Pronunciation Assessment in the Context of World Englishes

Dimova, Slobodanka

Published in:
Assessment in Second Language Pronunciation

Publication date:
2017

Document version
Peer reviewed version

Citation for published version (APA):
Pronunciation Assessment in the context of World Englishes

Slobodanka Dimova

University of Copenhagen
Slobodanka.dimova@hum.ku.dk

Word count: 8,137

Abstract: 267

The globalization of English has initiated a debate of acceptability of competing norms, be they endonormative or exonormative (Davies, 1999; Gill, 1999). In WE’s view, exonormative orientations towards the native-speaker may be discriminatory against non-native speakers (NNS) of English, so it recommends a pluralized and pluricentric notion of world Englishes (Kachru, 1992). ELF’s standpoint, conversely, rejects native speaker norms in favor of endonormative realizations of lingua franca varieties. These normative issues are exacerbated, both theoretically and practically, in the context of language assessment. The chapter will begin by outlining the early WE conceptualizations of pronunciation through the model of understanding in cross-cultural communication (Smith, 1976). These conceptualizations will then be contrasted with more recent ELF views of mutual intelligibility among the Expanding Circle users, for which “phonological intelligibility” is the most important criterion (Jenkins, 2006a). The chapter will continue with a discussion of criticism of the current practices in language testing and assessment that claim that the field has failed to adopt the WE perspective and realistically represent the variation of pronunciation norms in international communication. In addressing these criticisms, the chapter will argue that embracing WE or ELF, particularly in relation to pronunciation, is a challenging task due to the existing constraints guiding the design of valid tests which accurately represent the domains of target language use (Elder & Harding, 2008). The chapter concludes that despite these constraints, strides have been made towards encompassing a WE perspective in test construction and task design, especially in listening and speaking tests, though the strides are not expected to result in radical changes in the current language testing practices.
Introduction
This chapter addresses assessment-related pronunciation issues within the paradigm of World Englishes (WE), including English as Lingua Franca (ELF) which arguably has emerged from the WE framework. The rapid spread of English across the world, fueled first by colonialism and recently by internationalization and globalization, has resulted in emergence of different English varieties and diverging views over English norm selection for international uses.

According to WE, the focus on native-speaker (NS) norm in international communication in English is inappropriate because it undermines the multiplicity of varieties encountered in real-life communicative situations. While WE recommends pluralized and pluricentric notion of world English norms, ELF rejects native speaker norms in favor of endonormative realizations of the lingua franca varieties (Kachru, 1992).

The field of language testing and assessment has been accused of hesitating to adopt perspectives stemming from WE and ELF research, and therefore failing to realistically represent the variation of pronunciation norms in international communication in English (Canagarajah, 2006; Davidson, 2006; Jenkins, 2006b; Lowenberg, 2002). The main critique is language test developers’ over-reliance on native-speaker (NS) norm criteria and promotion of linguistic standards which exclude English varieties spoken in many contact situations (see Davies, Hamp-Lyons & Kemp, 2003). In today’s world, being multidialectal has become a prerequisite of English proficiency, which means that the desire to emulate only an ideal native speaker in learning and testing situations is unrealistic (Kachru, 1994; Sridhar, 1994). Considering the multitude of contexts in which standardized tests are used, the local validity of standardized test tasks calibrated against a target norm becomes questionable if the target norm does not coincide with the varieties spoken in the setting in which the scores are used (Lowenberg, 1993).

Pronunciation holds a prominent place in WE and ELF discussions about intelligibility of different English varieties. Research on factors affecting intelligibility (e.g., accent familiarity, attitudes) has major implications for language test design, particularly tests of speaking and listening. Despite the critiques, the
field of language assessment has been changing to reflect the sociolinguistic
conditions of international target-language domains. Moreover, research in
language testing has contributed to an improved understanding of how intelligibility
factors may affect the validity of testing procedures.

The chapter will begin by outlining early WE conceptualizations of
pronunciation through the model of understanding in cross-cultural communication
(Smith, 1976). These conceptualizations will then be contrasted with the more
recent ELF views of mutual intelligibility among the Expanding Circle users, for
which “phonological intelligibility” is the most important criterion (Jenkins, 2006a).
The chapter will continue with a discussion of the criticisms of the current practices
in language testing and assessment which claim that the field has failed to adopt the
WE perspective and realistically represent the variation of pronunciation norms in
international communication. In addressing these criticisms, the chapter will argue
that embracing WE or ELF, particularly in relation to pronunciation, is a challenging
task due to the existing constraints guiding the design of valid tests which
accurately represent the domains of target language use (Elder & Harding, 2008).
The chapter concludes that despite these constraints, strides have been made
towards encompassing a WE perspective in test construction and task design,
especially in listening and speaking tests, in which pronunciation is implied, though
the strides are not expected to result in radical changes in the current language
testing practices.

Historical and current conceptualizations of English for international
communication
The implications of norm selection have been acknowledged in standardized
language testing where fairness and avoidance of bias are critical factors. The
inclusion of pronunciation norms other than the “standard varieties” of English in
standardized tests could affect the level of bias against different groups of test-
takers and the washback effect on language teaching and learning. However,
developing assessment methods which incorporate pronunciation in non-standard
speech has proved extremely challenging for test developers, particularly due to lack of codification of emerging varities (Jenkins, 2006b; Taylor, 2006).

Considerations of bias ensue from the different conceptualizations of what constitutes International English (IE), i.e. to what degree IE includes or excludes various user groups. The World Englishes (WE) concentric circle model (Kachru, 1992) representing the spread of English can be applied to describe user group inclusion across the different paradigms. As Seidelhofer (2009) points out, traditionally IE was based exclusively on the Inner Circle varieties, i.e. first language (L1) English varieties, because it “is generally interpreted as the distribution of native-speaker Standard English rather than the way English has changed to meet international needs” (p. 237). In support of the post-colonial emancipation, on the other hand, much of the work in WE has focused on recognition and inclusion of Outer Circle varieties, i.e. post-colonial English varieties, while neglecting the Expanding Circle, i.e. English as a foreign language (EFL) varieties, which became the main focus of the English as a lingua franca (ELF) research.

WE and ELF scholars vehemently oppose the traditional approach to norm selection, i.e. a standard native English variety, criticizing its monocentricity; Standard English (SE) is centered solely on the educated native-speaker norm because of its prestige, recognizability, and spread. Despite the SE failure to embrace the developing international uses, it has been used as the most common model for learning and assessment because it represents a codified language system against which learners’ progress or proficiency can easily be measured (Lowenberg, 2002).

Arguably, the international uses of spoken SE could hypothetically lead to independent development of a monolithic form of spoken English which has been referred to as World Standard Spoken English (WSSE) (Gorlach, 1990; McArthur, 1987; 1998). Though independent, WSSE has apparently been strongly influenced by the U.S. variety of English (Crystal, 2003), which means that it remains a “single monochrome standard form” based on the native-speaker (NS) models used by non-native speakers (NNSs) (p.6; see also Quirk, 2014).

According to WE scholars, the SE approach is biased against the local norms whose role and status become undermined. An exonormative, pluricentric approach
to norm selection allows for the realistic representation of different varieties, including English varieties from the *Outer Circle*, i.e. post-colonial countries. This representation would help legitimize and strengthen the status of *Outer-Circle* varieties, and consequently lead to their codification. Descriptive analyses of language uses would improve our understanding of how language works in different contexts and provide a wider and more flexible interpretation of what forms are acceptable, unlike the rigid prescriptivism of standards (Nelson, 2011). While WE scholars advocate codification and standardization of *Outer Circle* “norm-developing” varieties, they describe *Expanding Circle* varieties as “norm-dependent”, English as a foreign language (EFL) varieties, which are learned for communication with native speakers (Bolton, 2004).

Given the predominant NNS to NNS oral interaction in international English uses, the traditional prescriptivism of English language teaching (ELT) in the *Expanding Circle*, which emphasizes the benefits of imitating the educated native speaker of SE, has become unacceptable. Therefore, the ELF paradigm emerged to support primarily the non-standard characteristics of English speaking through the “study [of] idiosyncratic features of English language usage which are showing signs of becoming systematic, function in communication between interlocutors, and potentially provide speakers with markers of identity in the social group with which they identify (and act as well as an alternative to ceremonially joining the Anglo-American sphere of influence when using English)” (Modiano, 2009, p. 209). Though seemingly a WE spinoff, the ELF paradigm has been criticized by WE scholars arguing that it excludes native speaker varieties and displays monolithic resemblance to Crystal’s WSSE, i.e. neglects the polymorphous nature of English (Rubdi & Saraceni, 2006). In defense, Jenkins (2006a) emphasizes the non-exclusive, pluricentric orientation of ELF:

ELF researchers do not believe in any such monolithic variety of English does or ever will exist. Rather, they believe that anyone participating in international communication needs to be familiar with, and have in their linguistic repertoire for use, as when appropriate, certain forms (phonological, lexicogrammatical, etc.) that are widely used and widely
intelligible across groups of English speakers from different first language backgrounds. That is why accommodation is so highly valued in ELF research. (p.161)

The phonological aspects, especially pronunciation, have received focused attention in ELF research, which is primarily based on corpus data. Jenkins (2000) has carefully developed the *Lingua Franca Core* (LFC), which is a list of pronunciation features she argues essential for intelligible communication in ELF contexts. The LFC is intended to redefine and re-conceptualize pronunciation error accepting the sociolinguistic facts of regional accent variation as opposed to regarding deviation from NS pronunciation erroneous. According to ELF, NS accents may be desirable, but are not imposed as a norm, but rather a point of reference and approximation. The increased use of oral English tests for uses in ELF contexts has led to revisions and re-definitions of speaking proficiency scales. For instance, the Test of English Proficiency of Academic Staff (TOEPAS), used for oral English certification of academic staff intending to teach in English-medium instruction (EMI) programs at a Danish university, moved from the “educated native speaker” norm reference to allow for more accent variation in the scoring rubrics, particularly at the top scalar levels (Dimova & Kling, 2015; Kling & Dimova, 2015).

At the core of these debates are not only the issues of inclusion in the representation of IE, but also what constitutes intelligible and effective cross-cultural/international communication.

*Intelligibility from WE perspective*

The ambiguity and prescriptivism of the early notions of intelligibility appear unacceptable for new conceptions of international communication in WE. Abercrombie (1949) argues that “language learners need no more than a comfortably intelligible pronunciation” (p. 120). In this argument, he defines ‘comfortably intelligible’ as “a pronunciation which can be understood with little or no conscious effort on the part of listener” (p.120). Catford (1950) confounds intelligibility in the effectiveness of interaction, measuring intelligibility levels based on appropriate interlocutor responses, and Bansal (1969: 14) measures
phonological intelligibility against normative standards, proposing that articulation and pronunciation of sounds have to be clear and correct and they should not pose any listening difficulties for the hearer.

Similarly, Kenworthy (1987) operationalizes intelligibility through the level of listener's understanding of an utterance (p. 13). In her intelligibility definition, the more words the listener is able to accurately identify without repetition and clarification, the more intelligible the speaker is. Clarity and accuracy as features of pronunciation can be found in band descriptors of L2 speaking scales. For instance, the pronunciation category in the IELTS band descriptors refers to a “range of pronunciation features with precision and subtlety” (IELTS, 2015).

Smith’s tripartite model for successful cross-cultural communication in WE (Smith & Nelson, 1985) outlines three different levels of understanding of utterances: intelligibility, comprehensibility, and interpretability. The model, also known as the Smith paradigm (Nelson, 2008, p. 301), places pronunciation in the intelligibility category, which is the least complex of the three. Unlike the early notions of intelligibility, this model distinguishes between intelligibility and comprehensibility, the first focusing solely on phonological aspects of language, while the latter includes the meaning of the utterance. The most complex category in the model is interpretability, referring to how the listener interprets the intended meaning behind the utterance. In other words, the three categories can be placed on a complexity continuum, ranging from intelligibility, which represents word/utterance recognition, comprehensibility, which represents “locutionary force”, and interpretability, which represents “illocutionary force” (Nelson, 2011).

According to Nelson (2011), intelligibility “is the level of language use with the fewest variables, as it involves just the sound system” (p. 32). He goes on to state that, “Far from being an issue only across ‘native’ and ‘non-native’ varieties, intelligibility is a concern across any varieties, whether broadly or narrowly construed” because the words tend not to be lexicalized applying the same phonology across different varieties (p. 33).

Intelligibility, though, does not solely rely on the phonological accuracy of speaker’s oral production or the hearer’s perceptions because it is co-constructed
through interaction between the speaker and the hearer (Gumperz, 1992; Smith & Nelson, 1985). According to WE and discourse scholars, intelligibility depends on a number of interconnected factors related to the speaker, the interlocutor, and the linguistic and social context (Field, 2003; Pickering, 2006). As Nelson points out, “being intelligible means being understood by an interlocutor at a given time in a given situation” (1982, p. 59).

In empirical research, intelligibility has often been operationalized as phonological recognition of words, and the most common measures have been self-reported intelligibility based on a Likert scale, cloze tests (e.g., every sixth word removed from the transcript), word by word transcriptions, and partial dictation. Unlike intelligibility, comprehensibility has been measured inconsistently because of operationalization difficulties. Though multiple-choice comprehension questions and story summarizing have also been used to examine the comprehensibility of non-native speech (Gass & Varonis; 1984; Varonis & Gass, 1982), the most common comprehensibility measure has been self-reported comprehension on a Likert-type scale.

For example, Munro and Derwing (1995) and Derwing and Munro (1997) designed methods for assessing intelligibility and comprehensibility. Intelligibility was measured by the level of accuracy with which native-speakers wrote each word they heard. Comprehensibility, on the other hand, was measured by subjects’ own perceptions of understanding represented on a nine-point Likert scale. Since the comprehensibility measure was based on listeners’ judgments, it was termed “perceived comprehensibility”.

Though different methods have been employed to delineate intelligibility and comprehensibility in empirical research (Munro & Derwing, 1995; Derwing & Munro, 1997), in much of the research of pronunciation, intelligibility and comprehensibility have not been clearly delineated and are referred to interchangeably and are inextricably represented (Nelson, 2011).

*Intelligibility from ELF perspective*
ELF focuses on the narrow sense of intelligibility (phonological intelligibility). Jenkins recognizes the primacy of the phonological aspects of speech in intercultural communication and argues that though the lexicogrammatical and pragmatic meanings are important, the very first noticeable characteristic of speech is pronunciation. Though seemingly different from the Smith paradigm, the ELF intelligibility paradigm arguably shares paradigmatic commonalities (Berns, 2008). Juxtaposing the two paradigms, Berns (2008) points to the different theoretical frameworks Smith and Jenkins apply to describe speech understanding, i.e. cross-cultural communication, and general linguistics and speech act theory, respectively. Despite the apparent correspondence between the two models, Berns posits that equating Smith’s intelligibility, comprehensibility, and interpretability with Jenkins’ accent, propositional content, and pragmatic sense may not be as simplistic as it seems due to the different theoretical underpinnings of the two models (p. 328).

Given the significance of pronunciation in intelligibility of international communication, identification of the contributory phonological features seems essential. The Lingua Franca Core (LFC) (Jenkins, 2000: 2002) is the most frequently recognized in current research. However, the early work by Gimson (1978, 1994) initiated the discussion by proposing the “rudimentary international pronunciation” system for NNS, which allowed for sound modification if they bear little influence on “minimum general intelligibility.” According to Gimson’s model (1994) most modifications could be licensed in the vowel system whereas only minor divergence from the NS norm is allowed for consonants (see also Field 2005). Jenner (1989) first proposed the notion of a pronunciation core, which was further developed by Jenkins (2000; 2002)

Through NNS corpus analyses of the phonological features associated with communication success or breakdowns, Jenkins has found that that intelligibility could be achieved if the following pronunciation characteristics are maintained:

1. Accurate pronunciation of most consonant sounds + one vowel (/ɜː/)
2. Preservation of most consonant clusters
3. Vowel length (especially before voiced/unvoiced consonants)
4. Appropriate word grouping and placement of nuclear stress
Unlike the emphasis on stress and rhythm in early intelligibility work (Bansal, 1969), Jenkins’ (2000; 2002) work suggests that word stress and tone are not core phonological elements of ELF.

While ELF promotes NNS to NNS communication, in speaking assessment, which commonly encompasses assessment of pronunciation and intelligibility, communication is mostly assumed between NNSs and NSs, so produced speech is expected to be “understandable” for the NS. In the ACTFL Proficiency Guidelines (2012), for example, “[E]rrors virtually never interfere with communication or distract the native speaker from the message” (ACTFL, 2012). Nevertheless, according to the rater certification manual, raters are not required to be native speakers (ACTFL, 2015). Though pronunciation is not explicitly used in the ACTFL level descriptors, in the highest level, the manual allows for accent variation at the highest level though it is expected to follow native-like patterns, “[A] non-native accent, a lack of a native-like economy of expression, a limited control of deeply embedded cultural references, and/or an occasional isolated language error may still be present at this level “ (p.4).

Uses of “intelligibility” in Second Language (L2) speaking proficiency scales
Pronunciation assessment is commonly integrated in speaking assessment, either in holistic speaking rubric descriptors or in pronunciation (or fluency) subscales of analytic speaking rubrics. Understanding the role of pronunciation in the larger speaking construct remains essential, though its operationalizations tend to be inconsistent and vague. While some scales lack pronunciation references (e.g., the Common European Framework of Reference (CEFR) (Council of Europe, 2001; North, 2000), other are “strikingly random in describing how pronunciation contributes to speaking proficiency” (Levis, 2006, p. 245, in reference to the ACTFL scale). For instance, the Test of Spoken English (TSE) rating scale (1995), often used for screening international teaching assistants (ITAs) at U.S. universities, includes pronunciation as a feature of speaking proficiency, yet the construct remains
underdeveloped, and the links to intelligibility or comprehensibility are absent (Educational Testing Service, 1995).

According to Isaacs and Trofimovich (2011), even when included, the inconsistency of pronunciation descriptors leads to construct underrepresentation. The use of the term pronunciation is inconsistent across scales for speaking assessment because it may simply refer to segmental features (i.e., errors that involve individual sounds) or include suprasegmental features (e.g., word stress, rhythm, intonation).

The term intelligibility is present among scalar descriptors of oral proficiency in several L2 tests (Dimova & Jensen, 2013). For example, the Test of English as a Foreign Language (TOEFL iBT) scoring rubric for speaking clearly links “pronunciation” and “articulation” with “intelligibility” and “listener effort” (ETS, 2015). Similarly, the International English Language Testing System (IELTS) Speaking Band Descriptors provide relations between “mispronunciation” at word and sound level, “L1 accent” and “understanding” “intelligibility” and “difficulties for the listener” (IELTS, 2015). It is worth noting that IELTS seemingly makes use of “understanding” in broader, more general sense, and “intelligibility” in narrower, more local sense to discuss pronunciation, which likens WE conceptualizations of intelligibility and comprehensibility.

The Oral English Proficiency Test (OEPT), a semi-direct screening test for oral English proficiency of international students assistants (ITAs) at an U.S. Midwestern university includes holistic scale descriptors specifically related either to listener requirements or speaker performance characteristics (OEPT, 2015). Though not measured separately, both terms, intelligibility and comprehensibility, are included in the speaker performance description, with intelligibility being affected by “marked L1 features.” Unlike the holistic approach to intelligibility in the OEPT scale, the ESL Placement Test (EPT), which also assesses prospective ITAs and NNS students, isolates intelligibility as its main measure. The first part of EPT is a three-minute interview in which the rater, who can be a NS or NNS, assesses the candidate’s unrehearsed speech on the basis of rater’s ability to understand every word that the interviewee utters (Isaacs, 2008), which is similar to Munro and
Derwing’s intelligibility measure (Munro & Derwing, 1995; Derwing & Munro, 1997). However, like pronunciation, intelligibility is rarely assessed in isolation in the context of L2 testing, as it is commonly embedded, either holistically or analytically, in rating scales of speaking. This means that intelligibility tends to be measured through raters’ subjective perceptions, which suggests that it is comprehensibility rather than intelligibility that is used as a criterion in these scales (Isaacs, 2008).

To sum up, pronunciation and intelligibility assessment are subsumed in assessment of speaking, frequently occurring among descriptors of the lower scalar levels and being related to raters’ subjective perceptions of produced speech. Despite research findings suggesting that accent and intelligibility are independent (Derwing & Munro, 1997; Smith & Rafiqzad, 1979), in speaking rubrics, accentedness and L1 influence seem to be commonly indicated as factors affecting intelligibility levels of L2 speech performances.

**Intelligibility factors in the WE context: familiarity and accentedness**

A number of studies in WE have focused on identifying the factors affecting the levels of intelligibility of different English varieties. Many of them compare varieties across the three circles of Englishes (inner, outer, and expanding), but studies comparing only inner circle varieties can also be found in the literature. Some of the investigated factors include familiarity, accentedness, and attitude towards an English variety. As early as in 1950, Catford suggests the need to identify a “threshold of intelligibility” (p. 14), i.e. how much exposure to a language or a variety a user needs in order to become familiar with it. In other words, users with high exposure to the variety experience greater intelligibility. More familiarity, in turn, may reduce resistance and influence the “perceived attitudes” towards the variety.

An influential study by Smith and Rafiqzad (1979), suggests that intelligibility and comprehensibility do not seem to be linearly correlated with degrees of foreign accentedness. In their seminal study involving 1,300 participants in eleven countries, they found that native-speaker phonology is not necessarily
more comprehensible than non-native phonology, which negates the widespread assumption of the supremacy of the NS accent. Findings are surprising because they show that the recordings of the American and the Hong Kong Chinese readers were least intelligible, while those of the Japanese, the Indian, and the Malaysian were among the top five most intelligible.

Despite Smith and Rafiqzad’s results, subsequent research findings seem to relate native and local varieties with higher intelligibility levels. The role of exposure to native varieties in the development of intelligibility is supported by Smith and Bisazza’s findings (1982). In their study, each of the three different forms of the listening comprehension test (Michigan Test of Aural Comprehension) was recorded by an Indian, a Japanese and an American speaker of English, and then administered to university students in English as a foreign language (EFL) (Japan, Taiwan and Thailand) and English as a second language (ESL) contexts (Hong Kong, India and the Philippines). According to their findings, the American speaker was easiest, while the Indian speaker was most difficult to understand. The researchers believe that these findings stem from the participants’ higher exposure to American English compared to the other two varieties because NS norms are preferred in EFL and ESL instruction.

Higher comprehensibility of American and British native speakers was also found by Ortmeyer and Boyle (1985). In addition, they found that proficiency levels significantly interacted with comprehensibility of NS or NNS accents. In their study, they administered listening comprehension and dictation tests, including recordings from an American, a British, a “clear” Chinese, and an “unclear” Chinese speaker to 228 students at the Chinese University of Hong Kong. Students, especially lower proficiency level, scored higher when listening to American and British English accents than when listening to Chinese accents.

Taking into account the similar findings from Smith and Bisazza’s (1982) and several other studies (Brown, 1968; Ekong, 1982; Wilkox, 1978), Flowerdew (1994) conclude that students find it difficult to comprehend “unfamiliar” accents. In other words, students are most likely to understand the accents of lecturers who share their first language or the accents used in “society at large”, i.e. those used in
instruction (e.g., American English in Taiwan). However, Tauroza and Luk (1997) and Pikho (1997) indicate that though comprehension seems definitely aided by accent familiarity, whether this is the local accent becomes a secondary issue. Based on findings from an experiment with 63 Hong Kong school students who listened to Hong Kong English and Received Pronunciation, Tauroza and Luk modified the familiarity hypothesis adding that non-local but familiar accents, rather than solely local accents, can also be comprehensible for L2 listeners.

To gain improved understanding of the relationship between familiarity and comprehensibility, Gass and Varonis (1984) deconstruct the familiarity concept into four variables. The four variables facilitating NS comprehension of NNS accents were: familiarity with the topic, familiarity with NNS speech in general, familiarity with a particular NNS accent, and familiarity with a particular NNS. Results confirm that all familiarity variables affect comprehensibility, though familiarity with the discourse topic seemed to greatly facilitate message interpretation.

While a number of studies included speakers and listeners from countries across the three circles of Englishes, or NS listeners and NNSs, Matsuura, Chiba, and Fujieda (1999) investigated the relationship between NNS familiarity on intelligibility of two different NS varieties, American English and Irish English—the former being more widely spread in Japan than the latter. Results pointed to the relationship between NS intelligibility and proficiency levels, but also to a discrepancy between subjects’ actual and perceived comprehensibility. In other words, higher language proficiency, rather than familiarity with the variety, is more strongly associated with intelligibility even though familiarity leads to higher perceived comprehension levels. The authors conclude that exposure may promote “less bias and more tolerance toward different varieties of English,” but this did not necessarily mean “better understanding of the message” (Matsuura et al., 1999, p.58).

In summary, research suggests that exposure to a particular variety leads to increased familiarity and positive attitudes, and possibly higher intelligibility levels. This means that accent familiarity must be considered during rater training and behavior analyses, especially with regard to assessment of pronunciation and
speech performance intelligibility. Given the different experiences test candidates have in terms of exposure to local English varieties, and hence different intelligibility levels, the selection of varieties to be included in listening comprehension tests needs careful scrutiny as it could potentially lead to bias against certain groups.

**New directions in assessment of L2 pronunciation in relation to listening and speaking**

Discussions about pronunciation assessment from the WE and ELF perspectives have been generally been rooted in the contexts of speaking and listening assessment. The main critiques of pronunciation assessment in relation to speaking have dealt with the reliance on NS pronunciation norms and lack of accommodation. Listening assessment has been blamed for NS pronunciation norm dominance and limited NNS accent variety representation in the listening tasks.

**WE and ELF perspectives on pronunciation in relation to speaking assessment**

Despite the critiques of the relatively conservative practices in speaking assessment, certain changes have been triggered by the current discussions in WE and ELF. References to NS competence in the assessment criteria are no longer predominant as they used to be in the past (Kling & Dimova, 2015) due to the elusive nature of the NS construct (Davies, 2002; 2003). Though these changes may seem recent, Weir (2013) claims that prominence of the “native speaker” construct in assessment of speaking began decreasing even at the beginning of the 1980’s.

The “deficit model” for NNS oral production has been abandoned and substituted by “can do” statements, which focus primarily on function and communication. For example, pronunciation criteria are distanced from NS imitation, accuracy and correctness, but rather comprehensibility and communicative effectiveness, allowing for more accent variation at the higher proficiency levels (Taylor, 2006). Moreover, NSs who have traditionally been responsible for rating of L2 speaking performances (Lazaraton, 2005; Lowenberg, 2002; Seidelhofer, 2001) despite irrelevance of the SE norms in many testing
situations (Graddol, 1997), are no longer viewed as the exclusive keepers of SE pronunciation norms -- an increased number of NNS raters have been involved in the rating process. These changes are certainly in line with WE and ELF propositions for inclusions of varieties.

A number of studies have compared NNS rater behavior to that of NS raters in terms of rating consistency and application different rating criteria and standards. Results from these studies suggest high scoring consistency between NNSs and NSs, though NNSs apply different rating criteria (Kim, 2009). In fact, though NNSs have the potential to reveal the main non-SE criteria utilized in the ELF context, they can adhere to SE norms even more prescriptively than NSs (Zhang & Elder, 2010).

Supporting previous findings from Carey, Mannell, and Dunn (2011), who found a tendency among IELTS examiners to rate pronunciation higher when they had prolonged exposure to test-takers’ L1, Winke, Gass, and Myford (2011; 2012) also found accent familiarity effect on raters’ behavior, leading to rater bias. They defined accent familiarity as having learned the test-takers’ L1. Results suggest that L2 Spanish raters and L2 Chinese raters were significantly more lenient with L1 Spanish and L1 Chinese test takers respectively. However, rater bias can be effectively minimized when appropriate rater training programs are implemented (Xi & Mollaun, 2009; 2011).

Lack of opportunities for accommodation in speaking tests has also been criticized as it underrepresents authentic situations (Jenkins, 2006b) -- in real life we constantly adjust our speech to improve intelligibility and accommodate our interlocutors. Taylor (2006) argues that examiners of Cambridge ESOL’s speaking tests are trained to follow certain interlocutor scripts to maintain fairness, and when it comes to pronunciation, they need to consider the following:

Examiners should try to put themselves in the position of a non-EFL/ESOL specialist and assess the overall impact of the pronunciation and the degree of effort required to understand the candidate. Clearly the emphasis is on making oneself understood rather than on being ‘native-like’. (p. 55)
Perhaps accommodation has been best addressed with the design and implementation of L2 speaking tasks based on a paired speaker assessment model (see Certificate of Proficiency in English, in Macqueen and Harding, 2009). Though paired assessment has traditionally been negatively viewed due to the potential variability of interlocutors (age, gender, proficiency level, etc.) and the possibility for one interlocutor to dominate the conversation (Iwashita, 1998; O’Sullivan, 2000), recent research points to the advantages of this assessment method. In particular, research suggests that interlocutor variability in pair work can be considered as strength, rather than a threat, because it allows for elicitation of negotiation in collaborative interaction, based on the same requirements for accommodation as in the real-world communicative exchanges (Ducasse & Brown, 2009; Taylor & Wigglesowrth, 2009; May, 2009).

**WE and ELF perspectives on pronunciation in relation to listening assessment**

The NS norm reliance critique in listening tests refers to the misrepresentation of the variety-rich target-language use (TLU) domains for which these tests are designed (e.g., universities, business, healthcare, aviation). Given that in the real world, communication will occur among NS and NNS of English varieties under different conditions, inclusion of NNS accents in listening tests could yield enhanced authenticity (Bejar et al., 2000; Harding, 2012). For example, English proficiency tests used for university admission at US and Australian universities, such as the TOEFL and the IELTS, employ a range of NS accents in the listening sections -- TOEFL iBT includes British and Australian in addition to the North American accents (see ETS, 2005), and IELTS covers British, Australian, North American, and New Zealand accents (see IELTS, 2015). However, students at the US and the Australian universities encounter different NNS in addition to the NS accents represented in the tests because of the number of international faculty members and international teaching assistants (ITAs). Moreover, these tests are increasingly used for admission purposes at English-medium instruction (EMI) programs in different European and Asian universities, where NNS, rather than NS, varieties are prevalent. Students in an EMI program at a Danish university are more likely to have
Danish speakers of English among their instructors than any of the NS varieties represented in the listening sections of these tests (Dimova & Kling, 2015).

From the washback perspective, inclusion of NNS varieties in listening tests may increase the amount of exposure to different NNS varieties in the English learning classroom (Jenkins, 2006b). Given that exposure to and familiarity with a variety improves intelligibility levels, and that learners are more likely to find themselves in NNS to NNS communicative situations, increasing the range of NNS varieties in classroom teaching would improve students’ communicative skills. Ultimately, this could lead to a wider acceptance and recognition of NNS varieties.

The effects of accent familiarity on listening input has been the focus of an ETS funded study which investigated the effect of nonstandard accents on listening task scores (Major et al. 2002). The researchers used a “Listening Comprehension Trial Test,” which was based on the TOEFL listening section and included lectures delivered in accented English by NS of Chinese, Japanese, Korean, and Spanish. The test was administered to listeners who were NS of the same languages. Findings suggest that irrespective of L1 background, listeners scored higher when they listened to the Spanish accented speech, which was similar to the scores associated with the American accented speech. Though these findings may lead to rejection of the assumption that one’s own NNES accent should be more intelligible than others NNES accents, they are inconclusive because of incomparable task difficulty and limited range of accents.

More recent research has found that though possible, the familiarity effect or shared-L1 advantage is not prevalent in all circumstances (Harding, 2012). These findings are based on a study in which 212 L2 listeners, among whom Mandarin Chinese L1 speakers and Japanese L1 speakers, listened to three versions of a listening subtest featuring Australian English accented speaker, a Mandarin Chinese accented speaker and a Japanese accented speaker. Results from DIF analyses suggested slight advantage for Japanese L1-shared speakers but clear advantage for Mandarin Chinese L1-shared speakers.

In terms of accent variability in listening input, a much earlier study (Powers, 1986) investigated the validity of TOEFL listening tasks by administering a survey to
146 academic university staff across different fields. The survey asked about the language and listening demands on NSs and NNSs. Findings suggest that coping with different accent and dialects was problematic for NNS, but speech rate was even more challenging. The implication is that introduction of accent variety in listening input may pose disadvantage to some candidates, especially if they are delivered with higher speaking speed.

Though not extensively, research in assessment of pronunciation in relation to speaking and listening assessment has added to an improved understanding of intelligibility, both from WE and ELF perspectives. Research contributions have mainly focused on familiarity and accent as important intelligibility factors in speaking raters’ behavior and test-takers’ performance on listening tests. However, more research is certainly needed to understand how accent variation in pronunciation assessment fits in the broader constructs of listening and speaking we attempt to measure.

**Conclusion**
This chapter has discussed the WE and ELF perspectives and research on norm selection and intelligibility with implications for assessment of pronunciation. The central argument in these discussions is the inadequacy of the NS model both for pronunciation teaching and testing purposes. Inclusion of *Outer Circle* and ELF models in teaching and testing is proposed to legitimize the status of the colonial varieties and establish realistic expectations about uses of English for international communication. Consequently, the increased exposure to different NNS pronunciation varieties would result in improved accent familiarity and, therefore, an increased degree of accent intelligibility.

The reliance on the NS pronunciation standards in language testing has been challenged by WE and ELF scholars, but the lack of systematic codification of the different outer-circle and ELF varieties creates constraints for their inclusion in high-stakes testing. Though presence of a range of NNS pronunciation varieties, particularly in listening tests, could enhance task authenticity and lead to positive washback, current research has yielded mixed results, rendering the findings
insufficient to confirm the validity of NNS variety inclusion. Concerns about test bias and practicality are yet to be adequately addressed (Taylor, 2006; Taylor & Geranpayeh, 2011).

Nevertheless, research in the field has contributed to the investigations related to English uses in international contexts, and despite the existing constraints, strides have been made towards encompassing a WE perspective in test construction and task design, especially in pronunciation assessment as implied in listening and speaking. However, more radical departures from current testing practices cannot be expected nor are they warranted until more stable definitions of ELF pronunciation norms are provided by researchers within these paradigms and the intelligibility concept is clarified.

Even though Nelson (2011) claims that intelligibility is the simplest level involving only the phonological system, the existing knowledge about intelligibility factors remains unsatisfactory. Some connection between accent familiarity and intelligibility has been established, but what constitutes familiarity (e.g., shared-L1, learnt-L2, exposure, attitudes) has received limited attention. Moreover, if intelligibility is co-constructed (Gumperz, 1992; Nelson, 1985), then the interconnected factors related to the speaker, the interlocutor, and the context need careful examination.

To conclude, a wider acceptance of accented, rather than native-like, pronunciation at the higher end of speaking rating scales may occur if we clearly recognize the characteristics of a highly intelligible accented speech. Obtaining a firmer grasp of the interactional nature of intelligibility could lead to design of tailor-made rater training programs which specifically address the differences of various rater groups (e.g., NS, NNS, shared-L1). Finally, identification of the intelligibility factors could assist in the design of listening inputs with a variety of highly intelligible NNS accents without compromising validity in listening assessment.

References


IELTS [International English Language Testing System]. Understand the listening test. Retrieved on 20 August 2015 from


Whitstable, England: IATEFL.


