FORUM

Whose norms? International proficiency tests in English

ALAN DAVIES,* LIZ HAMP-LYONS** and CHARLOTTE KEMP***

EDITORS’ NOTE: This forum paper is published to encourage scholarly debate and discussion. Comments on this paper by a panel of scholars, along with responses by the authors, will appear in a later issue of World Englishes.

ABSTRACT: The paper reports on research into the claim that widely used English language proficiency tests are unfair (unfair in the sense that a test favouring boys could be said to be unfair to girls). The global spread of English opens a major debate on the acceptability of competing norms, whether they should be exonomative or endonomative (Davies, 1999; Gill, 1999). An International English (IE) view insists that the only acceptable norms are those of native English speakers (NES). A strong World Englishes (WES) view maintains that to impose IE on users of WEs may be discriminatory against non-native English speakers (NNES). The theoretical and practical aspects of the issue come together in institutionalized English language proficiency testing. The paper reports on comparative definitions by expert local judges of the norms used in international (IELTS and TOEFL) and national (Hong Kong, Singapore, Malaysia, India, China) English language proficiency tests and considers in what ways the differing definitions have in practice influenced language test construction. It further reports on the findings so far with regard to three basic questions:
1. How possible is it to distinguish between an error and a token of a new type?
2. If we could establish bias, how much would it really matter?
3. Does an international English test privilege those with a metropolitan anglophone education?

INTRODUCTION

We began with the hypothesis that international English tests are biased: by that we mean that they systematically misrepresent the ‘true scores’ of candidates by requiring facility in a variety of English to which whole groups of candidates have not been exposed. Bias, therefore, is not about difference as such but about unfair difference. The argument about bias in international English tests is that these tests represent the old colonial Standard English of the UK, USA, etc., a kind of English that is not known, or only partly known, to many of those who have learnt English as an additional language, in particular those living in one of the so-called New English societies which have adopted a local (or locally emerging) variety of English – societies such as Singapore, Malaysia, India. We set out in this research to investigate this claim of bias. We realized, as will become clear during the paper, that there is much speculation and very little evidence that would support or refute the bias claim. Wishing to produce some hard data, we planned a project which would

* Department of Theoretical and Applied Linguistics, University of Edinburgh, 40 George Square, Edinburgh EH8 9LL, Scotland, UK. E-mail: a.davies@ed.ac.uk
** Language Testing Research Centre, University of Melbourne, VIC 3010, Australia. E-mail: lizhl@unimelb.edu.au
*** Institute of Education, University of Stirling, Stirling, FK9 4LA, Scotland, UK. E-mail: charlotte.kemp@stir.ac.uk

© Blackwell Publishing Ltd. 2003, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.
compare performance on international and local tests of English, our argument being that a lack of fit of comparative performance on these tests would give some credence in support of bias. We recognized that without some neutral measure to determine true ‘true scores’ we would never be in a position to assert or reject bias, since any differences we found could be attributable to real differences rather than to bias.

**INTERNATIONAL ENGLISH AND WORLD ENGLISHES**

While there is general agreement on the facts of the continuing contemporary spread of English (Fishman, Conrad and Rubal-Lopez, 1996; McArthur, 1998), there are polarized attitudes to this spread, which we shall indicate by the terms International English (IE) and World Englishes (WEs). These attitudes are often quite nuanced: here, to point up the discussion, we present them as stereotypes.

IE stands for the universalist view, there is one English which unites all those who use English. WEs, on the other hand, stands for the view that English has split post-colonially into a plurality of -lects: the English of Singapore, the English of India, the English of Malaysia, the English of Nigeria and so on. Each polarity has its supporters and its opponents. Thus IE is reckoned by its supporters to be enabling (Quirk, 1985, 1990), while its opponents bitterly object to its hegemonizing grip on the modern westward leaning world (de Beaufregarde, 1999).

The WEs view (Kachru, 1986, 1992) is that without attention to the local norms, all institutional uses of English are necessarily biased in favour of those for whom the metropolitan forms are native, on the grounds that IE does in practice equal Standard British English, Standard American English, etc. This applies equally to postcolonial WEs (Pakir, 1993) and to EFL learners and users whose primary need for English is in-country (for example in China) not international.

For holders of both views, IE and WEs, what is at issue is not the existence of variation but the role and status of language norms (Bartsch 1988; Davies, 1999). The IE view is strengthened both by the strict view of norm acquisition, viz. that it needs a large enough body of native speakers to take on its responsibility (Davies, 2003; but see Gradrol, 1999), and by the need many EFL learners have for a test to provide international recognition of their English proficiency (e.g., for certification, university entry, employment, immigration etc.).

The argument about the institutional role of English comes to a head in concerns about the role of IE in English language testing, especially proficiency testing in high stakes contexts, for example TOEFL, TOEIC, IELTS (Criper and Davies, 1988; Spolsky, 1993; Clapham, 1996). What is at stake here is whose norms are to be imposed. Bhatt (1995) upbraids Quirk for discrediting the use of non-native varieties of English as pedagogically acceptable models since these varieties are not adequately described. While agreeing that that is the case, Bhatt continues: ‘Quirk argues that in non-native contexts only the “Standard” (the “native” model) must be used in the teaching of English and further that non-native teachers must be in constant touch with the native language. The implications of this argument’, he adds, ‘are quite unfortunate and backward’ (Bhatt, 1995: 247).

To date, there is a remarkable dearth of empirical evidence to substantiate cries of language test bias (Coppieters, 1987; Birdsong, 1992). ETS have over the years produced research reports on the conduct of TOEFL (e.g., Clark, 1977) which have shown that
Whose norms? International proficiency tests in English

performance across national and linguistic groups is varied, systematically so. But that in itself does not support the bias case since it might well be that the group differences that are found are reflections of these groups’ true scores, just as a tape measure which shows that men are on the whole taller than women is not biased in favour of men. Much of the argument about bias takes the form of polemic rather than discussion of data.

Nelson (1995) condemns ‘the monocentric, probably ethnocentric view’ that a particular form of English is ‘correct’ and ‘right’ and that other forms are then by definition ‘wrong’. And Davidson (1994) comments on ‘the prevalent imperialism of major international tests of English’ and says ‘Several large English tests hold sway world-wide; tests which are clear agents of the English variety of the nation where they are produced. These tests maintain their agency through the statistical epistemology of norm-referenced measurement of language proficiency, a very difficult beast to assail’ (1994: 119–20).

Hill and Parry (1994) complain about the conservatism of those responsible for English language examinations in the face of these new WEs challenges. They remark that ‘one question . . . that [educators] continually face is the degree to which non-native learners in a particular country should be tested on local as opposed to metropolitan varieties of English’ (1994: 2). Lowenberg (1993) makes a similar point about the conservatism of the language testing profession:

in language testing, an implicit (and frequently explicit) assumption has long been that the criteria for measuring proficiency in English round the world should be candidates’ use of particular features of English which are used and accepted as norms by highly educated native speakers of English. (Lowenberg, 1993: 95)

Those who question that assumption would argue that the days when the only people who took English proficiency tests were members of the educated elite in their home country/language are past. They point out that if English proficiency tests are not to be localized, they must instead be based on a demonstrably common language core – language that can be shown to be shared by all the native varieties of English, and by curriculum/syllabus documents worldwide. However, Lowenberg’s analysis of the TOEIC test leads him to the following conclusion:

the brief analysis presented in this paper is sufficient to call into question the validity of certain features of English posited as being globally normative in tests of English as an international language, such as TOEIC, and even more, in the preparation materials that have developed around these tests. Granted, only a relatively small proportion of the questions on the actual tests deal with these nativized features; most test items reflect the ‘common core’ of norms which comprise Standard English in all non-native and native-speaker varieties. . . . But given the importance usually attributed to numerical scores in the assessment of language proficiency, only two or three items of questionable validity on a test form could jeopardize the ranking of candidates in a competitive test administration. (1993: 104)

Lowenberg challenges ‘the assumption held by many who design such English proficiency tests . . . that native speakers still should determine the norms for Standard English around the world’ (p. 104). More recently, he has followed up his earlier work with an analysis of newspaper style sheets, government documents and ESL textbooks in Malaysia, Singapore, Brunei and the Philippines, and found that these diverge from native speaker varieties at all levels from the morphosyntactic and lexical to pragmatic and discoursal conventions (Lowenberg, 2002).

One way to avoid using the global norms to which Lowenberg objects is to investigate to

© Blackwell Publishing Ltd. 2003
what extent local norms are appropriate both locally and beyond the local, and use this information in test development. Such an investigation is reported by Hill (1996) and Brown and Lumley (1998), both referring to the development of an English Proficiency Test for Indonesian teachers of English. Hill comments:

the majority of Indonesian learners will use English to communicate with other non-native speakers within South-East Asia. For this reason it was decided the test should emphasize the ability to communicate effectively in English as it is used in the region, rather than relate proficiency to the norms of America, Britain or Australia . . . this approach also aims to recognize the Indonesian variety of English both as an appropriate model to be provided by teachers and as a valid target for learners. (Hill, 1996: 323)

Brown and Lumley claim that in the Indonesian test development they had several aims in view. These aims, they maintain, were all fulfilled. They were:

- the judicious selection of tasks relevant to teachers of English in Indonesia,
- the selection of culturally appropriate content,
- an emphasis on assessing test takers in relation to local norms,
- the use of local raters, that is non-native speakers of English (whose proficiency was nevertheless of a high standard). (Brown and Lumley, 1998: 94)

Kenkel and Tucker (1989) mounted one of the few research studies in this field. They noted that ‘international students . . . have spent much of their lives acquiring and using their regional variety of English. These students bear some similarities . . . to speakers of Black English in the U.S.’ (1989: 202). And they concluded from their analysis of written essays by Nigerian and Sri Lankan students that ‘errors’ in their work should more accurately be called deviations from the native speaker norm. Recent empirical work that has attempted to marry the two approaches, the international and the local, has been carried out by Hamp-Lyons and Zhang. They maintain that:

The question of which English(es) should be privileged on tests is particularly problematic and interesting in academic contexts where traditionally ‘standard’ forms of English are the only ones accepted. (Hamp-Lyons and Zhang, 2001).

Taylor (2002) notes the importance of distinguishing between reception and production. She quotes from the UCLES/FCE Handbook:

candidates’ responses to tasks in the Cambridge EFL examinations are acceptable in varieties of English which would enable candidates to function in the widest range of international contexts. (UCLES, 2001: 6)

She also comments that:

candidates are expected to use a particular variety with some degree of consistency in areas such as spelling. (UCLES, 2001: 6)

In other words, a ‘Variety of English’ needs to be consistent within itself in order to be accorded the status of a separate and distinct world English. This policy has profound implications not only for learners/test-takers, who should be fully apprised of the range of acceptable language, which may be a surprise to them and their teachers, schooled to a view that ‘correct English’ means British English, American English, etc. It also has implications for raters of oral (and written) tests and for their rater training. Raters who are not native users of the variety being used by test-takers may find rating difficult. These
imlications are extensive and of potentially great impact, but to date they have not been reported on by UCLES.

ATTITUDES TO ENGLISHES

Positions on the WEs vs. IE question are social-affective rather than linguistic. For example, in Singapore local attitudes towards Singlish, the variety of local English with the lowest social status (the baselect) are condemnatory. Take these comments on Singlish made in radio broadcasts in Singapore by the former and the current Prime Minister:

Singlish is a handicap we must not wish on Singaporeans. (Lee Kuan Yew, 14 August 1999)

If we carry on using Singlish, the logical final outcome is that we . . . will develop our own type of pidgin English, spoken only by 3 million Singaporeans, which the rest of the world will find quaint but incomprehensible. (Goh Chok Tong, 22 August 1999)

At the same time there is fierce loyalty towards the local Singapore variety. Tommy Koh, representative of Singapore in Washington, made his feelings clear: ‘I should hope that when I’m speaking abroad, my countrymen will have no problem recognizing that I am a Singaporean’ (Koh, 1979).

Lukmani (2002) argues that many Indian speakers of English produce an interlanguage which is not systematic, either in grammar or discourse, except for a few examples such as ‘putting up a play’, ‘coping up with a situation’. Politically, she continues, no-one in India accepts the existence of Indian English as an acceptable written variety and there are no models on which to base it. In her view there should not be a disparity between national and international standards for testing: Standard English is the expression of the culture of middle-class India and the large number of people who speak a mesolect can use their other languages as well as English. The situation in India as Lukmani describes it may be comparable to that of Black English Vernacular (BEV) in the USA 20 years ago, when some academics claimed that it could be used for all functions but most educated Blacks were totally against it. To Lukmani the issue of standardization is very important, for without it there are no correct answers. Even so, Lukmani concedes, tests and examinations often feature Indian English because some testers, especially at school level, are not aware of their own lack of knowledge of Standard English. One might reply that if speakers are not aware that the English they speak is not ‘Standard English’, it must form a standard of its own— that is, a variety. And if this ‘Indian English’ appears on tests and exams in India, the test developers presumably believe they are using some kind of standard form in their test questions. We seem to have a conundrum here, as so often in this complex research area.

THE SEMINAR AND JUDGEMENT STUDY

In pursuing the judgemental aspects of our research, we organized a seminar, bringing together a small group of experts, each representing one country where bias might be present. Our seminar consisted of representatives from China, Singapore, Malaysia and India, each of whom presented local English test material. Together we carried out a content analysis (Criper and Davies, 1988) of a small number of these local tests as well as of two major international English tests, TOEFL and IELTS. Those participating in the

© Blackwell Publishing Ltd. 2003
seminar were: Yasmeen Lukmani (India); Bonnie Zhang (China); Jay Banerjee (Singapore); Nesamalar Chitavelu (Malaysia); Charlotte Kemp (UK), our research assistant; Tom Lumley (Hong Kong and Australia); Liz Hamp-Lyons (Hong Kong and USA); Alan Davies (Hong Kong and UK). All these have had input to this paper, but only the three named authors are responsible for any defects.

Before the seminar we asked each country representative to prepare a contextual briefing on the use of English in the country they were representing, and to provide some test papers from national or regional tests within that country. The materials for the seminar included the resulting briefing documents and test papers; a certain amount of analysis/critique of each test; plus more detailed analyses of the TOEFL and IELTS carried out by Charlotte Kemp.

**English language testing in Malaysia**

The situation of English in Malaysia, described by Chitavelu and investigated in depth by Kemp for the Seminar, is an interesting one. English was extensively used until the end of the 1970s, when Malayization policies led to the planned dominance of Bahasa Malaysia in education. In recent years it has gradually been acknowledged at government levels that this policy has led to a significant decline in the standard of English in Malaysia, and policy has been changed to bring English back into advanced educational contexts. Nevertheless, in modern Malaysia English does not have very deep penetration. Our group found that the Malaysian school exams use markedly Malaysian English, reflecting either on the general competence of school leavers in Malaysia or on the robustness of the national view of the local variety, or possibly both. In test papers we noted patterns of ‘error’ (from the Standard English point of view) and examples of local expressions such as:

- Across the South China sea, two exciting destinations await your discovery. Both Kuching and Kota Kinabalu have a lot to offer visitors. They are the cultural havens where food, dance, music and picturesque scenery can truly be enjoyed and appreciated. (patterned article misuse)

- Nature lovers planning to participate in the Bukit Fraser 2001 International Bird Race on Saturday can still register an hour before flag-off. (local lexis)

- A mere mention of Hong Kong conjures up images of a cosmopolitan city, neon lights, a shopper’s paradise, and all that is associated with life on the fast lane. (wrong preposition in fixed phrase)

Whether it has been realized by the test writers or not, these tests assume local knowledge.

The most interesting feature of the Malaysian school-level tests of English (SULIT) is that the test writers mix international Standard English and local varieties. This manifests itself in two ways:

1. in reported speech, for example:
   
   ‘The government cannot shy away from its responsibilities and expect the universities to fend on their own so soon.’ (The first part of the sentence is accurate if slightly archaic formal British English; however in the latter part of the sentence British Standard English would expect: ‘to fend for themselves’; and the phrase is a little idiomatic compared to the rest of the sentence.)

2. in reading comprehension tests, where the texts (taken from authentic British English material) are normally in British Standard English (BSE) while the questions (presumably written by local examiners) may be in the local variety. For example:
Whose norms? International proficiency tests in English

- ‘to draw the reader’s attention, the writer introduces his subject by’:
  (usually you ‘draw someone’s attention to something’)
- ‘mentioning Hong Kong conjures images of a cosmopolitan city’
  (‘conjures up’ would be expected in BSE)
- ‘pointing out that the surprising fact that more than forty per cent of Hong Kong is still rural’ (the first ‘that’ is superfluous).

In the Malaysian university level tests (MUET) the examples of English used might equally well be found in a UK test. We may surmise, then, that at school level (SULIT) there is an acceptance of aspects of Malaysian English in tests, but not at university level.

**English language testing in Singapore**

There seems no doubt that Singapore Colloquial English is a native variety of English. That is not in question. What is in question is whether that variety is now accepted as a standard model for use in education and formal domains in Singapore. The evidence would suggest that it is not. That Singlish is not accepted officially in tests of the written language is clear. What remains unclear is how far Singlish is accepted from candidates in tests of spoken English.

Singaporean examinations are resolutely exonomative in the norms they observe. Teachers seem to find the co-existence of the formal (written) and the informal (spoken) a compromise they can live with. This is well expressed by Teng Hui Huang:

the British standard form and the local variety (or Singlish) should not be treated as being mutually exclusive since they address two different domains, a formal and an informal one . . . As a teacher, of course, I will continue to recommend the standard British norm in formal instruction but that, however, does not mean that I will frown on the use of Singlish in other informal settings. (Teng, 1985: 3)

Politicians, however, are less ready for compromise, hence their condemnation of spoken usages such as the following, taken from speeches delivered by politicians launching the Speak Good English Campaign:

‘Thirteen boys are facing disciplinary action for insisting on wearing shaven heads.’ (NNS collocation)

‘Local types of English often sprout up in places where non-English speakers come into contact with English speakers.’ (Singlish phrasal verb).

‘Investors will hesitate to come over if their managers or supervisors can only guess what our workers are saying.’ (Singlish phrasal verb)

**English language testing in China**

As Bonnie Zhang explained in the Seminar, there has only been a small amount of discussion of the forms of English used by Chinese learners of English, known variously as Chinglish, Chinese English and China English, and their implications for English teaching and learning in China (Li, 1993; Jiang, 1995; Lin, 2002). Hamp-Lyons and Zhang (2001) investigated the reactions of two rater groups, English native speakers and Chinese native speakers, to the culture-bound rhetorical patterns in Chinese EFL learners’ examination essays and discuss the WEIs issues in writing assessment.

British and American standards are often mixed together without distinction: for
example, people may be familiar with ‘taxi’ (BrE) but not ‘cab’ (AmE) because of greater exposure to British English texts/textbooks and media; yet many students and even some teachers of English know how to write a ‘resumé’ (AmE) but not a ‘CV’ (BrE) because most of the people in China who apply for further studies in other countries aim at universities in the US. Although non-standard varieties of English or even incorrect/inappropriate use of English can sometimes be found in a test paper, Zhang felt that this mainly results from test developers’ ignorance or unawareness of the errors or the non-standardness of a form.

Zhang’s examination of NMET and CET test papers indicated that in terms of content or selection of materials for reading comprehension tests, a mixture of AmE, BrE, and China/Chinese English can be found in a single test paper depending on where the source text comes from. According to the instructions given to test developers there is only a very general requirement about text selection, i.e., texts can be selected from any English-language newspapers, magazines, the web, etc. regardless of the place of publication and it is sometimes recommended that at least one of the texts needs to relate to life or culture in China. Therefore, different item writers may have their own judgements of what Standard English is and the texts selected may vary significantly in one test paper. The issue that has emerged here is: is there only one single (consistent) standard/norm, i.e., the IE standard, applied in test development as is claimed/required/assumed by the policy-makers in China? This issue is also reflected in assessing speaking and writing, particularly writing. An examination of the topics used in the CET writing test for the past few years shows a lot of similarity with the topics in the TOEFL Test of Written English (mostly argumentative essays). Although topics are set by local test developers/designers and scoring is also done locally by Chinese native speaker teachers of English, an examination of the scoring guide/criteria used in the CET writing test in the TOEFL TWE presents further similarity in descriptions of performance levels. This is exactly what the policy-makers or the developers are aiming at: using AmE/BrE standards in assessing students’ performance. But are the local raters really following the same AmE/BrE standards in assessing students’ scripts as it is claimed/assumed they are doing? An investigation by Hamp-Lyons and Zhang (2001) suggested that while the local raters think they are applying native speaker norms, they tend to give more favourable comments on scripts with salient China English features, particularly discourse features, as they share the same background with the testees, than the English native speaker raters. In contrast, English native speaker raters may give different, sometimes contradictory, comments on the same scripts even though they are applying/referring to the same scoring guide.

We can tentatively conclude that English language test development practice in China is IE (AmE/BrE) norm dependent, but localized in many ways, e.g., selection of contents/texts, scoring, rater training, etc.

**English language testing in India**

During the Seminar, Lukmani maintained that ‘good’ English in India varies according to class and type of work situation. She proposed a model of three levels of English users:

1. 1 to 2 million with native-like English;
2. very large numbers with advanced fossilized English: fossilization will have occurred because of dominant feedback of the same fossilized error-ridden forms within/among this group;

© Blackwell Publishing Ltd. 2003
3. a small number of users of semi-pidgin who have difficulty with both reception and production.

Lukmani reported that teachers of English in India aim for Standard English, although it is not clear that teachers everywhere and at all educational levels provide a Standard English model for their students (some, undoubtedly, are in the middle group in her model). The high value placed on literature in India has resulted in language forms that tend to be formal and archaic, and these appear both in examination papers and in the students’ answers. Lukmani’s own position is strongly in favour of teaching and testing the standard model (by which she means the UK model), reflecting Lowenberg’s (1993) argument that ‘the devaluation of home-grown forms is strengthened by the effects of both local and international examining boards.’ However, Lukmani reported on a survey which showed that 50% of Indians want to aim for Standard (British) English while the other 50% accept the currency of Indian English, perhaps hardly surprising when only 1–2% can recognize Standard English.

To identify a consistent, separate Indian English we would need evidence of significant variation in phonology, lexis, syntax, genre and style (see Dasgupta, 1993). Yamuna Kachru (1988) describes some differences between written Hindi and written English. She points out that written Hindi was used solely for literary purposes and that prose genres did not appear in India until the 1900s with the advent of English-speaking missionaries and the artefacts of the military and industrial world they; and the colonizers brought with them. This history has produced a marked stylistic difference between writing about literature and writing about technical/factual material. Examining the university entrance test (of written English) supplied to the seminar by Lukmani, this dramatic difference in style could clearly be seen. It could also be seen that many test items were conservative in topic and item-type, bearing considerable resemblance to, for example, UCLES First Certificate in English questions from the 1970s. In our group’s review of exam papers from India we noticed the frequency of archaic style and lexical choices, which although certainly within Standard English, would be unlikely to occur in everyday Modern Standard British English.

Odlin (1989) points out that one of the most strikingly non-standard features of English as spoken by Indians is stress patterning (/di/visions; /them/selves). According to Odlin, the maintenance of stress patterns that are known by speakers not to be Standard English is a sign of the social significance to speakers of their own language, that is, it may be a deliberate language choice. Despite its being an official language in India, English is often found to be politically unpalatable, and the maintenance of a distinctly ‘Indian’ stress pattern may well be a subconscious form of resistance. According to Ethnologue, India has no native English speakers and only 11 million second language speakers – or just 1.1 per cent of the population (http://www.ethnologue.com/info.asp: 23.3.03). There is no all-India test of English, national or international, and yet there is a strong demand and expectation in employment for good English. Despite Lukmani’s own position, then, it seems that conditions in India are right for the development and acceptance of non-standard Englishes. Many universities devise their own English test and set its standards and criteria. Complexity arises at the rating stage, because teachers who mark/rate student work presumably vary in their own level of conformity to a ‘standard’ model, in their recognition and tolerance of non-standard forms, and therefore in their rating behaviours. It is clear that rating and raters are a key area requiring research if we are to fully...
understand whose norms are operationalized in national and local English proficiency tests (Hamp-Lyons and Zhang, 2001).

INTERNATIONAL TESTS OF ENGLISH

TOEFL

TOEFL (the sampler version, 2001) states in the rubric for Structure (Practice Questions and Review): ‘This section is designed to measure your ability to recognize language that is appropriate for standard written English.’ And of course what is meant is Standard American English. TOEFL is an American test, developed and run by Educational Testing Service in Princeton New Jersey, and was taken by approximately 700,000 people in about 100 countries in the year 2000. Differences between Standard American English and other metropolitan standards are often asserted to be minimal, and to be confined to certain well-known orthographic conventions and punctuation features; and also to the reference range of certain lexemes. However, we are not aware of any study which has measured the difference between Standard American English and other standard Englishes, and it is a claim which must be approached cautiously. Reference range is specific to a particular population so that the occurrence and distribution of a lexeme’s usage and meanings within a community is part of the sociolect of that community. For example, terms such as ‘bird-watching’ or ‘college’ have different referential meanings in British, American or Australian English. The American husband of one of the authors of this paper was embarrassed on his first day teaching in Australia when a student from another class asked if he could borrow a ‘rubber’.

Even straightforwardly referential terms such as ‘Harvard University’ or ‘radar-tracking stations’ might be problematic in contexts where the entities themselves do not exist. Identical objects are named differently in different inner-circle native Englishes and may indeed have other different names in national varieties of English: ‘stove’, ‘oven’ and ‘cooker’ all refer in some sense to the same object, but may not be recognized between inner-circle Englishes, or else the term is used somewhat differently in other Englishes.

An even more complex issue relates to vocabulary frequencies: until corpora are available for Singapore English, Hong Kong English, Indian English, etc., structured in ways that enable them to be compared, and further compared with corpora from inner-circle Englishes (including control for discourse domain, etc.), applied linguists are in no position to compare frequencies in world English contexts with those in British and American contexts.

In our discussions, we picked out several instances where US-specific knowledge might be required, or where US conventions of grammar might confound regional conventions, causing a problem for users of non-standard varieties:

(1) After serving as a teacher and principal in Mason City, Iowa, Carrie Chapman Catt became an organizer and lecturer for the woman suffrage movement of the 1980s.

Candidates were invited to locate the error in this sentence, choosing from ‘serving’, ‘and’ and ‘of the’. But some of us felt that ‘the right answer’ was the use of ‘woman’ (as opposed to ‘women’s’), while others pointed out that collocations such as ‘Department of Woman Studies’ are currently preferred among feminists. The item might have been clarified as a
grammar item and not a culture item if the sentence had said ‘the woman’s suffrage movement’ leaving number but not possessive to be corrected.

(2) Because of its size and proximity to the Earth, the surface of the Sun . . . . . . . . . . . . .
studied in detail.
(can be, being, it is, it can be)

Since, as we saw earlier, resumptive pronouns are acceptable in some New Englishes, there may be a possibility that speakers/users of some Englishes would find ‘it can be’ correct.

In summary, the problems we noted in the TOEFL sampler were few and minor; but we must again make the point that whether or not they will affect scores is a matter for empirical research and not of opinion.

IELTS

The IELTS test is a British/Australian test, originally developed (as the ELTS) by the British Council, but now developed and run jointly by the University of Cambridge Local Examinations Syndicate and IDP Australia with the British Council. IELTS was taken by 140,000 people in over 200 countries in the year 2000. IELTS follows British English conventions: for example, phrasal verb structures are British rather than American and the orthography consistently British (even the ‘s’ and not ‘z’ in verb spellings). In recent years IELTS development has aimed to widen the ‘norms’ used in test materials. For example, in the Listening module there is a range of voices, including Australian and American (Taylor 2002).

In the Seminar the group examined the Specimen Materials (1995). We encountered some culturally specific knowledge (‘miles’, ‘feet’ [as a distance measure], ‘Hiroshima in 1945’, ‘5 on the Richter scale’, ‘Fahrenheit’) but these were no more arcane or exotic than in any test text anywhere in which the local context is reckoned to be sufficient for understanding so that the reader does not need background knowledge, and in the version we reviewed answers to items were not critically dependent on exact comprehension of these lexical items. The vocabulary range in the specimen test was wider than in the TOEFL sampler: for example, we found ‘micro-plate, ‘eruption’, ‘tremor’, ‘volcanologists’, ‘avalanche’, ‘pulverised’, ‘spasmodic’, ‘viscous’, ‘vents’, ‘magma’, ‘decompression’, ‘lava domes’, ‘atmospheric particles’, ‘mantle’, ‘negligible’, ‘acumen’, and ‘graphology’ – as opposed to fewer and perhaps less demanding lexical items such as ‘lumber’, ‘geometric’, ‘aesthetic’, ‘global’, ‘rotates’, and ‘titanium’ in the TOEFL sampler. Once again, however, the test items were very limited in the demands they made, and few of these lexical items were essential to the comprehension required for arriving at a correct answer.

It is difficult to detect any features either in the questions (and texts) or in the rubrics that would elicit unusual or different answers from New English speakers. Paradoxically, what an analysis does yield is examples in the IELTS material of discoursal features unusual to a native speaker, for example:

1. ‘the unstructured interview . . . fares little better’
2. ‘a small hand camera’
3. ‘Hitherto, virtually all photographers developed and printed their own pictures.’

We are led to speculate that IELTS may incline towards ‘colonial Englishes’ by the use of its outmoded forms.

It appears, then, that in the international tests there is little that is obviously
discriminatory. This was, after all, what Lowenberg found. But without an empirical study, the extent (if any) to which New English speakers perform differently on an international language test and a local language test still remains to be established.

THREE QUESTIONS

As a result of the judgement study and our critical review of the literature, we have three questions which frame our ongoing research.

*How possible is it to distinguish between an error and a token of a new type?*

This is perhaps another way of asking whether we can ever distinguish between a stage of learning (where learners are acquiring a distant, that is non-local standard) and a stage of transition where a new local code is in process of formation.

Over 25 years ago Bill Crewe wrote:

(there are) certain features of Singapore English which most Englishmen and Americans feel intuitively are non-native – for example the syllable-timed rhythm, the universal tag question 'isn't it', certain intonation contours – but the difficulty is that the existence of native dialects possessing these or similar features would invalidate the point [that they are non-native] ipso facto . . . it is virtually impossible to establish a criterion of non-nativeness with regard to any features in any dialect which is not invalidated by the existence of a similar feature in a dialect within the acknowledged native speaker area. (Crewe, 1977: 100)

*If we could establish bias, how much would it really matter?*

‘Bias’ in the technical, language testing, sense by definition applies to a whole group not just to individuals: in this sense of bias, individuals cannot be biased against. Given the differences that exist among native speaking communities (think Scotland, Australia, different regions and groups in the USA) would a WEs speaker be more disadvantaged? This is a very complex question to explore, but it bears more likelihood of an outcome than most of the other questions.

*Does an international English test privilege those with a metropolitan anglophone education?*

Note that the issue does not normally arise for local tests for local consumption (as in our Indonesian teachers’ test example). However, it is interesting that very recently the Hong Kong University Grants Committee in opting for an exit English test for graduating students rejected the locally developed, custom-designed test, with Hong Kong accents and references, in favour of the international IELTS test.

WHICH STANDARD?

What is at issue in comparing international and local tests of English proficiency is which standard is under test. The question then becomes: does a WEs variety ‘own’ (in the sense of accept) a standard of its own which it appeals to in a local test? If not, then the assumption is that speakers of this WEs will be required to operate in the testing situation in the IE standard. That is precisely the point made strongly by Lukmani. But are such WEs speakers discriminated against in being required to do this? That is, as we have said, an empirical question.

© Blackwell Publishing Ltd. 2003
So far we have taken it for granted that the norms of a metropolitan Standard English (British, American, etc.) are clear and the codes well described. Standard English, claims Quirk, is the English we take for granted, English which is not strange or unusual or different in any way, what is sometimes referred to as the unmarked variety: ‘standard English is particularly associated with the English that is intended to have the widest reach, and in consequence it is traditionally associated most of all with English in not just a written form but a printed form’ (Quirk, 1985: 123). The Australian Pam Peters goes further and claims that whatever the accent with which the educated speak English, their speech just as much as their writing will use the grammar and so on of Standard English. She makes the interesting link between Standard English and international English and no doubt this link is influential in her widening the range of Standard English to the spoken language since so much of international English (telephone, radio, television) is spoken (Peters, 1995).

But the issue is not so clear-cut, as Bex and Watts ruefully conclude: ‘it soon became apparent that there was no general consensus as to what constituted Standard English’ (1999: 1). Now if there is uncertainty about the Old Variety of English (OVE) Standard Englishes, it is hardly surprising that there should be vagueness as to what constitute New Varieties of English Standard Englishes. What seems clear is that among the educated, both in the OVE and the NVE domains, the differences across the Standard Englishes may be small. To what extent those small differences matter linguistically, in terms of understanding, intelligibility and attitudes, is unclear. What we need to know, and so far do not, is to what extent these small differences matter operationally, in terms of test scores.

REFERENCES

Hill, Clifford and Parry, Kate (eds.) (1994) From Testing to Assessment: English as an International Language. London: Longman.

(Received 20 August 2003.)