2013

Language testing and international intelligibility : a Hong Kong case study

Andrew SEWELL
Lingnan University, Hong Kong

Recommended Citation

INTRODUCTION

In recent discussions of language testing, a major topic has been the appropriacy of competing norms. Davies, Hamp-Lyons and Kemp (2003, p.572) describe the opposing viewpoints as being "International English" (more-or-less homogeneous, based on exonormative, native-speaker models) and "World Englishes" (heterogeneous, based on endonormative or local models). This debate has a long history. Exchanges between the two camps led to the coinage of epithets such as "liberation linguistics" (Quirk, 1990) to describe attempts to decenter the English language. But as Canagarajah (2006) notes, the ground upon which the debate arose changed with the advent of "postmodern globalization" (Hall, 1997). In a world now characterized socioeconomically by multilateral interaction and the porosity of borders, the accompanying linguistic phenomena include hybridity and transcultural flows (Appadurai, 1996; Pennycook, 2007). While the impossibility of imposing a single standard is now taken almost for granted – except perhaps in the case of written English in academic contexts – the nation-based model of World Englishes depicted by Kachru's three concentric circles (Kachru 1992) is also incapable of capturing the complexity of local language use (Mahboob and Szenes, 2010).

Language testing is thus faced with the challenge of trying to do justice to these phenomena, while adhering to more established principles: reliability and validity, community acceptability, and pedagogical compatibility. Hong Kong presents an interesting case study. Despite being at the forefront of economic globalization, language teaching and testing remain largely anchored to native-speaker norms. With exonormative attitudes being widespread among teachers (see Tsui and Bunton, 2002) and students (see Li, 2009), there has been little discussion of the possibility of developing local norms for testing, perhaps along the lines described by Brown and Lumley (1998) in the case of Indonesia. A partial explanation lies in the widely-held community view that Hong Kong English equates with bad English, or that its use may reduce international intelligibility. As Taylor (2006, p. 58) observes, test users (including teachers, employers and admissions officers) generally expect tests "to contain English varieties that are widely recognized and used in the pedagogic contexts to which they relate". Among linguists, Hong Kong English is usually seen as an emergent variety (Setter, Wong and Chan, 2010), one which is at the nativization stage in Schneider's (2003) dynamic model. Hong Kong appears to be in the process of developing its own norms, and there is a justified concern that attempts to codify local norms may be premature.

But at the same time, there is a kind of linguistic schizophrenia (Kachru, 1977) involved in expecting speakers of English in Hong Kong to conform to native-speaker norms. As Groves and Chan (2010, pp. 48-49) observe, it is unrealistic to expect them "to conform to native speaker norms in all respects, especially considering their input is largely from other locals, whose English exhibits the same features". The "other locals" here are local language teachers, who are faced with a native-speaker model "that they themselves do not speak" (Kirkpatrick 2006, p. 79). There are potential advantages for teachers and students, then,

in considering how local norms could be reflected in teaching and testing materials. It should be emphasized that reflecting local norms does not mean replacing one code ("Standard English") with another (e.g., "Hong Kong English"). Given the inter-speaker variation that exists and the consequent lack of agreement regarding what constitutes a "norm", this would be impracticable.

Instead, one possibility is to consider which features of the local variety are most likely to affect international communication. There is a body of research (Jenkins, 2000; Deterding and Kirkpatrick, 2006) which suggests that certain phonological features of English play an important role in the maintenance of international intelligibility (usually defined as relating to communication between L2 speakers of English (Ferguson, 2006, p.165), but potentially also applicable to encounters between L1 and L2 speakers). These features constitute a "core", around which variation is permissible. The advantage of a "core" approach is that it can provide features-based guidelines for teaching and testing, so that curricula and syllabi can be reviewed for their relevance, and test criteria can be reviewed for their alignment with the requirements of international communication. Of course, this rests upon the premise that international intelligibility is a relevant criterion. This article will review and evaluate the concept in some detail, but there appears to be a *prima facie* case for the acceptance of certain features of the local variety in local speaking tests, if they can be shown to be a) widespread amongst users of that variety, and b) inconsequential for intelligibility.

A similar argument has been made for using L2 accents in listening tests (see Harding, 2011). This would seem to have the advantage of making tests "appropriate for the intended test-taking populations" (Joint Committee on Testing Practices, 1988; cited in Brown and Lumley 1998, p. 84). If the possibility of washback is accepted, there may also be pedagogical benefits, such as more efficient use of classroom time. However, as Taylor (2006, p. 58) reminds us, testing is very much "the art of the possible". Brown and Lumley's Indonesian case study found that a test based on local norms was "generally well received", but noted that "the concept of the non-native model is not yet fully accepted" (Brown and Lumley 1998, p. 93). The answer to the question of whether local language norms should be reflected in teaching materials and tests thus depends on various factors, and requires contextually-sensitive research.

This article first provides some background to language proficiency testing in Hong Kong, focusing on the spoken component of the Language Proficiency Assessment for Teachers of English (hereafter, LPATE). It then introduces the concept of international intelligibility (e.g., Jenkins, 2000) and considers its relevance for language testing. Its main purpose is to exploit the availability of examiner comments relating to the pronunciation component of the LPATE. By analyzing these comments, it attempts to assess the degree of alignment between the test descriptors and the criteria for international intelligibility; a lack of alignment might suggest that candidates are being penalized for using local language features

that do not threaten international intelligibility. Finally, the article discusses the implications for speaking tests of a similar nature.

LANGUAGE TESTING IN HONG KONG AND THE LPATE

In Hong Kong, the LPATE was introduced in 2001 as a benchmarking examination, to ensure that the language proficiency of local English teachers reaches a required standard. There are a total of five papers: Reading, Writing, Listening, Speaking, and a Classroom Language Assessment (or CLA). This article will focus on pronunciation, which is assessed in two papers (Task 1A of the Speaking paper and the CLA) using a *Pronunciation, Stress and Intonation* scale. In the LPATE examination, the term "pronunciation" appears to refer to segmental features such as vowels and consonants, while "stress and intonation" covers suprasegmental features such as stress and intonation; in this article, "pronunciation" is used more generally to refer to both segmental and suprasegmental features.

Task 1A of the LPATE involves reading aloud a prose passage, and usually takes around two minutes. The sample passage in the LPATE Handbook (Education Bureau, 2010, p.74) is an excerpt from a story by Haruki Murakami, and candidates are instructed to read it aloud "as meaningfully as possible" as if they were in front of a class of students. (There are potential problems with this task, including its authenticity as a classroom activity, the effects of orthography and unfamiliar lexis, etc., but these are beyond the scope of this paper.) The CLA is a test of candidates' language use in actual classroom settings, and a continuous period of 20 minutes' teaching is the minimum duration for assessment. The Speaking paper also features two other tasks, Task 1B (a recount of personal experience, or the presentation of arguments based on a stimulus) and Task 2 (a group discussion), but neither of these tasks involves a direct assessment of pronunciation.

The exact weighting of pronunciation (in the general sense of the term) in the overall distribution of marks is unknown, but its assessment in two of the five papers suggests that it is seen as an important part of an English teacher's professional competence. The LPATE Handbook (Education Bureau, 2010, p.70) describes pronunciation as an "important assessment category" in the Speaking test. A more detailed picture of the relative importance of pronunciation can be gained from considering the nature and application of the assessment scales. The LPATE is a standards-referenced assessment with six scales or descriptors of performance, the application of which depends on the paper and task. Tasks are assessed using more than one scale, and the same scales are used with more than one task. For example, Task 1A of the Speaking paper is assessed using the *Pronunciation, Stress and Intonation* scale and the *Reading Aloud with Meaning* scale. The CLA is assessed using four scales: *Pronunciation, Stress and Intonation, Grammatical and Lexical Accuracy and Range, The Language of Interaction* and *The Language of Instruction*. Summarized descriptors from the *Pronunciation, Stress and Intonation* scale are shown below in Table 1.

TABLE 1

LPATE band descriptor summaries for *Pronunciation, Stress and Intonation** from the LPATE Handbook (Education Bureau, 2010, p.71).

LPATE band	"Pronunciation, Stress and Intonation" descriptor
5	Reads in a fully comprehensible way with no systematic errors in pronunciation and uses stress and intonation in a very natural way.
4	Reads in a comprehensible way with few systematic errors in pronunciation and uses stress and intonation in a mostly natural way.
3	Reads in a generally comprehensible way, though may make errors in pronunciation. Uses stress and intonation to convey meaning, though may occasionally sound unnatural.
2	Does not read in a consistently comprehensible way due to errors in pronunciation, stress and intonation and speech is frequently hesitant.
1	Makes frequent errors in pronunciation, stress and intonation which cause confusion for the listener.

* Used in Task 1A of the Speaking paper and in the Classroom Language Assessment.

The LPATE Handbook states that Level 3 is the required level of proficiency in each paper (Education Bureau, 2010, p. 2). Grades are averaged in papers with more than one assessor (there are two in the Speaking paper, and either one or two in the CLA), and the minimum standard is set at 2.5 in any one scale and 3 in all other scales. LPATE assessors include both expatriate native speakers and local bilingual speakers with a high level of English proficiency; most are university teachers, and their training includes standardization using video and classroom observation (Coniam, 2003, p.58).

Of the 643 teachers who took the LPATE test in 2003, 333 failed (Morris, 2004, p. 19). Understandably, as the authorities required schools to redeploy or dismiss those failing the test, the LPATE became a topic of bitter controversy in Hong Kong. But much of the controversy has centred around the design of the test, including the requirement for candidates to pass in all five areas (Reading, Writing, Listening, Speaking and the Classroom Language Assessment), and its difficulty level, with reports that some native speakers have also failed to pass certain papers (Morris, 2004, p. 20). The test descriptors do not seem to have attracted much attention, although Coniam (2003, p.58) criticizes the vagueness of certain phrases within them, from a prospective candidate's viewpoint. Luk and Lin (2006) provide an account of an LPATE standardization meeting in which several commonly-occurring features of Hong Kong English – such as replacing /ð/ with [d] – were classed as "errors". This feature is unlikely to affect comprehensibility (see, e.g., Jenkins, 2000), and at first glance suggests a less-than-accommodating attitude towards local pronunciation features.

Accounts of the process of test development (e.g. Coniam and Falvey, 2002; Morris, 2004) reveal that the issue of local norms was in fact considered in some detail. It was agreed as early as 1997 that the model required from teachers would be recognizable as that of an L1

Cantonese speaker, while not deviating markedly from the "educated" forms of other varieties, such as Singapore English (Coniam and Falvey, 2002, p. 18). The reasons for this attitude towards Hong Kong English accent features were twofold: firstly, to avoid charges of neocolonialism during the sensitive period of the 1997 handover of sovereignty from Britain to China; and secondly, to avoid proposing an unattainable and therefore demotivating model for teachers (Coniam and Falvey, 2002, p. 28).

There is nothing in the descriptors themselves to suggest that the LPATE is unable to accommodate local accent features, while simultaneously requiring comprehensibility. They do, at least, avoid the assumption that the native speaker is the benchmark or target. As Taylor (2006, p.52) points out, many language tests no longer make reference to native-speaker competence, and some charges of norm inappropriateness (e.g. Jenkins, 2006) are therefore misdirected. The crucial question in this case is how terms such as "comprehensible", "errors" and "natural" are interpreted and applied by the assessors. The empirical part of this article will examine this issue, by comparing examiner comments with criteria for international intelligibility. This concept is explained in more detail in the following section.

INTERNATIONAL INTELLIGIBILITY

The need to maintain international intelligibility was one of the considerations that guided the development of the LPATE test, and oriented it towards "educated" or "standard" features. This is understandable in view of Hong Kong's official branding as "Asia's World City", and the importance of international trade, especially financial services, to the local economy. In this section I will summarize research in the area, including the use of terminology, consider the relative importance of some segmental and suprasegmental features for intelligibility, and briefly evaluate the general usefulness of the concept of intelligibility for language testing.

While intelligibility as generally accepted to be one of the important constructs in assessing spoken interaction, there is great difficulty in arriving at a precise definition; Isaacs (2008) concludes that it is an "evasive" concept, about which little is known. Many studies distinguish between intelligibility and comprehensibility, although the two are closely interrelated. For some, the former is more local in terms of processing and relates to word and utterance recognition (e.g. Smith, 1992), while comprehensibility relates to the understanding of overall meaning. A similar interpretation is taken by researchers who measure intelligibility "objectively" through transcription tasks, and assess perceived comprehensibility by asking listeners to rate speech samples for ease of understanding (e.g. Munro and Derwing, 1999; see also Trofimovich and Isaacs, 2012). In some contexts the two terms are used interchangeably (Field, 2005). Although the LPATE descriptors in Table 1 refer to "comprehensibility", most of the examiner comments refer to mispronounced words, phenomena that are usually assigned to the domain of intelligibility.

There tends to be a focus on the intelligibility of non-native speakers for native listeners in many studies (e.g. Munro and Derwing, 1999; Hahn, 2004). One of the first investigations of international intelligibility, in which the attention shifts to non-native listeners and speakers, was that of Jenkins (2000). The study is noteworthy in that it provides empirical data on interactions between non-native speakers from different language backgrounds. The phonological features that caused communication problems were noted, and the culmination of Jenkins's research was the Lingua Franca Core (LFC) of features that play an important role in the maintenance of intelligibility. The LFC stipulations (Jenkins, 2000; 2007) are:

- All consonant sounds are important, with the main exception of the dental fricative or TH sounds (these can be substituted); slight variations in the realizations of consonant sounds are permissible, but not if they cause confusion with other sounds.
- Consonant cluster simplification is permissible, but only if it is 'appropriate' (Jenkins 2000, p. 132). This means that final clusters can be simplified according to NS patterns, e.g. the deletion of /t/ and /d/ in words like *facts* and *pounds*, or in word sequences like *last week*. Sounds in initial clusters should not be deleted.
- Vowel length differences (such as between *ship* and *sheep*) are important, but variations in quality are not.
- Nuclear (or tonic) stress is important, in order to separate 'thought groups' and focus on important information.

It should be noted that the LFC research has several limitations. For example, the database is relatively small, and there is uncertainty as to the representativeness of the examples upon which the LFC is based (see the edited volume by Dziubalska-Kolaczyk and Przedlacka, 2005, for further discussion). On the other hand, there has been some support from subsequent research. In their study of international conversations at an ASEAN meeting, Deterding and Kirkpatrick (2006) found that all five of the phonological features that caused intelligibility problems – mostly consonantal substitutions – were core features of the LFC. There is also some theoretical support for the LFC findings, from the concept of functional load (see Brown, 1991; Munro and Derwing, 2006). Functional load can be defined as the information-carrying capacity of a phonemic opposition, and depends on factors such as frequency of occurrence and phonetic similarity. The fact that substitutions of the dental fricatives did not cause intelligibility problems in the LFC research is probably due to their relatively low functional load (see Brown, 1991, for rankings).

An important indication of the LFC is that not all aspects of the English phonological system are equally important for international intelligibility, including those which may appear in language teaching syllabi and language test descriptors. In general, the LFC places a lower priority on suprasegmental features (with the exception of nuclear stress). This goes against the grain of the view that suprasegmentals make a greater contribution to international intelligibility (Nihalani, 2010), and should thus be emphasized in pronunciation pedagogy (e.g. Morley, 1991; Hahn, 2004). But as Hahn (2004, p.203) observes, there has been little empirical evidence to support this view, except perhaps in the case of word stress. Hahn's

(2004) study found that misplaced word stress reduced comprehensibility, as measured by tests of content recall and overall comprehension, for NS listeners. Similarly, Tsuzuki and Nakamura (2009) found that misplaced word stress caused intelligibility (word recognition) problems for NS listeners. Field (2005) investigated international intelligibility (word recognition for NS and NNS listeners), and also concluded that word stress had important effects.

On the other hand, it has been claimed elsewhere that certain suprasegmental features, such as linking, stress-timed rhythm and vowel reduction, are unimportant for intelligibility; Deterding (2010) and Walker (2010) argue that they may actually reduce it. The case for emphasizing segmental features is also strengthened by considering the nature of language processing for L2 listeners. Jenkins (2000, p.78) notes that L2 users, who lack familiarity with the language and its customary usage, are forced to place greater reliance on phonological form in order to decode meaning. Dalton and Seidlhofer (1994, p. 26) point out that non-native speakers are less able to make use of contextual information and consequently "rely – often exclusively – on acoustic information alone".

The segmental/suprasegmental question requires further investigation, with language users of different proficiency levels. It seems likely, however, that these areas are interdependent (Brazil, 1994, in Jenkins 2000, p. 135). Field (2005) observes that alterations in word stress often lead to changes in vowel quality and/or quantity, which may in themselves cause intelligibility problems. There may also be ways in which the perception of nuclear stress (a core feature) depends upon word and/or sentence stress (a non-core feature, according to the LFC). Language teaching and testing need to acknowledge both areas and their interdependence, and although there is clearly a need for further research, the LFC provides a possible indication of ways in which teaching could be made more efficient, and testing aligned more closely with the requirements of international intelligibility. It helps to clarify the "feature/error" distinction, and implies that accent features which are inconsequential for intelligibility should not be classed as errors.

Finally, it is worth briefly considering the possible dangers of overstating the case for "intelligibility" in relation to language teaching and language testing. The identification of specific intelligibility-related features, as in the LFC research, can be seen as a systemizing enterprise with its roots in structural linguistics. From a more interactionist viewpoint, meaning is negotiated by participants in context-specific ways; in Bamgbose's well known formulation "it is people, not language codes, that understand one another" (1998, p.11). Intelligibility, then, is to some extent interactional between speaker and listener (Smith and Nelson, 1985), and what causes intelligibility problems in one context may not do so in another, making generalization difficult. Of course, it must be acknowledged that linguistic form is related to function and meaning, but also that certain reductions in form are more problematic than others (this is what the LFC research investigated).

If the concept of intelligibility assumes an overly structural or monolithic nature, there is a danger that the development of local norms will be inhibited (Canagarajah, 2006). In the case of Hong Kong, an argument could be made for the use of a local model and norm, rather than one informed by international intelligibility. In other words, such a model would prioritise *intra*national, rather than international, intelligibility, and would to some extent ignore the demands of lingua franca communication. But there are several problems with such a proposal. Firstly, and as mentioned above, Hong Kong's professed identity as an international city would make a "pure" local model unattractive to policymakers. Secondly, there is evidence that phonological problems can disrupt intranational, as well as international, communication (Bansal, 1990, p.229, cited in Jenkins, 2000, p.78). Research needs to avoid generalization at the level of "variety", in order to focus on the language features that affect communication (as in the LFC); language testing needs to deal with the local/global tension and negotiate the turbulent waters between excessive localization and excessive standardization.

DATA COLLECTION AND ANALYSIS

The LPATE test is somewhat unusual in that examiner comments are compiled and placed in the public domain via the website of Hong Kong's Education Bureau (the EDB); they are contained in reports published by the Hong Kong Examinations and Assessment Authority (HKEAA). These reports are edited versions of the examiners' collective comments, and provide general summaries of candidates' strengths and weaknesses, yet they are quite specific in terms of the language "errors" that are noted. In this study, examiner comments relating to pronunciation were collected from the reports from 2003 to 2011 (there were twelve reports in total, as the test was taken twice in some years). As mentioned above, pronunciation is assessed in two parts of the LPATE (Task 1A of the Speaking paper, and the CLA). The reports are divided into sections covering each of the five papers; within the Speaking and CLA sections there are paragraphs or sub-sections covering Pronunciation, Stress and Intonation.

The comments were first analyzed and separated into five categories for the purposes of this study: Vowels, Consonants, Consonant Clusters, Word Stress, and Connected Speech/Intonation. There was little or no difficulty involved in assigning comments to categories, as they usually include examples for clarification. For example, a comment about "teachers who turned a 'goat' into a 'go'" was assigned to the Consonants category, while a nearby comment about "teachers who turned a 'roast chicken' into a 'rose chicken'" (HKEAA, 2004) was assigned to the Consonant Clusters category. Any associated comments about resultant intelligibility or comprehensibility problems were also included in the analysis for the purposes of a subsequent word count. To give some idea of the comments that were included in the analysis, an example from each category is shown in Table 2, below.

TABLE 2

Examples of examiner comments.

Category	Example comment	Date of report
Vowels	Some candidates mispronounced the vowels of some words (e.g. 'won' and 'game') and failed to distinguish between long/short vowels (e.g.'sleep/slip'and 'bean/bin')	2007
Consonants	Confusions sometimes arose when candidates failed to distinguish between sounds such as /v/, /w/ and /r/ (e.g. veal/weal/real); /n/, /l/and /r/ (e.g. neighbour/labour; blanket/bracket)	March 2006
Consonant Clusters	Problems with consonant clusters were also noted as in 'clean' being pronounced as 'keen'	September 2005
Word Stress	One common problem was giving stress to the weak vowel sounds as in 'chocolate', 'carrot' and 'ceremony'	March 2005
Connected Speech and Intonation	At the sentence level, misunderstanding of areas such as stress and intonation were core problems and created communication difficulties. Candidates occasionally stressed the wrong words, or used too much stress and/or inappropriate intonation, or spoke all words with the same degree of stress. This meant that the meaning of the phrase or sentence was often not adequately conveyed	2010

When allocating the comments, positive or neutral comments (for example, "problems with stress… were usually relatively minor"; HKEAA, 2005, p. 14) were ignored, on the grounds that these observations probably did not reduce candidates' scores. The methodological assumption of the study is therefore that negative examiner comments represent "errors" or "unnatural" language in the *Pronunciation, Stress and Intonation* scale shown above in Table 1. Given that a major reason for publishing the comments is to guide prospective candidates, it seems logical that the HKEAA would wish to draw their attention to the language features that influence assessors' decisions. In the scale, reference is made to "errors" in pronunciation. The examiner comments also directly refer to "errors" or "problems", so it is reasonable to assume that these language features reduced candidates' scores. However, the main limitation of the approach is that the actual effects of these comments on candidates' marks are unknown. In particular, while it is fairly straightforward to analyze the types of comments, the degree to which these presumed "errors" affected the assessors' decisions remains uncertain.

Subsequent analysis of the examiner comments had both quantitative and qualitative aspects. In order to assess the relative priorities of the examiners, two measures were employed. Firstly, the number of reports in which the different types of comments appeared was counted. Secondly, word counts of the comments were taken in order to

compare their distribution across the five headings. This may appear to be a curious metric, but it is one that exploits the availability of examiner comments – a total of 2,207 words on the subject of pronunciation alone – and their detailed and systematic nature. The limitation of this method is that the number of words does not necessarily equate with degree of importance or effects on ratings; nevertheless, it appears to offer an approximate, relative indication of examiner priorities.

Figure 1 below shows the number of reports (out of a total of twelve) which contained each category of error. Figure 2 shows the distribution of examiner comments across the five categories, as measured by the number of words in each category.

FIGURE 1

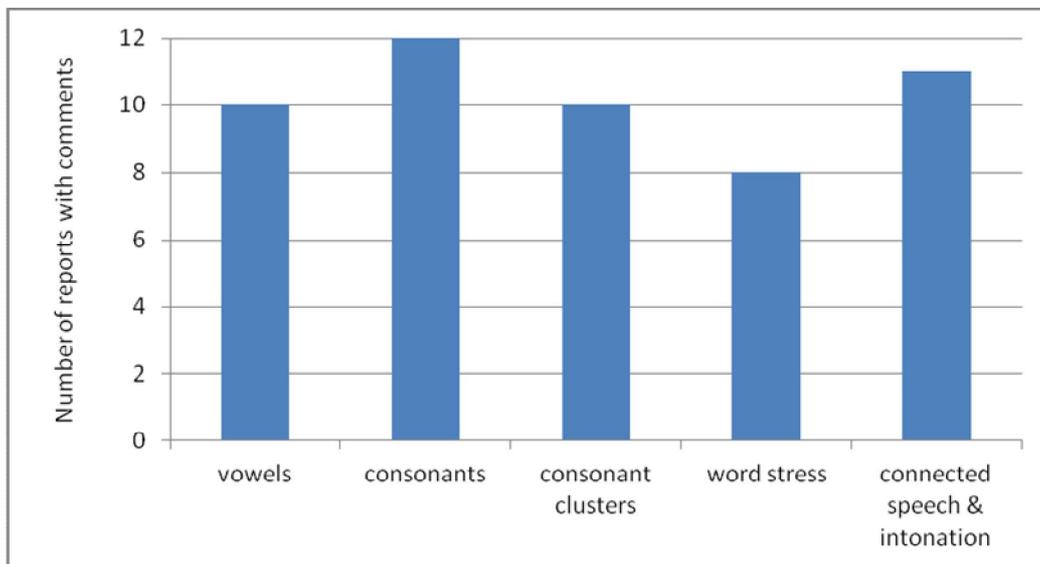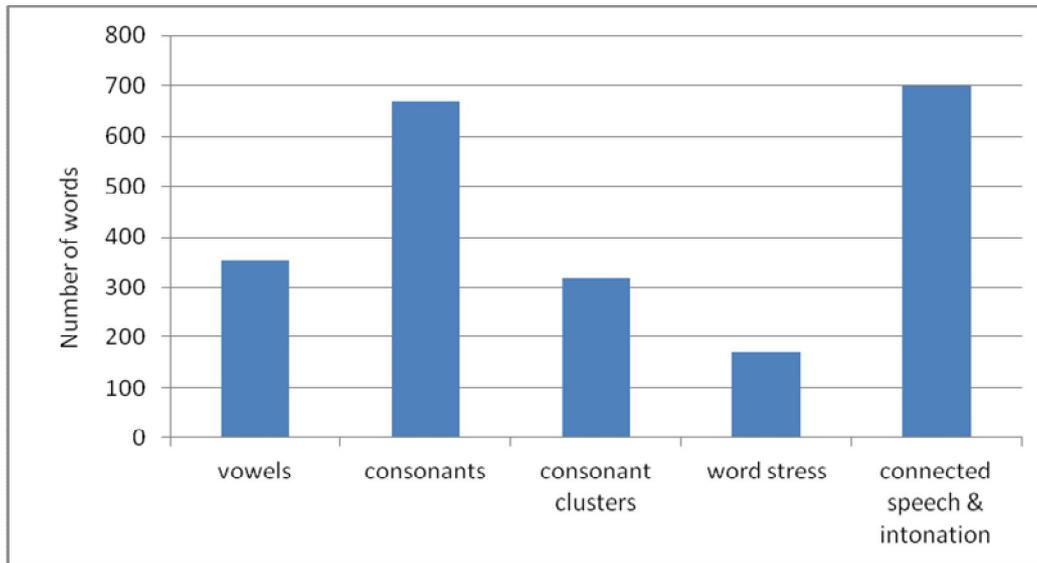The number of examiner reports containing comments in each category.

FIGURE 2

The distribution of examiner comments across categories, measured by the number of words in each category.

In terms of segmental and suprasegmental categories, the first three (segmental) categories therefore accounted for 60% of the comments, with suprasegmentals attracting the remaining 40% (measured in terms of the number of words). The examiners' comments thus seem to strike a balance between the two areas, although the appropriacy of the comments of course depends on their nature. Within the two superordinate categories (segmentals and suprasegmentals), consonants attracted more attention than vowels, and connected speech and intonation attracted more attention than word stress. More words were devoted to connected speech and intonation than any other category, by a small margin, but consonants and consonant clusters were also a major area of focus (if the two categories are combined into a single "Consonants" category, it then becomes the most commented-upon category).

The comments were then compared with the LFC criteria for international intelligibility (Jenkins, 2000, and above). Table 3 below shows the percentage of the comments within each category that agreed or disagreed with the LFC stipulations, or were uncertain in this regard (again, calculated in terms of the number of words).

TABLE 3
Analysis of examiner comments with regard to the LFC criteria.

Category	% of comments that: Agreed with LFC	Disagreed with LFC	Were uncertain	Total
Vowels	80.2	11.9	7.9	100
Consonants	98.9	1.1	0.0	100
Consonant Clusters	96.9	3.1	0.0	100
Word Stress	15.2	73.7	11.1	100
Connected Speech and Intonation	14.0	24.6	61.3	100

The assessment of "agreement" or "disagreement" with the LFC criteria was usually unproblematic, again because of the clear examples given by the examiners. Most cases of disagreement occurred within the Word Stress category; an instance of this can be seen in Table 2 above, where the examiner comment about "giving stress to the weak vowel sounds" clearly refers to a non-core feature of the LFC; the comment was therefore assigned to the "Disagreed with LFC" category in Table 3. Within Connected Speech and Intonation, a case of disagreement was "there seemed to be a general lack of attention to the linking of sounds as in 'think about it'" (HKEAA, 2008). According to the LFC, linking between words is not important for international intelligibility. When analyzing the Vowels category, a degree of interpretation was occasionally necessary. An apparent case of disagreement with the LFC included "the articulation of particular vowel sounds including /æ/ in 'bag', /e/ in 'berry' and 'beg', and /u / in 'cute'" (HKEAA, 2010). The comment was interpreted and analyzed as referring to precise vowel quality (a non-core feature), but without further details it is impossible to be certain. Such cases of ambiguity were rare, however.

Table 3 shows that there was a clear difference between the segmental and suprasegmental categories. Examiner comments in the three segmental categories (vowels, consonants and consonant clusters) aligned closely with the LFC criteria. Those in the suprasegmental categories (word stress, and connected speech and intonation), showed greater divergence from the LFC criteria. The following qualitative analysis of the comments within each category aims to discover the reasons for this. Additional explanation and critical analysis of the criteria will also be provided.

Vowels

The LFC makes a distinction between vowel length or quantity contrasts (important for international intelligibility) and variations in vowel quality (unimportant, except for the /ɜː/ vowel of *nurse*). The explanation provided by Jenkins (2000, p. 144) is that there is great inter-speaker and inter-dialectal variation in vowel quality (see Jenner, 1989), and so there is little need for learners to reproduce vowel qualities exactly. In addition, empirical evidence from the LFC research showed that variations in vowel quality did not lead to intelligibility problems, except in the case of the /ɜː/ vowel. It must be noted that the terms "quantity" and "quality" are problematic in this context, as both play a role in maintaining contrasts that may appear to be based on quantity alone (as noted by Jenkins, 2000, p.140).

Nevertheless, the high rate of agreement between the examiner comments and the LFC occurred because the majority of examiner comments refer to problems with the length or quantity contrast between /ɪ/ and /iː/ (the vowels in *live* and *leave*, for example). According to the examiners, communication was affected "when teachers failed to make a

clear distinction between words like 'slip/sleep', and 'leave/live'" (HKEAA, 2003, p.9). The likely reason for this lack of contrast is the influence of Cantonese, in which language this vowel distinction is allophonic, rather than phonemic (Chan and Li, 2000: 74; see also Flege, 1987). In Hong Kong English, therefore, these and certain other vowel distinctions tend to be perceptually and acoustically indistinguishable (Hung, 2000, p.340). As for the question of why communication was affected when speakers failed to make this distinction, the high functional load of the /ɪ, iː/ contrast may provide an explanation. In a study of the effects of functional load on ratings of accentedness and comprehensibility, Munro and Derwing (2006) concluded that contrasts with a high functional load had relatively large effects on both scales. According to Brown's (1991) ranking, the /ɪ, iː/ contrast has a high functional load, so the examiners' focus on this aspect of the candidates' pronunciation seems to be justifiable.

Most cases of disagreement related to the problem of not differentiating /e/ and /æ/ ("[w]eak candidates had difficulty distinguishing between /e/ and /æ/, leading some to confuse minimal pairs such as 'pen' and 'pan'" (HKEAA, 2008, p.11). According to the LFC, this is a quality difference and therefore unimportant, but no specific mention is made of the widespread tendency towards vowel merger in many L2 varieties of English. However, the LFC does mention the need for "consistency" in vowel production (Jenkins, 2000, p.159), which presumably refers to consecutive uses by the same speaker. The common problem of consistent lack of contrast is not addressed, but some commentaries imply that there are constraints on variation ("[t]he LFC focus on length is not to say that learners are free to use any vowel in any place"; Walker, 2010, p. 35).

The examiners' mention of this problem also seems wholly justified, especially as it is ranked highly for functional load by Brown (1991, p.114). An opposing view on the issue of vowel contrasts is given by Nihalani (2010), who maintains that failing to maintain contrasts is not as problematic as some believe: "[t]heir view [those of textbooks on English pronunciation]…sounds rather ethnocentric and displays intolerance of diversity" (Nihalani, 2010, p. 30). However, Abbott (1991: 233) gives the example of two Ugandan speakers for whom confusion arose between *destruction* and *distraction* because of the absence of contrast between /æ/ and /ʌ/ in their variety of English. Until such time as further empirical data become available, the examiners are probably justified in applying the precautionary principle, especially as the LPATE is an examination for teachers. The LFC guidelines in this area also need to be clarified.

Consonants

There was a very high rate of agreement (nearly 99%) between the examiners' comments and the LFC in the category of consonants. Compared with vowels, the scope for consonantal variation appears to be more constrained; this is also suggested by the relative

historical stability of the English consonantal system (Schreier, 2005, p. 1) and the limited amount of variation in cross-dialectal comparisons (Schwartz, 2005, p.189). Examiner comments in this category refer to substitution ("[c]ommunication was actually impeded when teachers failed to make a clear distinction between words like 'night/light', 'lumber/number', 'wise/vice'"; HKEAA, 2003, p.9) and to the deletion of single final consonants ("'Please dine with me' can easily turn into a request of a very different kind when it is uttered as 'Please die with me'"; HKEAA, 2003, p.9). Substitution was more likely to be mentioned than deletion (28 instances, as opposed to 14).

Again, most of these substitution and deletion patterns are traceable to the influence of the L1 (for example, Cantonese does not permit consonants to appear after diphthongs, explaining the pronunciation of *dine*). Because of their potential to cause intelligibility problems, and their noticeability, it is unsurprising that both the LFC and the examiner comments are intolerant of consonantal errors. A possible exception relates to the dental fricatives /θ/ and /ð/, which are often substituted with [f] and [d] in Hong Kong (and with other sounds elsewhere). According to the LFC this is unproblematic, probably because these contrasts have a relatively low functional load (Brown, 1991). But interestingly, these substitutions were hardly ever mentioned by the examiners (they account for the 1.1% of disagreement shown in Table 3). It is highly likely that the candidates used substitutions; Deterding, Wong and Kirkpatrick (2008, p. 154) found that a majority of Hong Kong speakers substituted [f] for /θ/ in final position. It therefore seems probable that frequency of occurrence was not the only trigger of examiner comments, and that judgments of relative severity or impact on communication are also being made.

In considering the relative severity of errors, as perceived by the examiners, it is interesting to note that the candidates' "errors" were sometimes thought to cause actual confusion. From Table 2, it can be seen that consonantal substitutions are one such case, but lack of vowel contrast and consonant cluster simplifications are also sometimes noted as causing confusion. If this occurred during the CLA, one can only speculate as to whether the "confusion" arose for the students in the classroom, who misinterpreted the meaning of the intended word, or for the examiners, who noticed a difference between the actual pronunciation and the expected pronunciation. If the former is the case, then the problem is indeed significant. If the latter, it could be argued that the examiners are failing to appreciate the reality of local norms. However, and as mentioned above, there are theoretical and empirical grounds for expecting certain substitutions and mergers to be problematic, even between users of the "same" local variety. Even if there was no actual confusion, the examiners may be justified in noting the potential for confusion, again in line with a precautionary approach.

Consonant clusters

According to the LFC, the maintenance of international intelligibility requires that initial consonant clusters should not be simplified, and medial and final clusters should only be simplified "according to L1 rules of elision" (Jenkins, 2000, p. 159). The reasons for this are probably connected with psycholinguistic processing constraints. Simplification in initial clusters is more likely to be problematic than in final clusters, because "[i]nformation lost at the beginning of words impedes word recognition whereas information lost at the end of words often occurs at little cost, word recognition being completed already (Schreier, 2005, p. 220). Final cluster simplification (as when *facts* is pronounced like *fax*) occurs very widely in native-speaker pronunciation, although it is influenced by factors such as the morphemic status of the cluster, the following segment (consonant or vowel), and the relative carefulness of speech (Schreier, 2009, p. 60). One would not generally expect examiners to penalize pronunciations such as /neks mʌnθ/ for *next month*, for example, but this depends on test purpose – examiners may require more careful speech in the language classroom.

The LPATE examiner comments are in fact very closely aligned with the LFC criteria (there was almost 98% agreement). Most of the comments (14 out of 20 examples) refer to the deletion of consonants in initial clusters ("as when 'fruit' ended up as 'foot', 'clouds' as 'cows', and 'phrases' as 'faces'"; HKEAA, 2004, p.13). The focus on deletion in initial clusters is therefore aligned with the LFC research, which specifically prohibits this (Jenkins, 2000, p.143). Regarding final clusters, the examiners note that words ending in -*ed* were often poorly pronounced (HKEAA, 2010, p.10). Again, deletion is mentioned more frequently, with only one case of epenthesis (this also accords with the LFC recommendation that addition is preferable to deletion; Jenkins, 2000, p.143). Although the likelihood of a native speaker deleting /t/ or /d/ in final clusters (and pronouncing *balanced* like *balance*) depends on various factors, and while the intelligibility effects of such an elision may vary, I have classified such examiner comments as being in agreement with the LFC. According to Schreier (2009, p. 60), bimorphemic words (such as *balanced*) are less likely to show final cluster simplification, perhaps because they carry grammatical information. Some of the elisions noted by the examiners (e.g., "[t]eachers who turned a 'roast' chicken into a 'rose' chicken"; HKEAA, 2004, p.13) may well reduce intelligibility and comprehensibility for learners who are unfamiliar with the collocation (or the dish) of *roast chicken*.

Word stress

Turning to suprasegmental features, the LFC suggests that word stress is a non-core feature, one that rarely caused intelligibility problems in the data (Jenkins, 2000, p. 150). This is controversial, and other research (Field, 2005; Tsuzuki and Nakamura, 2009) has drawn different conclusions. Jenkins herself notes that misplaced word stress may affect nuclear stress placement and other features, such as the aspiration of syllable-initial plosives, and concludes that it is "something of a grey area" (Jenkins, 2000, p. 150).

If the examiner comments are compared with the LFC criteria, the agreement rate is only 15% (see Table 2). The main reason for this is that examiner comments often refer to the candidates' pronunciations of multisyllabic words, for example "[w]rong syllable stress as in multi-syllabic words like 'informative' and 'superlative'" (HKEAA, 2009, p.13). It is likely that the candidates pronounced the first word as *inforMAtive*, as in the study of Hung (2005). The comments also mention "giving stress to the weak vowel sounds as in 'chocolate', 'carrot' and 'ceremony'" (HKEAA, 2005, p.14).

Does it matter if word stress is "misplaced" in these cases? With multisyllabic words such as *informative*, there are few grounds for believing that misplaced word stress affects intelligibility; Field's (2005) study focused on disyllabic words. In some other cases, it is possible that NS word stress patterns were maintained and what the examiners noticed was the avoidance of vowel reduction in unstressed syllables, whereby *chocolate* may be pronounced [ˈtʃɒkleɪt] rather than [ˈtʃɒklət]. This is a widely-noted feature of Hong Kong English (Bolton and Kwok, 1990; Deterding, Wong and Kirkpatrick, 2008), and of many other varieties (e.g. Singapore English; Low and Brown, 2005). It seems unlikely to interfere with intelligibility, as long as the stress pattern is maintained, and in fact it may even serve to increase intelligibility because the phonological form more closely resembles the spelling. Field (2005, p. 419) concludes that weak (or reduced) vowel quality "does not contribute importantly to intelligibility". Frequent, widespread occurrence and non-interference with intelligibility suggest that these pronunciations should not be penalized by teachers or examiners: "when local pronunciation features actually enhance international intelligibility, then it would be sensible to retain these features" (Kirkpatrick, Deterding and Wong, 2008, p. 360).

Once again, while some types of word stress "error" are unlikely to be problematic, it would be prudent to wait for further research data before recommending acceptance in examinations. It is possible that pronouncing "weak vowel sounds" as full vowels may affect perceptions of word stress and lead to confusion, but pitch is thought to be the most important cue for word stress (Roach, 2009, p.74). It may be the case, on the other hand, that the perception of nuclear stress (see below) is affected by word and sentence stress. The examiners may have had this interdependence in mind when they drew attention to the importance of "conveying the appropriate meaning through word or sentence stress and intonation" (HKEAA, 2007, p.9).

Connected speech and intonation

A discrepancy between the LFC criteria and the examiner comments can also be found in the case of connected speech and intonation. However, in this case the majority of comments could not be categorized as either agreeing or disagreeing with the LFC, mainly because insufficient detail was given. The LFC's recommendations in these areas are as follows: linking (as may occur in between the words *not at all*) is a non-core feature in

terms of international intelligibility, because it may interfere with the hearer's ability to recognize word boundaries (the LFC prioritizes the listener's ease of perception over the speaker's ease of articulation; Jenkins, 2000, p. 148). Pitch movement (for example, in different kinds of questions) also did not have significant effects on intelligibility in the LFC's empirical research. The available evidence (e.g. Levis, 1999) suggests that the correspondence between question type and pitch movement (yes/no questions rise, wh-questions fall) may not be as clear-cut as many pedagogical materials suggest. Nuclear or tonic stress, on the other hand, is thought to be crucial for intelligibility in the LFC research. Almost all of the intonational errors that caused miscommunication in Jenkins's (2000) study involved nuclear stress, which can be either unmarked (for example, the falling intonation on the last content word in word groups) or contrastive (as when the speaker wishes to draw attention to something; Jenkins, 2000, p. 153).

Turning to the examiner comments in these areas, there were some references to linking (e.g. "there seemed to be a general lack of attention to the linking of sounds as in 'think about it'"; HKEAA, 2008, p.12). As the LFC suggests, there is little reason to see this lack of linking as problematic, but it may depend on the proficiency levels of the participants. The absence of linking may slow down speech rates, and high-proficiency listeners (such as the examiners) may prefer a more rapid delivery. But low-proficiency listeners may well prefer to hear "unlinked" speech, as it makes word boundaries clearer. Low-proficiency speakers, on the other hand, may not be developmentally ready to use connected speech features such as linking. The negotiation of the "speaker-listener equilibrium" (Trudgill, 2005) and its interpretation in assessment is not a simple matter, especially when the complexities of international communication are taken into account.

Patterns of pitch movement that were perceived to be inappropriate also attracted attention from the examiners: "some candidates tended to use the rising tone for all question types, as in 'How do you spell it?'" (HKEAA, 2006, p.11). Although it may sound unfamiliar to native speakers, this pattern is common among L2 English speakers from many language backgrounds, probably because of the influence of the L1 (Jenkins, 2007, p. 124). The LFC designates pitch movement as a non-core feature, as it has little effect on intelligibility, but the main arguments against prioritization in either teaching or testing are mainly pedagogical: the rules governing pitch movement are so complex as to make it virtually unteachable (Walker, 2010, p. 39).

Some of the examiner comments refer to a lack of terminal fall in statements (HKEAA, 2007, pp.10-11), but this is generally seen to be an aspect of nuclear stress, which is a core feature of the LFC. Jenkins (2000, p. 153) defines nuclear stress as highlighting "the most salient part of the message, indicating where the listener should pay particular attention". The terminal fall at the end of statements, a case of unmarked nuclear stress, may help the listener to process the stream of speech. Other examiner comments refer to "intonation flatness" (HKEAA, 2009 p.13), and to non-contrastive stress and intonation patterns: "pausing slightly after individual words rather than after groups of words ('thought

groups'), leading to inappropriate stress and intonation patterns" (HKEAA, 2006, p.9). These may also affect the listener's ability to detect nuclear stress.

Evaluating the importance of connected speech and intonation is difficult because apparently independent areas may actually be interrelated. Schwartz (2005, p. 195) notes that the relationship between nuclear stress and intonation (i.e., pitch movement) is "inevitable",

and also implicates weak forms (e.g. pronouncing *that* with a schwa /ə/, when used as a relative clause marker). He concludes (p. 195) that claims about "unteachability" indicate that "the theoretical underpinnings of these phenomena might still need further development". A considerable degree of caution should therefore be exercised when interpreting research findings about intelligibility, and applying them to the field of assessment. On the other hand, caution is also necessary when deciding what constitutes the "natural" use of stress and intonation (see Table 1). What sounds "natural" to examiners may not be natural to local language users, and may not always equate with communicative effectiveness.

DISCUSSION

The negotiation of the local/global dimension is a challenging task for those involved in language testing. Bamgbose (1998) aptly describes the tension: they are "torn between the norms". There have been claims in some quarters that the focus on native-speaker norms creates bias against non-native speakers of English. Jenkins (2006, p. 49) suggests that as a minimum, examination boards should "refrain from penalizing NNS [non-native speaker] variants which are emerging through their frequent and systematic use as potential forms of future EIL [English as an International Language] forms". Lowenberg (2002) believes that the existence of these emerging norms casts doubt on the validity of tests (such as the TOEIC) that are based on Inner Circle norms. Khan (2009, p.204) focuses on the TOEFL examination and reaches a similar conclusion to that of Jenkins, namely that in order to promote effective pedagogy, assessment practices must be linked to "cultural and contextual realities".

However, Canagarajah (2006, p. 234) alerts us to the possible range of "contextual realities" mentioned by Khan: "[i]n extremely formal institutional contexts where inner-circle norms are conventional (such as in academic communication), one has to adopt the established norms". There are arguments for maintaining a conservative position on the selection of norms, although Canagarajah (2006, p. 236) also notes that the existence of local preferences for inner-circle norms is not in itself a valid argument against developing tests in local English. The arguments about competing norms are localized by Groves and Chan (2010) who believe (with Canagarajah) that the teaching of "standard English" in Hong Kong is necessary to maintain international intelligibility. But Groves and Chan (2010, p. 49) also propose a teaching guideline that gives recognition to some local features of English:

> For those variants that have become features throughout the whole spectrum of Hong Kong speakers, we recommend that both variants should be pointed out in the classroom – the standard English versus the Hong Kong English – in the same way that we today make learners aware of the differences between American English and British English.

This guideline also means that such differences need to be taken into account when tests are designed (Groves and Chan, 2010, p. 49). It seems that there are at least four criteria for the evaluation of feature norms in teaching and testing: systematicity or consistency of use in intra-speaker terms; distribution in inter-speaker terms (features that are used by a majority of speakers have a stronger case for acceptance), teachability (features that require a great deal of classroom time may not be worth prioritizing), and effects on intelligibility (features that do not reduce intelligibility should be accepted, other things being equal). One could add acceptability for local users to this list, but would then risk ignoring the existence of *de facto* local norms; local users are often critical of language features which they themselves employ.

Accordingly, in Hong Kong an aspect of local pronunciation that appears to have a strong claim for acceptability is the non-reduction of full vowels in di- and multi-syllabic words. This was mentioned by the LPATE examiners in several reports, but the use of full instead of "weak" vowels is widespread in Hong Kong English, and also occurs in many other varieties. It seems unlikely to reduce intelligibility, unless the word stress is altered (Field, 2005) and may actually enhance it (see Deterding, 2010). One of the original design goals of the LPATE was to allow for a Cantonese-influenced, yet intelligible accent (Coniam and Falvey, 2002), further supporting the argument for acceptance. Other suprasegmental features such as certain kinds of pitch movement (unlikely to affect intelligibility, and difficult to teach) could also be reviewed for their relevance. However, in the case of suprasegmental features it must be noted that the theoretical issues (such as the interrelationships between areas) are under-researched. There is also a need for research into the actual effects of variable feature use in international communication.

In such research, the attitudes of local stakeholders should not be dismissed as being merely "unenlightened" (as noted by Taylor, 2006, p. 52). While researchers in World Englishes and English as a Lingua Franca tend to advocate the acceptance of local norms, it is by no means certain that local users share their views. In order to increase the accountability of tests with regard to various stakeholders, Elder and Harding (2008, p. 34.3) remind us to consider the views of test-takers themselves. They further observe that while the arguments for contextually-sensitive tests are persuasive, "test users in those contexts are often the first to reject them for a range of reasons", including the lack of mobility potential (Blommaert, 2010) that may result. Once again, the counterargument to this is that tests may themselves play a part in either stultifying or encouraging the development and acceptance of local norms, as argued by Canagarajah (2006).

The study's finding that examiner comments tend to be aligned with the criteria for international intelligibility is interesting in itself. The examiners may possibly be aware of these criteria, but there is more likely a common underlying principle: the features that examiners notice and comment upon are generally those that affect intelligibility, and this is because of their salience. The notion of "salience" is complex (see Kerswill and Williams, 2002), but it essentially equates with noticeability. A candidate's pronunciation of a phoneme may differ from the examiners' representation of that phoneme (in fact it *will* differ, because all speakers have idiolectal variations). However, it is only when a phoneme is deleted, or when its realization is different enough for it to be perceived as a different phoneme that examiners will tend to notice it. Noticeability is also increased when such deletions or substitutions lead to changes in meaning, as when "'night sky' was turned into 'light sky'" (HKEAA, 2005, p. 14).

The comparison between the examiner comments and the intelligibility criteria suggests that what LPATE examiners notice and class as errors are generally the features that have been shown to reduce international intelligibility (according to the research of Jenkins (2000) and a consideration of factors such as functional load). The value of Jenkins's (2000) criteria for international intelligibility here is that they help to distinguish between intelligibility-reducing "errors" and intelligibility-preserving "features", and thus provide a possible guideline for speaking test descriptors (although once again, it must be emphasized that further research to verify, refine and localize the criteria is needed). While most of the LFC criteria appear to be supported by linguistic and psycholinguistic factors, the data from this study suggest that vowel quality is one area of the LFC that needs to be clarified.

CONCLUSION

Hong Kong's LPATE illustrates the difficult balancing act that language testing has to perform. There are linguistic, pedagogical and sociocultural issues that need to be addressed; the test should not show major incompatibility with linguistic theory, pedagogical considerations, or societal attitudes. The recent attention given to local varieties (the World Englishes standpoint) and international intelligibility has made this balancing act slightly harder, by emphasising the inevitability of local language features and their relationship with intelligibility. As Davies *et al.* (2003. p. 582) observe, it is difficult to distinguish between a "stage of learning" and a "stage of transition", where a new local code is being formed. Shifting norms, possibly accompanied by shifting attitudes towards them, must also be taken into account. But language testing is not necessarily restricted to merely responding to such changes. Because of the potential washback effects of testing, there may be ways in which test criteria can influence pedagogy, and thus be a cause (rather than simply an effect) of language attitudes (see Canagarajah, 2006).

The study is limited by its reliance on publicly-available materials, and may have oversimplified the processes of interpretation that lie behind the comments, while possibly overstating their effects on candidates' scores. In general, the LPATE examiner comments

suggest that an appropriate balance is being maintained between the competing norms. The comments were generally compatible with the international intelligibility criteria, and reflect the test's purpose as an assessment of teachers' language proficiency. The relevance of certain suprasegmental features in speaking tests was questioned, but the need for further theoretical investigation and empirical research was noted. Further research into international intelligibility (reflecting the actual international use of English, and not restricted to native-speaker perceptions), the distribution of non-standard features within and across varieties of English, and the relative effects of segmental and suprasegmental features, will all help to inform the evolution of speaking tests in a globalizing world.

REFERENCES

Abbott, G. (1991). A new look at phonological "redundancy". In Brown, A. (Ed.). *Teaching English pronunciation: A book of readings* (pp. 225-234). London and New York: Routledge.

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

Bamgbose, A. (1998). Torn between the norms: innovations in world Englishes. *World Englishes* 17/1, 1-14.

Bansal, R.K. (1990). The pronunciation of English in India. In Ramsaran, S. (Ed.). *Studies in the pronunciation of English: A commemorative volume in honour of A.C. Gimson*. London: Routledge.

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

Bolton, K., and Kwok, H. (1990). The dynamics of the Hong Kong accent: social identity and sociolinguistic description. *Journal of Asian Pacific Communication* 1/1, 147-172.

Brazil, D. (1994). *Pronunciation for advanced learners of English (Teachers' Book)*. Cambridge: Cambridge University Press.

Brown, A. (1991). *Pronunciation models*. Singapore: Singapore University Press.

Brown, A., and Lumley, T. (1998). Linguistic and cultural norms in language testing: A case study. *Melbourne Papers in Language Testing* 7/1, 80-96.

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

Chan, A.Y.W. and Li, D.C.S. (2000). English and Cantonese phonology in contrast:

explaining Cantonese ESL learners' English pronunciation problems. *Language, Culture and Curriculum* 13/1, 67-85.

Coniam, D. (2003). Perceptions of a multimedia syllabus – making the demands of a performance test more accessible. *System* 31, 55-70.

Coniam, D., and Falvey, P. (2002). Selecting models and setting standards for teachers of English in Hong Kong. *Journal of Asian Pacific Communication* 12/1, 13-38.

Dalton, C., and Seidlhofer, B. (1994). *Pronunciation*. Oxford: Oxford University Press.
Davies, A., Hamp-Lyons, L., and Kemp, C. (2003). Whose norms? International proficiency tests in English. *World Englishes* 22/4, 571-584.

Deterding, D. (2010). Norms for pronunciation in Southeast Asia. *World Englishes* 29/3, 364-377.

Deterding, D., and Kirkpatrick, A. (2006). Emerging South-East Asian Englishes and intelligibility. *World Englishes* 25/3-4, 391-409.

Deterding, D., Wong, J., and Kirkpatrick, A. (2008). The pronunciation of Hong Kong English. *English World-wide* 29/2, 148-175.

Dziubalska-Kolaczyk, K. and Przedlacka, J. (Eds). (2005). *English pronunciation models: A changing scene*. Bern, Switzerland: Peter Lang.

Education Bureau (2010). *Language Proficiency Assessment for Teachers (English Language) Handbook (2010).* Hong Kong: Education Bureau of the Hong Kong Special Administrative Region of the People's Republic of China. Available at: http://www.edb.gov.hk/FileManager/EN/Content_1382/language%20proficiency%20assessment%20for%20teachers%20%28english%20language%29%20handbook%202010.pdf Accessed October 1st, 2011.

Elder, C., and Harding, L. (2008). Language testing and English as an international language: Constraints and contributions. *Australian Review of Applied Linguistics* 31/3, 34.1 - 34.11.

Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly* 39/3, 399-423.

Ferguson, G. (2006). *Language Planning and Education*. Edinburgh: Edinburgh University Press.

Flege, J. (1987). The production of 'new' and 'similar' phones in a foreign language: evidence of the effect of equivalence of classification. *Journal of Phonetics* 15, 47-65.

Groves, J., and Chan, H.T. (2010). Lexical traps in Hong Kong English. *English Today* 26/4, 44-50.

Hahn. L.D. (2004). Primary stress and intelligibility: research to motivate the teaching of suprasegmentals. *TESOL Quarterly* 38/2, 201-223.

Hall, S. (1997). The local and the global: Globalization and ethnicity. In King, A. (Ed.) *Culture, Globalisation and the World System* (pp. 41-68). Minneapolis: University of Minnesota Press.

Harding, L. (2011). *Accent and listening assessment*. Frankfurt: Peter Lang.

HKEAA (Hong Kong Education and Assessment Authority) (2003). *Language Proficiency Assessment for Teachers (English Language) 2003 (March)*. Available at:
http://www.edb.gov.hk/index.aspx?langno=1&nodeid=1365#
Accessed October 1st, 2011.

HKEAA (Hong Kong Education and Assessment Authority) (2004). *Language Proficiency Assessment for Teachers (English Language) 2004 (September)*. Available at:
http://www.edb.gov.hk/index.aspx?langno=1&nodeid=1365#
Accessed October 1st, 2011.

HKEAA (Hong Kong Education and Assessment Authority) (2005). *Language Proficiency Assessment for Teachers (English Language) 2005 (March)*. Available at:
http://www.edb.gov.hk/index.aspx?langno=1&nodeid=1365#
Accessed October 1st, 2011.

HKEAA (Hong Kong Education and Assessment Authority) (2006). *Language Proficiency Assessment for Teachers (English Language) 2006 (March)*. Available at:
http://www.edb.gov.hk/index.aspx?langno=1&nodeid=1365#
Accessed October 1st, 2011.

HKEAA (Hong Kong Education and Assessment Authority) (2007). *Language Proficiency Assessment for Teachers (English Language) 2007*. Available at:
http://www.edb.gov.hk/index.aspx?langno=1&nodeid=1365#
Accessed October 1st, 2011.

HKEAA (Hong Kong Education and Assessment Authority) (2008). *Language Proficiency Assessment for Teachers (English Language) 2008*. Available at:
http://www.edb.gov.hk/index.aspx?langno=1&nodeid=1365#

Accessed October 1st, 2011.

HKEAA (Hong Kong Education and Assessment Authority) (2009). *Language Proficiency Assessment for Teachers (English Language) 2009*. Available at: http://www.edb.gov.hk/index.aspx?langno=1&nodeid=1365#
Accessed October 1st, 2011.

HKEAA (Hong Kong Education and Assessment Authority) (2010). *Language Proficiency Assessment for Teachers (English Language) 2010*. Available at: http://www.edb.gov.hk/index.aspx?langno=1&nodeid=1365#
Accessed October 1st, 2011.

Hung, T. (2000). Towards a phonology of Hong Kong English. *World Englishes* 19, 337-356.

Hung, T. (2005). Word stress in Hong Kong English: A preliminary study. *Hong Kong Baptist University Papers in Applied Linguistics* 9, 29-49.

Isaacs, T. (2008). Towards defining a valid assessment criterion of pronunciation proficiency in non-native English-speaking graduate students. *The Canadian Modern Language Review* 64/ 4, 555–580.

Jenkins, J. (2000). *The phonology of English as an International Language*. Oxford: Oxford University Press.

Jenkins, J. (2006). The spread of EIL: a testing time for testers. *ELT Journal* 60/1, 42-50.

Jenkins, J. (2007). *English as a Lingua Franca: Attitude and identity*. Oxford: Oxford University Press.

Jenner, B. (1989). Teaching pronunciation: the common core. *Speak Out!* 4, 2-4.

Joint Committee on Testing Practices (1988). The code of fair testing practices in education. Washington, DC: American Psychological Association.

Kachru, B. (1977). Linguistic schizophrenia and language census: a note on the Indian situation. *Linguistics* 15/186, 17–32.

Kachru, B. (1992). Models for non-native Englishes. In Kachru, B. (Ed.). *The Other Tongue: English Across Cultures* (pp. 48-74). Urbana, Illinois: University of Illinois Press.

Kerswill, P., and Williams, A. (2002). 'Salience' as an explanatory factor in language change. In Jones, M. and Esch, E. (Eds.). *Language change: The interplay of internal, external and extra-linguistic factors* (pp. 81-110). Berlin: Mouton de Gruyter.

Khan, S.Z. (2009). Imperialism of international tests: an EIL perspective. In Sharifian, F. (Ed.). *English as an International Language: Perspectives and Pedagogical Issues* (pp. 190-205). Clevedon, UK: Multilingual Matters.

Kirkpatrick, A. (2006). Which model of English: native speaker, nativized or lingua franca? In Rubdy, R. and Saraceni, M. (Eds.). *English in the World: Global rules, global roles* (pp. 71-83). London and New York: Continuum.

Kirkpatrick, A., Deterding, D., and Wong, J. (2008). The international intelligibility of Hong Kong English. *World Englishes* 27/3-4, 359-377.

Levis, J. (1999). The intonation and meaning of normal yes/no questions. *World Englishes* 18/3, 373-380.

Li, D.C.S. (2009). Researching non-native speakers' views toward intelligibility and identity. In Sharifian, F. (Ed.). *English as an International Language: Perspectives and pedagogical issues* (pp. 81-118). Clevedon, UK: Multilingual Matters.

Low, E.L. and Brown, A. (2005). *English in Singapore: An introduction*. Singapore: McGraw-Hill.

Lowenberg, P. (2002). *Southeast Asian norms: implications for assessing proficiency in English as a global language*. Paper presented at the AILA Conference, Singapore, December.

Luk, J., and Lin, A. (2006). Uncovering the sociopolitical situatedness of accents in the World Englishes paradigm. In Hughes, R. (Ed.). *Spoken English, Applied Linguistics and TESOL: Challenges for theory and practice* (pp. 3-22). Basingstoke: Palgrave Macmillan.

Mahboob, A., and Szenes, E. (2010). Construing meaning in World Englishes. In Kirkpatrick, A. (Ed.). *The Routledge Handbook of World Englishes* (pp. 580-598). Abingdon, UK: Routledge.

Morley, J. (1991). The pronunciation component in teaching English to speakers of other languages. TESOL Quarterly, 25/3, 481-520.

Morris, P. (2004). Teaching in Hong Kong: Professionalization, accountability and the state. *Research Papers in Education* 19/1, 105-121.

Munro, M. and Derwing, T. (1999). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning* 49, 285–310.

Munro, M. and Derwing, T. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. *System* 34/4, 520-531.

Nihalani, P. (2010). Globalization and international intelligibility. In Saxena, M., and Omoniyi, T. (Eds.) *Contending with globalization in World Englishes* (pp. 41-62). Bristol: Multilingual Matters.

Pennycook, A. (2007). *Global Englishes and transcultural flows*. London: Routledge.

Quirk, R. (1990). Language varieties and standard language. *English Today* 21, 3-10.

Raschika, C., Butterfint, Z. and Song, L. (2008). *Intonation and prosody as markers of intercultural identity*. Paper presented at the 41st BAAL Symposium, University of Swansea.

Roach, P. (2009). *English phonetics and phonology* (4th ed.) Cambridge: Cambridge University Press.

Schneider, E. (2003). The dynamics of new Englishes: from identity construction to dialect birth. *Language* 79/2, 233-81.

Schreier, D. (2005). *Consonant change in English worldwide*. Basingstoke: Palgrave Macmillan.

Schreier, D. (2009). How diagnostic are English universals? In Filppula, M., Klemola, J. and Paulasto, H. (Eds) *Vernacular universals and language contacts: Evidence from varieties of English and beyond* (pp. 57-79) New York and London: Routledge.

Schwartz, G. (2005). The Lingua Franca Core and the phonetics-phonology interface. In Dziubalska-Kolaczyk, K. and Przedlacka, J. (Eds). *English Pronunciation Models: A changing scene* (pp. 177-198). Bern, Switzerland: Peter Lang.

Setter, J., Wong, C.S.P., and Chan, B.H.S. (2010). *Hong Kong English*. Edinburgh: Edinburgh University Press.

Smith, L., and Nelson, C. (1985). International intelligibility of English: Directions and resources. *World Englishes* 3, 333-342.

Smith, L. (1992). Spread of English and issues of intelligibility. In Kachru, B. (Ed.) *The other tongue: English across cultures* (pp. 75-90). Urbana, Illinois: University of Illinois Press.

Taylor, L. (2006). The changing landscape of English: Implications for language assessment.

ELT Journal 60/1, 51-60.

Trofimovich, P. and Isaacs, T. (2012). Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition* 15/4, 905-916.

Trudgill, P. (2005). Finding the speaker-listener equilibrium: segmental phonological models in EFL. In Dziubalska-Kolaczyk, K. and Przedlacka, J. (Eds.) *English Pronunciation Models: A Changing Scene* (pp. 213-228) Bern, Switzerland: Peter Lang.

Tsui, A.M., and Bunton, D. (2002). The discourse and attitudes of English language teachers in Hong Kong. In Bolton, K. (Ed.) *Hong Kong English: Autonomy and creativity* (pp. 57-78). Hong Kong: Hong Kong University Press.

Tsuzuki, M., and Nakamura, S. (2009). Intelligibility assessment of Japanese accents. In Hoffmann, T., and Siebers, L. (Eds.). *World Englishes: Problems, properties and prospects.* Amsterdam and Philadelphia: John Benjamins.

Walker, R. (2010). *Teaching the pronunciation of English as a Lingua Franca.* Oxford: Oxford University Press.