on a communicatively oriented EAP test.

Language Testing, *10*(3), 337–353.

Hamp-Lyons, L., & Davies, A. (2008). The Englishes of English tests: Bias

revisited. *World Englishes*, *27*(1), 26–39.

Harris, R., Griffin, T., & Williams, P. (2002). *Sustainable tourism: A global

perspective*. Oxford: Routledge.

Hayes, A. F. (n.d.). SPSS, SAS, and Mplus Macros and Code. Retrieved January 26,

2016, from http://www.afhayes.com/spss-sas-and-mplus-macros-and-

code.html

Hayes-Harb, R., Smith, B. L., Bent, T., & Bradlow, A. R. (2008). The interlanguage

speech intelligibility benefit for native speakers of Mandarin: Production

and perception of English word-final voicing contrasts. *Journal of

Phonetics*, *36*(4), 664–679.

Heatley, A., Coxhead, A., & Nation, I. (2002). RANGE and FREQUENCY programs.

Retrieved November 21, 2016, from

http://www.victoria.ac.nz/lals/resources/range

Henning, G. (1987). *A guide to language testing: Development, evaluation,

research*. Cambridge, MA: Newbury House Publishers.

Holliday, A. (2005). *The struggle to teach English as an international language*.

Oxford: Oxford University Press.

Home of the IELTS English Language Test. (n.d.). Retrieved September 16, 2016,

from https://www.ielts.org/

House, J. (1999). Misunderstanding in intercultural communication: Interactions

in English as a lingua franca and the myth of mutual intelligibility. In

Teaching and learning English as a global language (pp. 73–89). Tübingen: Stauffenburg.

House, J. (2002). Communicating in English as a lingua franca. *EUROSLA Yearbook*, *2*(1), 243–261.

House, J. (2003a). English as a lingua franca: A threat to multilingualism? *Journal of Sociolinguistics*, *7*(4), 556–578.

House, J. (2003b). Misunderstanding in intercultural university encounters. In G. Kasper, S. Ross, & J. House (Eds.), *Misunderstanding in social life: Discourse approaches to problematic talk* (pp. 22–56). London: Longman.

House, J. (2012). Teaching oral skills in English as a lingua franca. In L. Alsagoff, S. L. McKay, G. Hu, & W. A. Renandya (Eds.), *Principles and practices for teaching English as an international language*. New York: Routledge.

Hu, G. (2012). Assessing English as an international language. In L. Alsagoff, S. L. McKay, G. Hu, & W. Renandya (Eds.), *Principles and practices for teaching English as an international language* (pp. 123–143). New York: Routledge.

Huang, R. (1984). How to achieve higher inter-rater strategy for anyone who wants to keep it simple. Presented at the American Educational Research Association Annual Meeting, New Orleans.

Hudson, T., Brown, J. D., & Detmer, E. (1995). *Developing prototypic measures of cross-cultural pragmatics* (Vol. 7). Honolulu: National Foreign Language Resource Center.

Huh, M. Y. (2006). *North Korea's negotiation behavior toward South Korea: Continuities and changes in the post inter-Korean summit era*. Korea Institute for National Unification.

Hurn, B. J. (2011). Simulation training methods to develop cultural awareness.

Industrial and Commercial Training, *43*(4), 199–205.

https://doi.org/10.1108/00197851111137816

Hymes, D. (1967). Two types of linguistic relativity (with examples from

Amerindian ethnography). In W. Bright (Ed.), *Sociolinguistics, Proceedings*

of the UCLA Sociolinguistics Conference, 1964 (pp. 114–167). The Hague:

Mouton.

Hymes, D. (1972). On communicative competence. In J. B. Pride & J. Holmes

(Eds.), *Sociolinguistics: Selected readings* (pp. 269–293). Harmondsworth:

Penguin.

ICEF Monitor. (2015, November 5). The state of international student mobility in

2015. Retrieved May 8, 2017, from

http://monitor.icef.com/2015/11/the-state-of-international-student-

mobility-in-2015/

IELTS - Why accept IELTS scores? (n.d.). Retrieved September 16, 2016, from

https://www.ielts.org/ielts-for-organisations/why-accept-ielts-scores

IELTS Media Centre 2013. (n.d.). Retrieved November 10, 2015, from

http://www.ielts.org/media_centre/media_centre_2013.aspx

Interagency Language Roundtable. (2012, May 18). ILR skill level descriptions

for competence in intercultural communication. Retrieved May 24, 2017,

from http://www.govtilr.org/skills/Competence.htm

Itomitsu, M. (2009). *Developing a test of pragmatics of Japanese as a foreign*

language (Unpublished doctoral dissertation). The Ohio State University.

Iyldyz, L. (2011). Rethinking the validity of the L2 proficiency concept: Lessons

    for EIL. In R. Nunn & S. Sivasubramaniam (Eds.), *From defining EIL*

    competence to designing EIL learning (pp. 31–44).

James, C. (2013). *Errors in language learning and use: Exploring error analysis*.

    London: Routledge.

Jenkins, J. (2000). *The phonology of English as an international language*. Oxford:

    Oxford University Press.

Jenkins, J. (2002). A sociolinguistically based, empirically researched

    pronunciation syllabus for English as an international language. *Applied*

    Linguistics, *23*(1), 83–103.

Jenkins, J. (2003). *World Englishes: A resource book for students*. New York:

    Routledge.

Jenkins, J. (2004). The ABC of ELT…"ELF." *IATEFL Issues*, *182*(9).

Jenkins, J. (2005). Teaching pronunciation for English as a lingua franca: a

    sociopolitical perspective. *The Globalisation of English and the English*

    Language Classroom, 145–158.

Jenkins, J. (2006a). Current perspectives on teaching world Englishes and English

    as a lingua franca. *TESOL Quarterly*, *40*(1), 157–181.

Jenkins, J. (2006b). Points of view and blind spots: ELF and SLA. *International*

    Journal of Applied Linguistics, *16*(2), 137–162.

Jenkins, J. (2006c). The spread of EIL: A testing time for testers. *ELT Journal*,

    60(1), 42–50.

Jenkins, J. (2012). English as a lingua franca from the classroom to the classroom.

    ELT Journal, *66*(4), 486–494.

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

Johnson, R. C., & Riazi, A. M. (2016). Validation of a locally created and rated writing test used for placement in a higher education EFL program. *Assessing Writing*. https://doi.org/10.1016/j.asw.2016.09.002

Joseph, J. E. (2004). *Language and identity: National, ethnic, religious*. New York: Palgrave Macmillan.

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

Kachru, B. B. (1986). *The alchemy of English: The spread, functions, and models of non-native Englishes* (1st ed.). Oxford: Pergamon Institute of English.

Kachru, B. B. (1994). The speaking tree: A medium of plural canons. In J. E. Alatis (Ed.), *Educational linguistics, cross-cultural communication and global interdependance* (pp. 1–17). Washington, D.C.: Georgetown University Press.

Kachru, B. B., & Smith, L. (1985). Editorial. *World Englishes*, *4*, 209–212.

Kachru, Y., & Nelson, C. L. (2006). *World Englishes in Asian contexts* (Vol. 1). Hong Kong: Hong Kong University Press.

Kachru, Y., & Smith, L. E. (2008). *Cultures, contexts, and world Englishes*. New York: Routledge.

Kane, M., Crooks, T., & Cohen, A. (1999). Validating measures of performance. *Educational Measurement: Issues and Practice*, *18*(2), 5–17.

Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, *112*(3), 527–535.

Kane, M. T. (1994). Validating interpretive arguments for licensure and

   certification examinations. *Evaluation & the Health Professions*, *17*(2),

   133–159.

Kawai, K. (1950). Mokusatsu, Japan's response to the Potsdam declaration. *The*

   Pacific Historical Review, 409–414.

Kirkpatrick, A. (2010). *English as a lingua franca in ASEAN: A multilingual model*.

   Hong Kong: Hong Kong University Press.

Knapp, K. (2015). English as an international lingua franca and the teaching of

   intercultural communication. *Journal of English as a Lingua Franca*, *4*(1),

   173–189.

Kramsch, C. (1986). From language proficiency to interactional competence. *The*

   Modern Language Journal, *70*(4), 366–372.

Kramsch, C. (1993). *Context and culture in language teaching*. Oxford University

   Press.

Kramsch, C. (1995). The cultural component of language teaching. *Language,*

   Culture and Curriculum, *8*(2), 83–92.

Kramsch, C. (1997). Guest column: The privilege of the non-native speaker.

   Publications of the Modern Language Association of America, 359–369.

Krippendorff, K. (2004). Reliability in content analysis. *Human Communication*

   Research, *30*(3), 411–433.

Kumaravadivelu, B. (2012). Individual identity, cultural globalization, and

   teaching English as an international language. In L. Alsagoff, S. L. McKay,

   G. Hu, & W. A. Renandya (Eds.), *Principles and practices for teaching*

   English as an international language (pp. 9–27). New York: Routledge.

Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The sources of four commonly reported cutoff criteria what did they really say? *Organizational Research Methods*, *9*(2), 202–220.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174.

Lane, H. (1963). Foreign accent and speech distortion. *The Journal of the Acoustical Society of America*, *35*(4), 451–453.

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, *28*(4), 563–575.

Learnvest, C. C. (2015, December 6). Five personality tests hiring managers are using that could make or break your next job interview. Retrieved November 28, 2016, from http://www.msn.com/en-nz/money/careersandeducation/5-personality-tests-hiring-managers-are-using-that-could-make-or-break-your-next-job-interview/ar-BBl1TRB#page=1

Lingley, D. (2006). Apologies across cultures: An analysis of intercultural communication problems raised in the Ehime Maru incident. *Asian EFL Journal*, *8*(1), 97–122.

Lowenberg, P. (2012). Assessing proficiency in EIL. In A. Matsuda (Ed.), *Principles and practices of teaching English as an international language* (Vol. 25, pp. 84–102). Bristol: Multilingual Matters.

Lussier, D., Golubina, K., Ivanus, D., Skopinskaja, L. W. S., Chavdarova-Kostova, S., Wiesinger, S., & de la Maya Retamar, G. (2007). Guidelines for the assessment of intercultural communicative competence. In I. Lazar, M. Huber-Kriegler, D. Lussier, G. S. Matei, & C. Peck (Eds.), *Developing and*

assessing intercultural communicative competence: A guide for language teachers and teacher educators (pp. 23–39). Strasbourg and Graz: European center for modern languages and Council of Europe publishing.

Mackenzie, I. (2014). *English as a lingua franca: Theorizing and teaching English*. Abingdon: Routledge.

Major, R. C., Fitzmaurice, S. F., Bunta, F., & Balasubramanian, C. (2002). The effects of nonnative accents on listening comprehension: Implications for ESL assessment. *TESOL Quarterly*, *36*(2), 173–190.

Matsuda, A. (2006). Negotiating ELT assumptions in EIL classrooms. In J. Edge (Ed.), *(Re-) Locating TESOL in an age of empire* (pp. 158–170). Hampshire: Palgrave Macmillan.

Matsuda, A. (Ed.). (2012). *Principles and practices of teaching English as an international language*. Multilingual Matters.

Matsuo, C. (2012). A critique of Michael Byram's intercultural communicative competence model from the perspective of model type and conceptualization of culture. *University of Fukuoka, Faculty of Humanities*, *44*(2), 347–380.

Matsuura, H., Chiba, R., & Fujieda, M. (1999). Intelligibility and Comprehensibility of American and Irish Englishes in Japan. *World Englishes*, *18*(1), 49–62.

Mauranen, A. (2003). The corpus of English as lingua franca in academic settings. *TESOL Quarterly*, *37*(3), 513–527.

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

McArthur, T. (1987). The English languages? *English Today*, *11*(3), 9–13.

McArthur, T. (1998). *The English Languages*. Cambridge: Cambridge University Press.

McCrostie, J. (2009, August 11). TOEIC no turkey at 30: Once-struggling test flying high in Japan as corporate partners take on larger role. *The Japan Times*. Retrieved from http://www.japantimes.co.jp/

McKay, S. L. (2002). *Teaching English as an international language: Rethinking goals and approaches*. Oxford: Oxford University Press, USA.

McKay, S. L. (2005). Teaching the pragmatics of English as an international language. *Guidelines*, *27*(1), 3–9.

McKay, S. L. (2009). Pragmatics and EIL pedagogy. In F. Sharifian (Ed.), *English as an international language: Perspectives and pedagogical issues* (Vol. 11, pp. 227–241). Bristol: Multilingual Matters.

McKay, S. L. (2012). Principles of teaching English as an international language. In L. Alsagoff, S. L. McKay, G. Hu, & W. A. Renandya (Eds.), *Principles and practices for teaching English as an international language*. New York: Routledge.

McKay, S. L., & Bokhorst-Heng, W. D. (2008). *International English in its sociolinguistic contexts: Towards a socially sensitive EIL pedagogy*. New York: Routledge.

McNamara, T. (2000). *Language testing*. Oxford: Oxford University Press.

Meara, P. M., & Milton, J. (2003). *X_Lex: The Swansea vocabulary levels test v.2.00*. Swansea, UK: Lognostics.

Meierkord, C. (1996). *Englisch als Medium der interkulturellen Kommunikation: Untersuchungen zum non-native-/non-native-speaker-Diskurs*. Frankfurt am Main: Peter Lang.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (pp. 13–103). New York: Macmillan.

Messick, S. (1996). Validity and washback in language testing. *Language Testing*, *13*(3), 241–256.

Mesthrie, R., & Bhatt, R. M. (2008). *World Englishes: The study of new linguistic varieties*. Cambridge: Cambridge University Press.

Modiano, M. (1999a). International English in the global village. *English Today*, *15*(02), 22–28.

Modiano, M. (1999b). Standard English (es) and educational practices for the world's lingua franca. *English Today*, *15*(04), 3–13.

Morrison, L. (2016, October 31). Native English speakers are the world's worst communicators. Retrieved November 28, 2016, from http://www.bbc.com/capital/story/20161028-native-english-speakers-are-the-worlds-worst-communicators

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, *45*(1), 73–97.

Munro, M. J., Derwing, T. M., & Morton, S. L. (2006). The mutual intelligibility of L2 Speech. *Studies in Second Language Acquisition*, *28*(01), 111–131.

Nault, D. (2006). Using world literatures to promote intercultural competence in Asian EFL learners. *The Asian EFL Journal Quarterly*, *8*(2), 132.

Nelson, C. L. (2008). Intelligibility since 1969. *World Englishes*, *27*(3-4), 297–308.

Nelson, C. L. (2011). *Intelligibility in world Englishes: Theory and application* (1st ed.). New York: Routledge.

Nerriere, J.-P., & Hon, D. (2009). *Globish the world over: A book written in Globish*. Lexington: International Globish Institute.

Norton, B. (1997). Language, identity, and the ownership of English. *TESOL Quarterly*, *31*(3), 409–429.

Norton, B., & Tang, C. (1997). The identity of the non-native ESL teacher on the power and status of non-native ESL teachers. *TESOL Quarterly*, *31*(3), 577–580.

Nunn, R. (2007). Redefining communicative competence for international and local communities. *Asian EFL Journal*, *9*(4), 77–110.

Nunn, R. (2011). From defining to developing competence in EIL and intercultural communication. *Journal of English as an International Language*, *6*(1), 21.

Nunnally, J. C. (1978). *Psychometric theory (2nd ed.)*. New York: McGraw-Hill.

Phillipson, R. (1992). *Linguistic imperialism*. Oxford: Oxford University Press.

Phillipson, R. (2003). *English-only Europe: Challenging language policy*. London: Routledge.

Pinker, S. (2011, November 15). Direct vs. indirect use of language. Retrieved November 1, 2016, from http://wordyenglish.com/lit/language_and_human_nature.html

Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, *29*(2), 191–215. https://doi.org/10.1006/jpho.2001.0134

Power, C. (2005, March 6). Not the Queen's English. Retrieved October 12, 2016, from http://www.newsweek.com/not-queens-english-114797

Powers, D. E. (2010). The case for a comprehensive, four-skills assessment of

    English-language proficiency. *R & D Directions*, *14*. Retrieved from

    https://www.ets.org/Media/Research/pdf/TC-10-12.pdf

Prodromou, L. (2007). A reader responds to J. Jenkins's "Current perspectives on

    teaching world Englishes and English as a lingua franca." *TESOL Quarterly*,

    41(2), 409–413.

Quirk, R., Widdowson, H. G., & Cantù, Y. (Eds.). (1985). *English in the world:*

    Teaching and learning the language and literatures. Cambridge:

    Cambridge University Press for the British Council.

Rajagopalan, K. (2010). The soft ideological underbelly of the notion of

    intelligibility in discussions about 'World Englishes.' *Applied Linguistics*,

    31(3), 465–470.

Rampton, M. B. H. (1990). Displacing the 'native speaker': Expertise, affiliation,

    and inheritance. *ELT Journal*, *44*(2), 97–101.

Richards, J. C., & Schmidt, R. W. (2013). *Longman dictionary of language teaching*

    and applied linguistics (4th ed.). New York: Routledge.

Roever, C. (2001, August). *A web-based test of interlanguage pragmalinguistic*

    knowledge: Speech acts, routines, implicatures. University of Hawaii at

    Mānoa, Hawaii.

Roever, C. (2004). Difficulty and practicality in tests of interlanguage pragmatics.

    In D. Boxer & A. D. Cohen (Eds.), *Studying speaking to inform second*

    language learning (pp. 283–301). Clevedon: Multilingual Matters Limited.

Roever, C. (2005). *Testing ESL pragmatics*. Frankfurt: Peter Lang.

Roever, C. (2006). Validation of a web-based test of ESL pragmalinguistics.

    Language Testing, *23*(2), 229–256.

Roever, C. (2013). Technology and tests of L2 pragmatics. In N. Taguchi & J. Sykes (Eds.), *Technology in interlanguage pragmatics research and teaching* (Vol. 36, pp. 215–233). Amsterdam: John Benjamins.

Salo-Lee, L. (2006). Intercultural competence in research and practice: Challenges of globalization for intercultural leadership and team work. In N. Aalto & E. Reuter (Eds.), *Aspects of intercultural dialogue: Theory, research, applications* (pp. 79–94). Köln: Saxa.

Schmidt, K. M., & Embretson, S. E. (2013). Item response theory and measuring abilities. In I. B. Weiner, J. A. Schinka, & W. F. Velicer (Eds.), *Handbook of psychology: Research methods in psychology* (2nd ed., Vol. 2, pp. 451–473). Hoboken, New Jersey: Wiley.

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, *8*(4), 350–353.

Sciolino, E., & Gordon, M. R. (1990, September 23). Confrontation in the Gulf: U.S. gave Iraq little reason not to mount Kuwait assault. *The New York Times*. Retrieved from http://www.nytimes.com/1990/09/23/world/confrontation-in-the-gulf-us-gave-iraq-little-reason-not-to-mount-kuwait-assault.html

Seargeant, P. (2010). Naming and defining in world Englishes. *World Englishes*, *29*(1), 97–113.

Seidlhofer, B. (2001). Closing a conceptual gap: The case for a description of English as a lingua franca. *International Journal of Applied Linguistics*, *11*(2), 133–158. https://doi.org/10.1111/1473-4192.00011

Seidlhofer, B. (2003). *A concept of "international english" and related issues: From "real english" to "realistic english"?* Strasbourg: Council of Europe.

Seidlhofer, B. (2004). Research perspectives on teaching English as a lingua franca. *Annual Review of Applied Linguistics*, *24*, 209–239.

Seidlhofer, B. (2005). English as a lingua franca. *ELT Journal*, *59*(4), 339–341. https://doi.org/10.1093/elt/cci064

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

Sercu, L. (2002). Autonomous learning and the acquisition of intercultural communicative competence: Some implications for course development. *Language Culture and Curriculum*, *15*(1), 61–74.

Sercu, L. (2004). Assessing intercultural competence: A framework for systematic test development in foreign language education and beyond. *Intercultural Education*, *15*(1), 73–89. https://doi.org/10.1080/1467598042000190004

Sharifian, F. (2003). On cultural conceptualisations. *Journal of Cognition and Culture*, *3*(3), 187–207.

Sharifian, F. (2008). Distributed, emergent cultural cognition, conceptualisation and language. In R. M. Frank, J. Zlatev, R. Dirven, & T. Ziemke (Eds.), *Body, language, and mind: Sociocultural situatedness* (Vol. 2, pp. 241–268). Berlin/New York: Mouton de Gruyter.

Sharifian, F. (2009). Cultural conceptualizations in English as an international language. In F. Sharifian (Ed.), *English as an international language: Perspectives and pedagogical issues* (Vol. 11, pp. 242–253). Bristol: Multilingual Matters Limited.

Sharifian, F. (2013). Globalisation and developing metacultural competence in learning English as an international language. *Multilingual Education*, *3*(1), 1–11.

Smith, L. E. (1976). English as an international auxiliary language. *RELC Journal*, *7*(2), 38–42. https://doi.org/10.1177/003368827600700205

Smith, L. E. (1983). Some distinctive features of EIIL vs. ESOL in English language education. In *Readings in English as an international language* (pp. 13–20). Oxford: Pergamon Press.

Smith, L. E. (1992). Spread of English and issues of intelligibility. In B. B. Kachru (Ed.), *The other tongue: English across cultures* (2nd ed., pp. 75–90). Chicago: University of Illinois Press.

Smith, L. E., & Bisazza, J. A. (1982). The comprehensibility of three varieties of English for college students in seven countries. *Language Learning*, *32*(2), 259–269.

Smith, L. E., & Nelson, C. (1985). International intelligibility of English: Directions and resources. *World Englishes*, *4*(3), 333–342.

Smith, L. E., & Rafiqzad, K. (1979). English for cross-cultural communication: The question of intelligibility. *TESOL Quarterly*, *13*(3), 371–380.

Snyder, S. (1999). *Negotiating on the edge: North Korean negotiating behavior*. Washington, D.C.: US Institute of Peace Press.

SoundCloud. (n.d.). SoundCloud. Retrieved October 9, 2015, from https://soundcloud.com/

Sridhar, S. N., & Sridhar, K. K. (1994). Indigenized Englishes as second languages: Toward a functional theory of second language acquisition in multilingual contexts. In R. K. Agnihotri & A. L. Khanna (Eds.), *Second language*

acquisition: Socio-cultural and linguistic aspects of English in India (pp. 41–63). London: Sage.

Strevens, P. (1981). What is "Standard English"? *RELC Journal*, *12*(2), 1–9.

Sue, D. W., & Sue, D. (1990). *Counseling the culturally different: Theory and practice* (2nd ed.). New York: Wiley.

Sullivan, P. (2000). Playfulness as mediation in communicative language teaching in a Vietnamese classroom. In J. P. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 115–131). Oxford: Oxford University Press.

Tagliabue, J. (2006, December 8). English is the EU's lingua franca. Retrieved May 16, 2017, from

http://www.taipeitimes.com/News/editorials/archives/2006/12/08/2003339612

Taylor, L. (2002). Assessing learner's English: But whose/which English (es)?'. *Research Notes*, *10*, 18–20.

Taylor, L. (2004). IELTS, Cambridge ESOL examinations and the Common European Framework. *Research Notes*, *18*, 2–3.

Thambipillai, P. (1992). Negotiating styles. In K. S. Sandhu (Ed.), *The ASEAN Reader* (pp. 72–75). Singapore: Institute of East Asian Studies.

Timmis, I. (2002). Native-speaker norms and international English: A classroom view. *ELT Journal*, *56*(3), 240–249.

TOEFL iBT: About the Test. (n.d.). Retrieved February 21, 2016, from

https://www.ets.org/toefl/ibt/about

Trudgill, P. (1999). Standard English: What it isn't. *Standard English: The Widening Debate*, 117–128.

United Nations, Department of Economic and Social Affairs, Population Division. (2016). *International migration report 2015: Highlights (ST/ESA/SER.A/375)* (pp. 1–32). New York: United Nations. Retrieved from http://www.un.org/en/development/desa/population/migration/public ations/migrationreport/docs/MigrationReport2015_Highlights.pdf

Walker, J. (2013, March 1). The world's English mania. Retrieved October 10, 2016, from https://www.youtube.com/watch?v=ZpILR21GWao

Walt, C. V. der. (2000). The international comprehensibility of varieties of South African English. *World Englishes*, *19*(2), 139–153. https://doi.org/10.1111/1467-971X.00165

Wang, H., Choi, I., Schmidgall, J., & Bachman, L. F. (2012). Review of Pearson test of English academic: Building an assessment use argument. *Language Testing*, *29*(4), 603–619.

Weir, C. J. (2005). *Language testing and validation*. London: Palgrave Macmillan.

Wen, Q. (2012). Teaching English as an international language in mainland China. In A. Kirkpatrick & R. Sussex (Eds.), *English as an international language in Asia: Implications for language education* (Vol. 1, pp. 79–93). Dordrecht: Springer Science & Business Media.

Widdowson, H. G. (1994). The ownership of English. *TESOL Quarterly*, *28*(2), 377–389.

Widdowson, H. G. (2012). ELF and the inconvenience of established concepts. *Journal of English as a Lingua Franca*, *1*(1), 5–26.

Wright, S. (2007). The right to speak one's own language: Reflections on theory and practice. *Language Policy*, *6*(2), 203–224.

Yamashita, S. O. (1996). *Six measures of JSL pragmatics* (Vol. 3). Honolulu:

    University of Hawaii Press.

Yong, O. K. (2009). At close quarters with the drafting of the ASEAN charter. In T.

    Koh, R. Manolo, & W. Woon (Eds.), *The making of the ASEAN charter* (pp.

    107–115). Singapore: World Scientific Publishing.

You-Hee Kim. (2009). An investigation into native and non-native teachers'

    judgments of oral English performance: A mixed methods approach.

    Language Testing, *26*(2), 187–217.

    https://doi.org/10.1177/0265532208101010

Young, R., & He, A. W. (1998). *Talking and testing: Discourse approaches to the*

    assessment of oral proficiency (Vol. 14). Amsterdam: John Benjamins

    Publishing.

Appendix A: Text of EIL test battery

Using English as an international language

My name is Douglas Meyer, a PhD student at Victoria University of Wellington. This test of the skills I think are needed for effective international communication in English is part of my thesis.

About this test - This test is voluntary and confidential. It will take a little over an hour to complete. If you need to take a break and return to the test later, you can. Your test answers are saved automatically after each section. Please use Google Chrome or Firefox web browser for best results.

There are four parts, one of which is not online.

PART I: Listening. Headphones are a good idea to help listen to the different voices from around the world.

PART II: Intercultural communicative competence. Read the situations and decide the best way to respond.

PART III: Strategic competence. When you don't understand someone, what do you say?

PART IV: Speaking. Have a voice recorder ready to record your voice. Most computers and Smart Phones have one.

-Remember this test is not connected to your ELTO course and grades.

If you wish to receive personal feedback, please complete all sections of the test and provide your full name and E-mail address. In exchange for your time you will receive: a test score at the end (excluding the speaking section) 30 minutes of my time to proof-read any of your English writing (via E-mail) a score and feedback on your pronunciation (if you send in the speaking section).

Ethics and Privacy

The Human Ethics Committee of Victoria University of Wellington has given permission for this research. By completing this online test, you are agreeing to be part of this research. Privacy is important, and all your data will be kept confidential in a locked computer file. You do not need to give your name, but if you'd like feedback on your results, please provide your E-mail address below. Any questions you have can be E-mailed to <douglas.meyer@vuw.ac.nz>

Q1.4 About You: (name optional). If you want personal feedback and a complete score from this test (including the speaking section), please provide the details below. You can use a nickname if you like. In any case, all data will be kept private.

Name

Citizenship

Email address

Q1.5 Gender

male

female

Q1.6 My age group is:

20-29

30-39

40-49

50 +

Q1.7 Which languages can you speak well (in your opinion)? Rank them below. Number 1 is the language you speak most fluently, number 2 is your second most fluent language, and so on. If you are very fluent in two or three languages, enter this in number one.

1. Most fluent language

2. Second language

3. Third language

4. Fourth language (or more)

Q1.8 If you have lived in another country (for more than six months) please indicate where.

Country 1

Country 2

Country 3 or more

Q1.9 Is English your first language or mother tongue? In other words, are you a native speaker of English? If so, click on 'yes' and scroll down to the bottom of the page and click on the red >> (NEXT) button on the bottom right.

Yes

No

Q1.10 If you have taken an English language test previously, please enter that information below. This information will be kept private in a locked computer file. Also, this information will help compare this test to other major tests.

IELTS overall band:

TOEIC total score:

TOEFL iBT total score:

Other (please enter test name and total score)

Q1.11 If you have lived in an English-speaking country for a short time, for about how many months did you do so? Move the sliding bar below to the correct number. Number of months:

Q1.12 Optional: If you have lived in an English-speaking country for a longer time, please indicate so here.

1-3 years

3-6 years

More than 6 years

All of my life.

PART I: RECEPTIVE INTELLIGIBILITY

This is a listening test consisting of three smaller parts: Section A: Word recognition. Section B: General comprehension. Section C: Interpretability (i.e. indirect language use). You will hear a variety of voices from around the world for each test sub-section. It may take a moment for audio clips to load. Using Internet Explorer has been problematic, so please use another Web browser.

Section A: Word Recognition

This section of the test will measure your ability to hear single words out of context. You will hear similar-sounding words spoken in different accents. For each number, you will see three words under the audio clip. Click on the orange 'play' button and listen. You will hear only one of the words below. Click on the one word you hear. You have about 7 minutes to answer all 25 questions. Note that it may take a moment for several short audio clips to load. Internet Explorer has created problems in this section. Consider Google Chrome or Firefox.

Q2.3
path
bass
bath
pass

Q2.4
pin
bin
vin
fin

Q2.5

took

tug

dug

dog

Q2.6

race

rice

lace

lice

Q2.7

ship

sheep

cheap

chip

Q2.8

bash

bass

pass

path

Q2.9

fine

pine

vine

bine

Q2.10

dart

tart

dard

tard

Q2.11
alive
arrive
arise
alas

Q2.12
cute
cut
caught
coat

Q2.13
boss
posh
both
push

Q2.14
face
pace
base
vase

Q2.15
dust
tusk
dud
task

Q2.16

flute

fruit

root

rude

Q2.17

fat

fate

vat

wait

Q2.18

three

tree

thee

she

Q2.19

fax

vax

packs

backs

Q2.20

debt

dead

bed

bet

Q2.21

lather

rather

laser

razor

Q2.22

pitch

peach

pith

pit

Section A (continued): Mr. Kambi calls the Airport Shuttle Service

Mr. Kambi is traveling to Asia next week to give a speech on disaster preparation. Mr. Kambi has made a change to his flight plans, and is now talking on the telephone to confirm his new arrival time with Ms. Tran. First, scroll down and look at questions 21-25 below. Then play Mr. Kambi's conversation. As you listen, type in the short answers to each question below the audio clip. You have 2 minutes for this section. Spaces and capital letters are not necessary.

Q2.25 What is Mr. Kambi's customer number?

Q2.26 Originally, Mr. Kambi was planning to arrive on which day?

Q2.27 What is Mr. Kambi's new flight number?

Q2.28 What time is his flight arriving?

Q2.29 What colour limousine does Mr. Kambi not want?

Section B: Comprehension

This section will test your listening comprehension skills. You will hear a speech about disaster management by Mr. Kambi. Before listening, take one minute to read the questions below. Then listen to the speech and answer the multiple-choice questions about the speech. You will have a total of four minutes once you begin this section.

Q3.3   What is the alarming trend Mr. Kambi is talking about?

The lack of disaster management.

Unprecedented damage caused by tropical storms.

There are more and more disastrous storms.

Countries in the region holding a disaster management conference.

Q3.4   Why does Mr. Kambi think the government has not addressed the issue of tropical storms?

The government is still reviewing weather data on climate change.

It is cheaper for the government to do nothing.

The government does not have enough experience to prepare for storms.

It is too expensive to prepare for storms.

Q3.5   What does Mr. Kambi think is the largest damage caused by the recent tropical storm?

The storm resulted in $829 million in damage and 529 deaths.

The largest damage occurred the night before the storm.

The storm resulted in 829 deaths and $529 million in damage.

Public trust in the government.

Q3.6   In which area does the government most need to invest, according to Mr. Kambi?

Updating weather stations.

Providing gear and training search and rescue personnel.

Building search and rescue centers.

Providing real-time flood information for citizens.

Q3.7 5. How can the government best reduce suffering, according to Mr. Kambi?

By spending more money to prepare for disasters.

By giving priority to the public and private sectors.

By pressuring citizens to fix this problem.

By getting funding from the private sector.

Section C: Interpretability

This section will test how well you can understand indirect English meaning.

Consider the following example between two housemates at dinner time:

Bob: Do you like the soup I made?

Kim: Wow, these are nice soup bowls. Are they new?

Kim probably doesn't like the soup, so she is saying something nice about the bowls. By being indirect, she is trying not to hurt Bob's feelings.

Now listen to the 10 short conversations below between speaker 1 and speaker 2. Then choose the best answer which shows speaker 2's intended meaning. There is no time limit for this section.

Context: After Mr. Kambi's speech, several diplomats and members of the government and guests are chatting and drinking in the conference center. They are talking about Mr. Kambi and his speech. Scroll down and read the context and responses before you play the audio clip.

<audio clip>

Q4.3   Many people are chatting, drinking and moving around the conference center. Two diplomats are discussing Mr. Kambi's speech. Listen to the audio clip above. What is speaker 2's intended meaning?

    A. His department spent a lot of money on the conference center.

    B. His hearing is usually bad, but the speech volume was just right tonight.

    C. He did not like Mr. Kambi's speech.

    D. The sound system was beyond his expectations

<audio clip>

Q4.5   At a dinner table, two diplomats are talking about Mr. Kambi's speech. Listen to the audio clip above. What is speaker 2's intended meaning?

    A. He likes his chicken well-done.

    B. He doesn't like either the meal or the speech.

    C. He thinks it is important to prepare a better dinner for the conference.

D. He thinks Mr. Kambi is a tough man.

<audio clip>

Q4.7   At another table, two cabinet members are also discussing Mr. Kambi's speech. Listen to the audio clip above. What is speaker 2's intended meaning?

A. He thinks spending must be increased.

B. He is worried about the weather.

C. He doesn't want to talk about spending.

D. He wants to change the topic of conversation.

<audio clip>

Q4.9   At the conference center bar, two business people are discussing possible results of increased spending by the government. Listen to the audio clip above. What is speaker 2's intended meaning?

A. She is looking forward to a trip.

B. She agrees completely.

C. She thinks there might be a possibility to sell more rescue gear in the future.

D. She is worried about losing her job.

<audio clip>

Q4.11 Near the conference stage, a Korean and Chinese diplomat are drinking and talking about the translation of the speech this evening. Listen to the audio clip above.What is speaker 2's intended meaning?

A. She didn't understand her translator.

B. She feels that her language skills are getting worse.

C. She thinks the translation was excellent.

D. She wants a new translator.

<audio clip>

Q4.13 At the conference room bar, two officials are talking about the possible impact of Mr. Kambi's speech on their government. Listen to the audio clip above.   What is speaker 2's intended meaning?

A. Mr. Kambi will be getting a promotion soon, and needs to change jobs.

B. The government is very unhappy with Mr. Kambi.

C. The government will hire Mr. Kambi to supervise disaster management spending.

D. Mr. Kambi's suggestions will create new jobs.

<audio clip>

Q4.15 At the conference room bar, a news reporter is interviewing a local politician about the possible changes in government spending. What is speaker 2's intended meaning?

A. He feels it is necessary to increase funding.

B. He thinks citizens must give their input on this matter.

C. He is more concerned about the economy.

D. He doesn't know if funding will be increased or not.

<audio clip>

Q4.17 At the conference dinner, two scientists are discussing the increase of tropical storms in the region, and the chance for more funding to research climate change. Senior government officials are standing within listening range, but S1 doesn't see them. What is speaker 2's intended meaning?

A. He is certain the government will increase their funding.

B. He is more interested in her colleague's research.

C. He is changing the topic to avoid trouble.

D. He disagrees strongly with her colleague.

<audio clip>

Q4.19 Search and rescue personnel from the military are also at the conference. Two of them are discussing their lack of proper equipment to do their job. Listen to the audio clip above.   What is speaker 2's intended meaning?

A. He thinks new funding is not coming anytime soon.

B. He means that patience is necessary when dealing with government funding.

C. He is worried about pilot safety.

D. He is hopeful that more funding will come soon.

<audio clip>

Q4.21 The mayor of Palnos City is chatting with a senior cabinet member about Mr. Kambi's speech. The mayor wants a search and rescue center built in his city, not on the coast. Listen to the audio clip above. What is speaker 2's intended meaning?
   A. He understands the mayor is worried about losing the election.
   B. He will make sure a search and rescue center is built in Palnos City.
   C. He is concerned the mayor is too focused on creating jobs in Palnos City.
   D. He is doubting the mayor's loyalty to the government party.

This test is about half-complete. If you wish to take a break, simply close your browser. The test is saved automatically, and you can resume the test from where you left off at a later time. You can do this at the end of any section of the test, but please return later to complete all of the sections. If you wish to keep going, click on the red 'Next' button below. The last two sections take about 12 minutes each, and the speaking section is just 4 minutes.

PART II: INTERCULTURAL COMMUNICATIVE COMPETENCE
In this section of the test, you will read about intercultural situations. Read these situations and choose the most interculturally appropriate response.

Q6.2   You have just started a new job in an international workplace, and have co-workers from all over the world. One Friday evening, your co-workers invite you to join a large group of them going to a Brazilian restaurant bar near your office. Upon arrival, you see a middle-aged lady (the owner) greeting all the customers with a single kiss greeting on the cheek when customers arrive. The owner now turns to you with open arms.

What could you do to BEST show your intercultural competence?

A. Give the owner a quick hug and pat on the back.

B. Smile politely but keep your distance.

C. Extend your hand for a handshake when she comes near.

D. Learn to relax and accept it, even if it makes you a little uncomfortable.

Q6.3   You are sent to a country to manage a project for the first time. During your first week, you are surprised at how different the workers are from in your country. You see the workers here take long breaks in the afternoon, stand around chatting for long periods, and generally falling behind schedule every day. You have tried both yelling at them and pushing them to finish on time. Nothing seems to change.

What could you do to BEST show your intercultural competence?

A. Bring in workers from own country to train the local workers.

B. Learn to understand how things are done from the local point of view.

C. Slowly bring the workers around to your way of thinking with bonuses.

D. Hire a bilingual local manager to achieve the results you want.

Q6.4   One day, you receive a phone call from a mid-level Chinese government official. You used to be good friends with him, and worked together on a joint development project a few years ago. He was very supportive, and helped you to get through the local bureaucracy and paperwork. But you haven't heard from him for a long time. Now he is calling to ask you for a really big favour, which is not part of your current job. If you do help, it'll take a lot of time and effort on your part.

What could you do to BEST show your intercultural competence?

A. Give him the phone number of a secretary who might be helpful.

B. Help him any way possible to maintain the relationship.

C. Provide help only if you think you will be working in China again.

D. Decline his request if you think it will take too much of your time.

Q6.5   You are working at a conference, guiding guests from registration to their seats in the main hall. As one of the guests arrives wearing a turban (religious head wear), you remember the conference center rules, which say all guests must remove hats while indoors.

What could you do to BEST show your intercultural competence?

A. Do not allow the guest to enter with his turban on. Rules must be followed.
B. Allow the guest to be seated as usual. The rules do not apply in this case.
C. Ask the guest to remove his turban and leave it at the coat check-in desk.
D. Sit the guest at the back of the conference hall, so he won't be seen by many people.

Q6.6   You are a section manager at an international company, which has just won a big work contract. To celebrate, you are organizing a party at a restaurant bar after work. Ali and Ahmed are two new staff members. They want to join the party and get to know the other staff, but Ali and Ahmed are observing Ramadan (a period of fasting for Muslims) this month.

What could you do to BEST show your intercultural competence?

A. Tell them they are welcome to join the party after Ramadan.
B. Find an ethnic restaurant with a vegetarian menu.
C. Delay the beginning of the party until after sunset.
D. Warn Ali and Ahmed that co-workers will be drinking alcohol at the restaurant bar.

Q6.7   You are visiting Kenya as part of a team helping with a large aid project. Unfortunately, the Customs paperwork for machine parts has been delayed for three months, and you have no idea why. A colleague suggests you go to the Customs office, find the person in charge, and give him some 'tea'. You

discover that 'tea' means a small amount of 'gift money' or 'appreciation' that will motivate the official to take care of your Customs paperwork.

What could you do to BEST show your intercultural competence?

A. Go to the Customs Office and find the person in change. Quietly give him/her an envelope containing some money.
B. Be patient and try to understand that things take longer in a developing country. It's better not to make trouble for anyone.
C. Go to the Customs Office, find the official involved, and tell him you will consider how to appreciate his/her efforts.
D. Go to the Customs Office and find another Customs official who will complete the paperwork without 'tea'.

Q6.8   You are visiting a small Pacific island on vacation. One part of your trip involves a 3-hour ferry trip to another island. When you arrive at the ferry terminal, you see there is no line (or queue) for tickets. Instead, there is a noisy crowd of people around the ticket booth, making you a little uncomfortable. You check your watch and see that the ferry is leaving in 10 minutes.

What could you do to BEST show your intercultural competence?

A. Wait for the crowd to quiet down and then get a ticket for a later ferry.
B. Walk into the crowd and make sure you buy your ticket.
C. Find another way to get to the other island.
D. Push your way to the front of the line and get your ticket.

Q6.9   You are studying abroad in a western country. Over a year, you make new local friends. One of them invites you to a dinner party for ten people, which you happily accept. You arrive on time with a small gift, but you are a little surprised by your host. After greeting you at the door, your host says, "Make yourself at home" and then returns to chat with other guests. In your

country the host does everything possible to make the guest comfortable, but here, you feel alone in a crowded room.

What could you do to BEST show your intercultural competence?

A. Try to keep busy by helping in the kitchen with the food and drink.
B. Relax and enjoy meeting the other guests on your own.
C. Talk to your host more so that he/she might take better care of you.
D. Sit quietly and wait for others to start a conversation with you.

Q6.10 You work for a company that is trying to do business with a similar company in Indonesia. You and your boss fly to Indonesia to meet with the head of this company and discuss areas where both companies can work together. However, after the second day of meetings, your boss is clearly impatient with the slow pace of the meetings, and starts to push the Indonesian company to into making some decisions.

What could you do to BEST show your intercultural competence?

A. Ask your boss to be patient, explaining that it will take time to develop a good business relationship with the Indonesian company.
B. Agree, and pressure the Indonesian company that the chance to make money is now, not later.
C. Plan future meetings at a bar, where both sides can get to know each other over drinks.
D. Throughout any meetings, make sure to clarify things and ask for feedback often.

Q6.11 You work in an international workplace for a growing company in India. At a meeting, one of the local co-workers (Mrs. Singh) is asked if she can complete a project by August 1st. She says yes, she understands that it can be finished by then. After that, Mrs. Singh mentions that her daughter will be getting married in July to a man from a nice family.

243

What could you do to BEST show your intercultural competence?

    A. Ask her if she would rather give the project to a co-worker.

    B. Assume that everything is fine and the project will be completed on time.

    C. Set a new deadline so that it is possible for Mrs. Singh to finish on time.

    D. Remember to buy something nice (such as a coffee maker) as a wedding gift.

PART III: STRATEGIC COMPETENCE

Sometimes in conversation, you may not understand someone because they speak too quickly, have an unfamiliar accent, or simply do not speak clearly. Other times, they may use expressions or idioms which you do not know. This section is testing how well you know just what to say in order to improve your understanding in a conversation. You will see and hear ten short conversations that are not clear. Audio clips of the conversation are available. Please choose the best response which fits in the space in each conversation.

Q7.3 Two people are talking on the telephone about an international conference in three months.   <audio clip below>

Mr. C: So, we are confirmed for booking the Riverside Hotel on July 1st, 2nd, and 3rd, right?

Mr. W: Yes, that's right. The hotel will be undergoing renovation June 29th, so it will be noisy through to the 6th. The main conference hall will be getting a new carpet, on the 1st and 2nd, and half of the executive guest rooms are being re-modeled from the 3rd.

Mr. C: Uhmm, ...

Mr. W: Our conference schedule should not be affected.

What can Mr. C say to best clarify the conversation? Make sure your answer fits both the context and Mr. W's response.

A. What do you mean about the executive guest rooms?

B. But the hotel will host us for those three days, yes?

C. How noisy will it be?

D. Please explain 'renovation'.

Q7.5   Two conference organizers are talking on the phone. They are discussing possible agenda topics for the international meeting next month, but Mr. R is having trouble speaking clearly.   <audio clip below>

Mr. A: So the report on the agricultural assistance project will be ready next
        month?
Mr. R: It should be, but who knows? At this stage, one cannot say. There
        have been many delays due to the change of government recently.
Mr. A: . . . .
Mr. R: I said I hope it'll be ready next month.

What can Mr. A say to best clarify this conversation?   Make sure your answer fits both the context and Mr. R's response.

A. What does that mean?

B. What is it that you want to mean?

C. Pardon? Can you say that again?

D. Can we put that on the agenda?

Q7.7   A work site manager is showing a visiting project leader the construction site of a new school.   <audio clip below>

Mrs. T: As you can see, the workers are ahead of schedule.
Mr. C: Why are the workers using shovels instead of a backhoe?
Mrs. T: . . . .
Mr. C: It's a machine with an engine and a very large shovel.

What can Mrs. T say to best clarify this conversation?   Make sure your answer fits both the context and Mr. C's response.

    A. What do you mean by a backhoe?
    B. What I wanted to say is that we will finish early.
    C. Can you say that again?
    D. I don't understand you.

Q7.9   Mrs. R is meeting his friend, Ms. L at the airport. Unfortunately, Ms. L has been delayed, and has only just now exited Customs.   <audio clip below>

Ms. L: Hi, it's been a long time!
Mrs. R: Too long, my friend. How are you?
Ms. L: Good, good. Sorry I'm late, there was some trouble with the airport… uhm…finding a new…ahh. . . . .
Mrs. R: Do you mean the jet way, the passenger walkway?
Ms. L: Yes, that's it. They had to find a replacement, and it took a long time.

What can Ms. L say to best clarify this conversation?   Make sure your answer fits both the context and Mrs. R's response.

    A. What do you call that tunnel to the airplane?
    B. Yes, it was trouble.
    C. I mean to say the tunnel?
    D. I don't know what it is.

Q7.11 Mr. G is in a bank abroad, changing his money into the local currency.
Mr. G: Hi, I'd like to change my Euros into dollars, please. <audio clip below>

Bank officer: Sure, no problem. If you are not a customer of this bank, it'll be
        necessary to do some paperwork first. Can I see some I.D?
Mr. G: . . . .
Bank officer: Yes, that's fine.

What can Mr. G say to best clarify this conversation? Make sure your answer fits both the context and bank officer's response.

    A. What does I.D.mean?

    B. Did you say identification?

    C. Pardon me, I'm not sure I understand.

    D. Do you mean identification...like my passport?

Q7.13 Two friends are attending an international conference. After the final day of the conference, they are chatting in the local shopping mall.<audio clip below>

Mrs. M: Let's get something to eat at this restaurant. It looks nice.

Mr. L: Hmm...there are no pictures on their menus. Can you read any of it?

Mrs. M: Well, it's all Greek to me!

Mr. L: . . . .

Mrs. M: Yeah. Let's ask for help.

What can Mr. L say to best clarify this conversation? Make sure your answer fits both the context and Mrs. M's response.

    A. Do you prefer Greek food?

    B. So, what you are saying is you don't understand?

    C. Can you say that again?

    D. I'm not sure I understand.

Q7.15 Two people are preparing for a joint presentation at a regional conference on international trade.   <audio clip below>

Mr. C: So, actually we have 25 minutes to speak, and not 30, right?

Mr. K: Yes, and that includes time for questions from the audience.

Mr. C: Well, let's get the presentation outline finished. Time and tide wait
      for no man!

Mr. K: . . . .

Mr. C: Time and tide wait for no man means we had better get the job done
        soon.

What can Mr. K say to best clarify this conversation?   Make sure your answer fits
both the context and Mr. C's response.

    A. Tell me again, ok?
    B. Can you say that once more, please?
    C. You mean what?
    D. I'm not sure I know this.

Q7.17 In preparation for an official welcoming ceremony, two friends are
discussing the best way to organize the list of speakers.   <audio clip below>

Mr. Y: I think we should have the younger speakers begin the ceremony, then the
        senior officials should speak at the end. What do you think?
Mr. H: Hmmm…ok. Now, let's see. The projector screen is in the middle, so…
Mr. Y: Oh, uhmm, where should we place the. . . . . . ., they stand behind it and
        speak on stage?
Mr. H: Oh, you mean the podium?

What can Mr. Y say to best clarify this conversation?   Make sure your answer fits
both the context and Mr. H's response.

    A. What I'm saying is...
    B. What do you call that thing...
    C. I mean this thing...
    D. I don't know the word...

Q7.19 Mr. S and Mrs. E are friends from different countries who also work for
the African Union. They are talking on Skype about getting together for dinner
after an African Union conference.   <audio clip below>

Mr. S: So, why don't we make a reservation at that new restaurant in Parktown?

Mrs. E: I hear it's really expensive….

Mr. S: That's ok, it's been months since we last met. Come on, let's have fun!

Mrs. E: Ja, well no fine.   Mr. S: . . . .

Mrs. E: It's like "Sure, whatever you like."

What can Mr. S say to best clarify this conversation?   Make sure your answer fits both the context and Mrs. E's response.

    A. You can say it again?
    B. What do you mean by that?
    C. So what you are saying is that you don't want to go there?
    D. What's that again?

Q7.21 Two former colleagues haven't met for a long time. They are chatting on the phone and talking about old friends.   <audio clip below>

Mr. W: Did you hear about David?

Mr. H: No, what's he doing these days? Last I heard he was being transferred.

Mr. W: Yeah, well he's still spewin' about his OS posting to Hong Kong.

Mr. H: . . . .

Mr. W: You could say that.

What can Mr. H say to best clarify this conversation?   Make sure your answer fits both the context and Mr. W's response.

    A. And this is good?
    B. What do you mean by 'spewin'?
    C. So what are you saying?
    D. You mean he's not too happy about it?

Part IV: Productive Intelligibility

For this part of the test, a voice recorder is needed. You may use your smart phone, a computer app, or any digital voice recorder. There are three short sections, resulting in a single 3-4 minute audio recording of your voice. Save this recording and send the audio file (for scoring) to: douglas.meyer@vuw.ac.nz

READ ALL OF THE INSTRUCTIONS BELOW BEFORE BEGINNING YOUR RECORDING. If you do not want feedback or scoring on this part of the test, scroll down to the bottom and click on the red >> button to skip ahead to the end.

Q8.2 SECTION A
Read each number, followed by the next four words. Then pause for 3-4 seconds and continue. There is no need to practice, just speak normally.

Q8.3 "Hello. My name is . . . . from the country . . . . I'll begin now.

1. path	bass	bath	pass
2. face	pace	base	vase
3. took	tug	dug	dog
4. race	rice	lace	lice
5. ship	sheep	cheap	chip
6. bash	bass	pass	path
7. fine	pine	vine	bine
8. debt	dead	bed	bet
9. lather	rather	laser	razor
10. cute	cut	caught	coat

...That's the end of section A."

PAUSE your recording here and briefly think about what to say for Section B.

SECTION B

Choose ONE topic from the list below and speak about it for about 1 minute.

Q8.5 Topic 1: Talk about your country. Where is it? What is your hometown like? What do you like (or not like) about your country? How is the weather there this time of year? What is most interesting about your country?

Topic 2: Talk about your job. What is your role and title, and in which government department do you work? Is there anything you like (or don't like) about your job? What is a typical day like? What projects have you been working on recently?

Topic 3: Talk about a difficult meeting you had in the past. Why was it difficult? What was being discussed? What point made the meeting difficult? Who was at the meeting? What was the result?

Topic 4: Talk about a foreign guest or VIP you met as part of your job. Who did you meet? Was he/she an expert in something? What was your responsibility or duties for this guest? Were you a guide or interpreter?

Topic 5: Talk about the culture or customs in your country. How are they the same or different than other countries in the region? Do you have any special festivals or events? Do you think young people still enjoy/do these activities?

Say "THIS IS THE END OF SECTION B"
PAUSE the recording here and think about how to respond to the next five questions.

SECTION C

This section is measuring your ability to use English indirectly. This skill is useful in difficult situations, where most people try not to upset others. For each context, just say "For number 1, I would say ..." and "For number 2, I would say ..." etc. Try to limit yourself to only 2 to 4 sentences. You can prepare by writing your answers on paper first (if you like).

Q8.7 1. One day at work your boss suddenly asks you to attend an international conference far away during national holidays. Unfortunately, you have planned a big family trip to attend your sister's wedding. Since you are organizing her wedding, it is very important that you attend.

Objective: Indirectly say no to your boss, but do not say exactly why you cannot go to the conference.
Your boss says, "I need you to fly to New York for the conference on climate change" You say " . . . . "

Q8.8 2. You have just spent a week escorting an important visiting guest around your country, visiting factories and reviewing joint development projects. Your guest, a wealthy businessman was not very sensitive or aware of the culture in your country. He has made several rude comments and clearly does not know much about your country and the joint projects you have been working on.

Objective: The guest has just made another rude comment. What can you say indirectly to make him understand these comments are not welcome?
Your guest says, "How many lazy bureaucrats do we have to bribe to get our paperwork finished for my project?"
You say " . . . . "

Q8.9 3. Imagine you come from a rich family and have an excellent education from a famous university. Your father is a powerful man with connections to business people and high-ranking officials in the government.   However, you are also modest and hard working. You want to work your way to the top of your department on your own, rather than take advantage of your family's powerful connections.

Objective: Indirectly explain your past without telling the whole truth about all the opportunities you enjoyed coming from a wealthy family. A co-worker asks you, "So, how did you get started in this department?"

You say " . . . . "

Q8.10 4. You are manager of your department with 24 people working under you, some of whom are your personal friends. Last month, the head of the ministry conducted a work performance review of all the departments. You were informed that one of your staff (and a close friend) must be fired for poor work performance. You call him into your office to tell him the bad news.

Objective: As the department manager, it is part of your job to fire staff. What do you say to this staff member?   He walks in and says, "Hey, what's up? What did you want to talk about?"
You say " . . . . "

Q8.11 5. You are the head of a foreign aid project in a very poor area. Jobs and money are really needed in this region. One day, a rich businessman wants you to write a false report saying that his company passes all health, safety and environmental standards.   He hands you an envelope full of cash.

Objective: Reject the bribe in such a way that does not threaten to end of the aid project. The businessman says, "I think we can agree that this project will benefit everyone."
You say " . . . . "

Q8.12 This is the end of the speaking section of the test.   You should have a single voice recording of 3-4 minutes. Send the audio file to the researcher at <douglas.meyer@vuw.ac.nz> for scoring.

This is the end of the test of English as an international language.
The next page will display your score and all of the answers for the questions completed online. Remember, this score does not include your speaking sample. To get a complete score, write down your score on the next page. Send

in your speaking sample for scoring, and later add that score to your present score.

Thank you for your time.

Appendix B: Summary of survey of test takers

An online survey was created in order to better understand the target language domain and how the participants use of EIL at work. This survey elicited information about three areas: bio data, English usage at work, and methods of fixing communication problems in English. This survey is just one of the bottom up approaches used in this research (the other two bottom-up approaches are classroom observations and interviews

Participants

Sixty-eight current and former government officials of the ELTO program responded to the survey, with a gender ratio of 56% male to 44% female. The majority (62%) were in aged in their thirties. The second-largest age group were people in their 20's, at 24% of respondents.

Table B.1: Nationalities of respondents

Nationality	Number	Nationality	Number
Benin	2	Myanmar	1
Cambodia	16	Niger	1
Cameroon	2	Timor Leste	5
Lao	4	Togo	1
Mongolia	1	Vietnam	14
Morocco	5	Other	1

The majority of respondents (69%) were from either Cambodia, Lao, or Vietnam. Many respondents spoke three languages, and several speak four languages, a reflection on the nature of the multilingual region where most respondents live, and the need for language skills in their workplace.

Survey results

Work context: nearly 50% reported English as being used at work by either themselves or their colleagues. The top responses for ways in which English was used are, in order:

- Reading emails, letters or reports   69%
- Speaking with foreign guests   54%

- Translating documents   54%
- Speaking in meetings   43%
- Speaking with business people   40%
- Planning projects   39%

A follow-up item of other tasks requiring English use generated a list which included writing project proposals, hosting international conferences, listening to newscasts, accompanying foreigners on field trips, interpreting, negotiating agreements, technical training courses, official functions, and presentations.

English language problems in the workplace were categorized as speech acts, such as requests, apologizing, complaining, expressing opinions, and so forth. The speech acts respondents indicated were somewhat vexing are: understanding different accents (41%), expressing opinions (37%), and responding to/making requests (34%).

Strategies

By far, the most common strategy used by respondents to overcome problems in EIL communication was to get the general meaning from spoken discourse rather than understand every word (63%). Other popular strategies include clarification (50%), rephrasing (49%), and speaking slowly (47%). Avoiding a clarification in the hopes it will become apparent later was the least popular strategy, with just 39% willing to let it pass.

Intercultural communication

This survey did not gather useful data on cultural barriers to communication. The survey was lacking a sufficient number of items and sufficient depth to investigate to what degree EIL misunderstandings were perceived to be the result of cultural differences.

Conclusion

The survey was successful in three ways. First, sufficient numbers of ELTO participants responded to the survey, lending it important credibility and validity. Secondly, the survey tool did what it was designed to do: determine how the participants used EIL at work, define common problems in

communication they face, and strategies used to overcome difficulties. Finally, it provided a sufficient number of contexts, situations, and tasks that will feed into the test battery design.

Appendix C: Summary of classroom observations

Appendix C is a brief summary of classroom observations carried out between September and November 2013. The researcher was looking for cases of communication breakdown and the circumstances that led to both the origin of the problem and its resolution. The participants were approximately 60 bi- and multilingual government officials from developing Asian and African countries, ages ranging from 25 to 45 years old. Their language ability has been established at the level of IELTS band score 4.5 to 5.5 (intermediate).

Purpose of observations:

- to identify instances of communication breakdown use of conversation repair strategies (if any).
- to identify the source of the misunderstanding
- to draw samples from the target-language domain
- to help design authentic items, tasks, and contexts for the assessment battery

Table C.1: Observation schedule

Date	ELTO Class	Activity	Observation Summary
26 Sept.	Asia	Pair work	Passive listening
25 Oct.	Asia	Pair work	Passive listening
31 Oct.	Africa	Split reading	Back-and-forth negotiation for meaning
1 Nov.	Asia	Pair work	Back-and-forth negotiation for meaning
4 Nov.	Africa	Pair work	Laughing, gesturing, back channeling
7 Nov.	Africa	Split reading	Overlapping agreement/disagreement
8 Nov.	Asia	Interclass	More clarifications, giving examples
14 Nov.	Asia-Africa	Interclass	Noticed Asians had small gestures, Africans big gestures)
15 Nov.	Asia-Africa	Interclass	Good comprehension, back-and-forth

Data collection: pen and paper notes were made while quietly observing the students working on classroom activities and assignments in pairs or small groups. Recordings were not deemed necessary because it may have disrupted the class and affected natural behaviors. Any audio recordings would be overwhelmed by background noise. Moreover, the classroom activities were not under the control of the researcher, so any data collected would be incidental rather than elicited by any tool of research design.

Observation results
- Positive attitudes, mutual respect, and open minds facilitated conversations, despite a lack of familiarity with each other's cultures.
- Asian participants tended to avoid open disagreements during activities.
- Many language errors were allowed to pass (without correction) in small group work.
- Asian classes tended towards passive listening with little back channeling, requests for clarification, or repetition).

Specific observations
Linguistically, all participants used a number of strategies to overcome communication difficulties. This list includes repetition, back channeling, confirmation checks and clarification checks. The split reading task elicited a lot of conversation repair, including the following examples:
- *You understand?*
- *I explain you how...*
- *You don't know?*
- *They can create a...how do you call it?*
- *You don't know?*
- What is it you want to mean?

These examples encapsulate EIL efforts to repair conversation, as interlocutors actively negotiate meaning and make efforts to understand each other. This cooperative effort to keep conversation going is demonstrated in this

example, where a participant suggested lexical items as a means of mutual support.

> A: "We need to…" [frustrated gesture]
> B: "Keep peace?"
> A: "Yes, and also continue to watch…"

These general and specific classroom observations were made with the intent of capturing cases of communication breakdown and adapting them as authentic scenarios for test battery tasks.

Conclusions drawn from observations

- The learner-friendly classroom environment was geared towards successful outcomes, eliciting few cases of communication breakdown.
- In case of communication breakdown, meaning was quickly and successfully negotiated.
- Observations confirmed that EIL users prioritize meaning making, pragmatic strategies, and the priory of intelligibility over accuracy.

Appendix D: Profile of EIL raters

All inter-raters are PhD students at Victoria University of Wellington in New Zealand at the time of this thesis.

Rater A is from the Netherlands originally, where she learned English, French and German. She did her Masters in the U.K. and lived there 15 years. She has traveled to 17 countries and has lived in Netherlands, Switzerland, the UK, and New Zealand.

Rater B is from Kenya, where English was learned as a colonial language alongside two domestic languages and one regional language. He has taught English and Sociolinguistics for over 14 years at schools and universities in Nairobi, and has written a book on women's rights and lived in two countries.

Rater C is from China. She has studied at the University of York in the U.K. for one year MA program and has recently completed her PhD thesis on second language grammar instruction. She has taught English for about 13 years in China and now teaches Chinese and English as foreign languages in New Zealand.

Rater D is from Canada, and has over 15 years of English language teaching experience in South Korea, Japan, and New Zealand. He has lived and worked in four different countries and traveled to over ten countries. He speaks English and some Japanese, a little French, and can read Korean.

Appendix E: Text of productive intelligibility test

There are three sections to this speaking test:

1. Word/utterance recognition: pronouncing similar-sounding words.

2. Comprehension: providing a 1-2 minute speech sample on a familiar topic.

3. Interpretability: using English in an indirect manner.

The following are the online instructions informing test takers how to complete the speaking section of the EIL test battery.

SECTION A: Word/utterance recognition

Read each number, followed by the next three words. Then pause for 3-4 seconds and continue. There is no need to practice, just speak normally.

"Hello. My name is . . . . from the country . . . . I'll begin now.

1. path   bass   bath   pass

2. face   pace   base   vase

3. took   tug   dug   dog

4. race   rice   lace   lice

5. ship   sheep   cheap   chip

6. bash   bass   pass   path

7. fine   pine   vine   bine

8. debt   dead   bed   bet

9. lather   rather   laser   razor

10.cute   cut   caught   coat

...That's the end of section A

PAUSE your recording here and briefly think about what to say for Section B.

SECTION B: Comprehensibility

Choose <u>ONE</u> topic from the list below and speak about it for about 1 minute.

Topic 1: Talk about your country. Where is it? What is your hometown like? What do you like (or not like) about your country? How is the weather there this time of year? What is most interesting about your country?

Topic 2: Talk about your job. What is your role and title, and in which government department do you work? Is there anything you like (or don't like) about your job? What is a typical day like? What projects have you been working on recently?

Topic 3: Talk about a difficult meeting you had in the past. Why was it difficult? What was being discussed? What point made the meeting difficult? Who was at the meeting? What was the result?

Topic 4: Talk about a foreign guest or VIP you met as part of your job. Who did you meet? Was he/she an expert in something? What was your responsibility or duties for this guest? Were you a guide or interpreter?

Topic 5: Talk about the culture or customs in your country. How are they the same or different than other countries in the region? Do you have any special festivals or events? Do you think young people still enjoy/do these activities?

Say "THIS IS THE END OF SECTION B"

PAUSE the recording here and think about how to respond to the next five questions.

SECTION C: Interpretability

This section is measuring your ability to use English indirectly. This skill is useful in difficult situations, where most people try not to upset others. Read each context and then say the best response in only 1 to 3 sentences. You can prepare by writing your answers on paper first (if you like).

1. One day at work your boss suddenly asks you to attend an international conference far away during national holidays.
Unfortunately, you have planned a big family trip to attend your sister's wedding. Since you are organizing her wedding, it is very important that you attend.

Objective: Indirectly say no to your boss, but do not say exactly why you cannot go to the conference.

Your boss says, "I need you to fly to New York for the conference on climate change"

You say " . . . . "

2. You have just spent a week escorting an important visiting guest around your country, visiting factories and reviewing joint development projects. Your guest, a wealthy businessman was not very sensitive or aware of the culture in your country. He has made several rude comments and clearly does not know much about your country and the joint projects you have been working on.

Objective: The guest has just made another rude comment. What can you say indirectly to make him understand these comments are not welcome?

Your guest says, "How many lazy bureaucrats do we have to bribe to get our paperwork finished for my project?"

You say " . . . . "

3. Imagine you come from a rich family and have an excellent education from a famous university. Your father is a powerful man with connections to business people and high-ranking officials in the government.

However, you are also modest and hard working. You want to work your way to the top of your department on your own, rather than take advantage of your family's powerful connections.

Objective: Indirectly explain your past without telling the whole truth about all the opportunities you enjoyed coming from a wealthy family.

A co-worker asks you, "So, how did you get started in this department?"
You say " . . . . "

4. You are manager of your department with 24 people working under you, some of whom are your personal friends. Last month, the head of the ministry conducted a work performance review of all the departments. You were informed that one of your staff (and a close friend) must be fired for poor work performance. You call him into your office to tell him the bad news.

Objective: As the department manager, it is part of your job to fire staff. What do you say to this staff member?

He walks in and says, "Hey, what's up? What did you want to talk about?"

You say " . . . . "

5. You are the head of a foreign aid project in a very poor area. Jobs and money are really needed in this region. One day, a rich businessman wants you to write a false report saying that his company passes all health, safety and environmental standards. He hands you an envelope full of cash.

Objective: Reject the bribe in such a way that does not threaten to end of the aid project.

The businessman says, "I think we can agree that this project will benefit everyone."
You say " . . . . "

This is the end of the speaking section of the test. Send the audio file to the researcher at <douglas.meyer@vuw.ac.nz> for scoring.

Appendix F: Scoring rubrics for productive intelligibility

Four raters from Europe, Asia, Africa, and North America evaluated the productive intelligibility test by listening to the audio-recordings of voice samples on four separate occasions. For each of the three sub-sections of the test (valued at a maximum of 20 points each), the three scoring rubrics below were used. Brief instructions for the raters follow each rubric.

Table F.1: Scoring word/utterance recognition

Score	Description of ability
0	The spoken words are unintelligible or indistinguishable to the listener.
1	The spoken words are intelligible and recognizable to the listener.
2	The spoken words are intelligible, recognizable, distinct and clearly articulated for the listener.

There are 10 items for a maximum of 20 points for this section. Focus on identifying their weak pronunciation areas, such as [r] and [l], or perhaps [th] and [s] to provide feedback.

Table F.2: Scoring comprehension

Score	Description of ability
18~20	Speaking is completely smooth, fluent, and easily understood throughout. No errors in accuracy that impact comprehension.
15~17	Speaking is usually smooth and fluent. Pronunciation is easily understood despite occasional lapses. Very few minor errors in accuracy.
12~14	Speaking is sometimes smooth and fluent. Pronunciation usually understood with very minor strain. Minor errors in accuracy are noticed, but easily ignored.
9~11	Speaking is generally smooth, but fluency is lacking. Pronunciation occasionally causes the listener strain. Accuracy is increasingly affecting comprehension.
6~8	Speaking is rarely smooth or fluent, but generally understandable. Pronunciation often causes strain. Accuracy errors are common, but basic communication is understood.
3~5	Speaking is not smooth, sometimes unintelligible, and lacks fluency. Pronunciation is poor overall. Accuracy errors are frequent, negatively affecting comprehension.
0~2	Speaking is largely unintelligible, with significant problems affecting fluency and pronunciation. Accuracy seriously impairs comprehension.

Using the rubric, choose a grade which best matches the description of ability. Then quickly convert that grade into a raw point score (that will fit in with the scoring format of all other test sections). The maximum score for comprehension will be 20 points.

Table F.3: Scoring interpretability

Points	Description of language use
4	The use of indirect language is very appropriate for the given context. Response is limited to 1-3 sentences and achieves the objective completely.
3	The use of indirect language is appropriate for the given context. Response is limited to 1-3 sentences and achieves the objective to a moderate degree.
2	The use of indirect language is somewhat inappropriate for the given context. Response exceeds the 1-3 sentence limit but eventually achieves the objective.
1	The use of indirect language is not appropriate for the given context. Response is either far too brief or too long, and the objective is not achieved.
0	Non-responsive or unintelligible response.

For responses which exceed 3 sentences (but are otherwise fine) subtract 1 point. Maximum score: 20 = 5 items at maximum 4 points each.

BLANK   PAGE