

**Linguistic and cultural norms in language testing:
A case study**

Annie Brown
The University of Melbourne
Tom Lumley
Hong Kong Poly University

Abstract

International communication within the ASEAN region is commonly conducted through the medium of English. Users of English within these countries, many of whom are likely to have relatively little contact with native speakers of English, may not have the same need as people living within an English speaking society to aspire to a 'standard' form of the language (such as British or American English); it may be more appropriate for them to develop a communicative competence employing the socio-linguistic and cultural norms of the region. These norms relate as much to the topics or content of language use as to formal aspects of expression.

Taking such an approach in the development of tests of English proficiency involves the rethinking of basic assumptions concerning both selection of test content and the notion of the native speaker as the 'ideal'.

This paper considers these issues in the context of a test of English proficiency recently developed for use with English teachers in Indonesia, based upon the view that:

1. input provided as stimulus material must be culturally appropriate in both content and form, and
2. output produced by test candidates must be evaluated without using the native speaker as the norm.

The practical implications of such an approach to English language proficiency testing are examined. These include the use of authentic local texts, both written and spoken, involving non-native speakers from the region as well as native speakers, and also non-native speakers as judges of proficiency, employing assessment criteria

which refer to the local context. Reactions of those involved in the trial testing situation, as test taker and as rater, are also examined.

1. Introduction and rationale

One effect of the increasingly important role of English in both international and intranational communication is that English has become a compulsory subject in secondary and even primary schools in countries where it has not traditionally had widespread use ("Expanding-Circle" countries, Kachru 1990). A consequence of this increasing need for English language teaching is a growing demand for English language *teachers*, a demand, however, which is also accompanied by concerns about the proficiency of the teachers recruited. A means of addressing these concerns is to institute assessment of the English language proficiency of trainee language teachers. This may be for purposes of registration, to ensure that minimum standards are adhered to, or for diagnostic purposes, relating to the provision of in-service language upgrading for teachers. In this context, it was decided in late 1992 to develop a test of English language proficiency for pre-service teachers of English in Indonesia. This was a joint project conducted by staff at the Language Testing Research Centre of the National Languages and Literacy Institute of Australia (NLLIA-LTRC) at the University of Melbourne and the SEAMEO Regional Language Centre, Singapore (RELC), and also involved the participation of teacher training institutions (IKIPs) in Java. This paper discusses issues of cultural appropriateness addressed during the development of the test, including the test content and the procedures through which candidate performance is judged.

2. Existing tests

Why was it necessary to develop a new test? There are many widely-used English language tests, such as those developed in Britain (e.g. the Cambridge First Certificate [FCE] and the Certificate of Proficiency in English [CPE]), in Britain and Australia (the International English Language Testing System [IELTS]), or in the USA (e.g. the Test of English as a Foreign Language [TOEFL] and the Test of English for International Communication [TOEIC]). These were, however, considered unsuitable in this context for a number of reasons.

Firstly, these tests either aim to measure general proficiency in English or else they are tests of English for Academic or Business Purposes. Neither of these purposes fits well with the requirement of assessing the kind of language proficiency needed by English teachers working in Indonesian secondary schools.

Secondly, partly because they assume that those who take them will wish to use English in countries where English is a native language, they generally have some idealised notion of the 'native speaker' as the ultimate goal, and therefore the highest level of proficiency.

This native speaker ideal is reinforced by the typical use only of assessors who have a strong claim to being 'native' speakers of English. While the purpose of this is presumably to ensure that candidates' performance is both understandable by and acceptable to people within the relevant context (i.e. the country to which the candidate aspires), it promotes an Anglocentric view of what is acceptable. In the case of Indonesia, furthermore, it may well be the case that few teachers and students will ever visit countries where English is the native language, and that they are far more likely to need English for interaction with people from their own region.

Thirdly, the content of such tests tends to be rooted in the Western culture of their country of origin. Indeed, many tests, such as TOEFL and CPE, are intentionally and explicitly based on the norms of a particular country as, for the most part, these tests are designed for people intending to function within the country of origin of the test. Thus the orientation of such tests is to the native speaker setting and reflects the demands of communication with native speakers of English in English speaking countries.

We then considered using existing locally developed tests, particularly those administered by teacher training institutions (IKIPs) in Indonesia to their students. While it was anticipated that this move might overcome the problems mentioned above, it brought with it two other problems. Firstly, such tests generally focus on achievement of the course objectives which typically include mastery of the grammar or other aspects of knowledge of language, rather than on the kind of communicative competence required by teachers using English in the classroom. Secondly, as a result of this primary purpose, the tests have been constructed by

* * * * *

lecturers within the institutions for their own use, and have not been subject to any process of validation or standardisation across institutions.

The decision was made therefore that a test should be developed specifically for use in the context outlined above. The process of reviewing and rejecting existing tests had clarified two further issues which needed to be addressed before development of the new test specifications could begin. Firstly, what model of English should the test take as its basis? In other words, against what variety of language (a 'native' standard or not) are the learners to be assessed? And, secondly, what cultural content is appropriate to the test?

3. Varieties of English and cultural content

While reference is commonly made to the 'native' standard and the 'native speaker', the 'educated native speaker' typically being the benchmark against which English learners are implicitly or explicitly assessed, there is in fact no single 'native' standard (Davies 1991). Not only are there differences between English-speaking countries in terms of what is acceptable, but these differences may also extend to varieties within the one country (Scots English and English English, for example). When one considers the standards to be applied in a country such as Indonesia, the question of *which variety* arises.

Furthermore, in a country such as Indonesia, it is questionable whether the aim of school-based English language teaching is attaining international communicative competence. It is likely that only a relatively small proportion of learners will need to interact with native speakers in a native speaker environment. Most of their contact with English will be on a local or regional level. It is also likely that only a small proportion of the teachers will have had much, if any, contact with or exposure to native speakers of English. Few of them, for example, will have lived or studied in English-speaking countries and most will have learnt English from non-native speakers. Thus, even those teachers who are highly competent in the language are likely to be competent in the 'local' variety, not in a native speaker one.

It appears, then, that native speaker varieties of English are somewhat irrelevant to the context we are looking at here, i.e. the English taught and learned in secondary schools in Indonesia. Examination of the research literature indicated that despite copious discussion of varieties of English, little has in fact been written concerning their relevance to the assessment of language proficiency (perhaps because assessment requires an explicit description of standards, and these varieties are ultimately considered inferior in some way).

The role of local varieties and their appropriateness in testing is, however, addressed indirectly in literature relating to the ethics of testing. The Code of Fair Testing Practices in Education (Joint Committee on Testing Practices 1988) has been proposed as the basis for a code of ethics for the International Language Testing Association (ILTA). It names as one of its principles for developing and selecting tests that "test users should select tests that meet the purpose for which they are to be used and that are *appropriate for the intended test-taking populations*" [emphasis added].

Davidson (1993: 114) argues that "Given the astonishing phenomenon of world growth in English varieties and the consequent irrelevance of 'native speaker', part of the test's standard is ... the linguistic norm it promotes", and continues "The study of World Englishes would suggest ... that there is no reason to pursue a dominant norm, and hence no reason to allow any test to do so It would be detrimental to believe that the linguistic standard promoted by a language test is somehow divorced from other considerations of testing ethics; in the same spirit, it would be unwise to develop and promote EFL tests without attention to the linguistic norms to which those tests adhere."

As far as cultural content of the test is concerned, Kachru points out that for learners of English motivated by instrumental concerns, such as those using it as an *auxiliary, library or link* language, "acquiring a second culture is not the main motivation for learning the language" (1983: 75). We would argue that this is certainly the case in Indonesia; most students in Indonesian schools will be unlikely ever to come into contact with the native speaker culture, let alone aspire to be part of it. As has been pointed out by Hamp-Lyons (1983: 45-6), "it is no longer true that to learn (English) one must learn the culture". Similarly, the majority of English teachers in Indonesia,

while they may wish to have some understanding of the culture of the native speaker of the language they are teaching, are also unlikely to aspire to belong to it. Thus in teaching and testing English in Indonesia it would seem more appropriate to draw on cultural content familiar to the learners, allowing them to interact with texts which are relevant to their interests and likely future language use.

It was felt, therefore, that in this context it was appropriate to retain a sense of the setting of the test, and to base it on a model which was both culturally and linguistically appropriate. The resulting test should thus be based on:

- a recognition of the practical limitations of the teaching context, where local teachers will never attain, in large proportions, native-like communicative competence and cannot therefore be expected to pass on a native model to learners; and
- a recognition of the Indonesian variety of English as a valid model.

4. 'Localising' the test

Once it had been determined that the test would be 'localised', the next step was to develop test specifications. This involved consideration of how 'localisation' would be achieved in terms of the test content. A reconsideration of the reasons for which currently available tests were rejected helped in defining what *was* needed:

- The test should include content specific to the target context, i.e. the domains in which Indonesian teachers of English could be expected to require English.
- The test must be firmly rooted within the local Indonesian context; there is no justification for orienting it to native speaker culture and norms. Therefore, texts and tasks must be appropriate to the context of use of the target population for the test.

- Assessment must be carried out in relation to acceptable performance *within the future context of use*, i.e. the ideal is the 'good' local user rather than the native speaker.

5. Contexts of use

A performance-based model of assessment (Wesche 1987, McNamara 1996) appeared to be the most appropriate test format for determining whether the trainee teachers had the requisite skills to use the language effectively in carrying out their duties as teachers, rather than simply possessing knowledge *about* the language. Such a model would have a further advantage over more traditional discrete item knowledge-based tests in that a secondary intention of the test developers was to provide a good model for language assessment and teaching through its communicative orientation.

We needed therefore to define the types of situations and topics teachers would be expected to be able to deal with. This involved two questions:

1. What kinds of 'texts' (spoken and written) are likely to be used within the classroom, i.e. what are learners likely to be presented with or expected to produce?
2. What do teachers need to do through the medium of English?

With regard to the first of these questions, we started from the assumption that a good communicative curriculum would involve the use in the classroom of authentic texts of the type that are likely to be relevant to Indonesians within their own country (Widdowson 1983). These text types will necessarily be limited, as the range of English texts and discourse encountered by learners of English in Indonesia will typically be narrower than those which a native speaker might encounter. Johnson (1990) gives an indication of the types of texts (spoken and written) likely to be the most relevant to speakers of what he terms International English (IE). The notion of IE is relevant to discussion of non-native varieties of English as these are likely to share some of its characteristics. He notes that users of IE are "necessarily limited in the range and delicacy of the functions they can perform" (1990: 308), and summarises its features as follows:

-
- It is not used for solidarity/identity with others;
 - It is not used for social distinctions within the group;
 - It is used for "formal education";
 - The ideational component is rich in CALP rather than BICS;
 - Speakers are less able to change level of familiarity; it is more "formal";
 - It is used for transactional goals rather than interpersonal goals;
 - It is used in more context-reduced situations.

Thus he points out that the type of communicative competence of speakers of IE would "be restricted largely to the transactional functions of the academic, commercial and professional world, and to discourse strategies appropriate to context-reduced communicative situations" (1990: 308).

In relation to the second question (What do teachers need to do through the medium of English?), the professional demands of English language teachers are likely to require them to interact with English more frequently than, and with different purposes from, the majority of Indonesian users of English. We were able to identify three areas of English use for teachers:

- gathering and preparing the requisite teaching texts. This would involve the perusal and selection of available authentic English language materials, both written and spoken, and their adaptation for teaching;
- communicating in English within the classroom with the students in order to:
 - present content,
 - talk about language; and
 - get things done (see Elder 1994 for a more detailed analysis of what this involves);
- taking part in professional training and development activities. The skills required to participate in professional

development activities are also largely defined within Johnson's (1990) listing above, and include reading academic journals, discussing methodology- and syllabus-related issues, etc.

6. Test design

Two integrated tests were designed, Reading/Writing and Listening/Speaking. In general, the texts and tasks, both receptive and productive, reflected the kinds of language claimed earlier (following Johnson 1990) to characterise English language use in Expanding Circle countries, i.e., fairly formal, cognitive, academic, context-reduced language or language used for transactional purposes.

6.1 Reading/Writing Test

The test was designed to sample from two domains of language use, classroom-related tasks and professional development activities.

Communicative language teachers need to be able to select texts and prepare appropriate accompanying activities for use in class (classroom-related tasks), and ideally such texts need to be part of the schematic world of the *learner* (Widdowson 1979) in order to promote learning. As one aim was to promote good teaching practice, it was considered necessary to draw on *local* English materials. Appropriate sources for reading texts of this kind included English language magazines and newspapers.

Because the texts were taken from local sources, we were able to assume their cultural familiarity to the test takers, thus addressing the problem in test design of ensuring test takers have adequate background knowledge to complete the tasks. For example, one of the texts chosen from the first version of the test was selected because it dealt with a natural product found almost exclusively in the South-East Asian region. It was also deemed to represent the conceptual level of a middle or higher secondary school text, and therefore the kind of material appropriate for use with secondary school students.

It was also considered necessary to present texts relevant to the academic domain, or professional development activities in which

teachers participate *outside* the classroom, such as in-service activities. Here the content relates to English language teaching, a discourse domain with which trainee teachers can be assumed to be already familiar through their training, and which they are expected to maintain through in-service activities and, possibly, further study.

Reading texts for the test were selected from the types of reading materials teachers would be likely to make use of in the professional development context, i.e. journals or other publications addressed to English language teachers.

The writing tasks also reflected the two domains of English language use - the classroom and professional development. Writing tasks were integrated with the reading texts in two ways. Firstly, the reading text becomes the input or stimulus for the writing task. Tasks representative of the classroom domain included, for example, providing explanations for lexical items or composing comprehension questions using a specific reading text; in other words, preparing a text for classroom use/teaching purposes. Secondly, the role of the reading text was to provide the background information needed to complete the task in the more extended piece of writing. This required candidates to complete a task such as commenting on teaching methodology in response to a journal article. This type of task represented the academic or in-service requirements of English language teaching.

6.2 Listening/Speaking Test

As with the reading, the listening texts were designed to be relevant to the classroom demands of teachers as well as representative of the professional development activities they may participate in. Text types included brief informal conversations, short monologues (e.g. radio news items) and extended discussions (including both native speakers and non-native speakers of English). These discussions were either on topics appropriate for use in the classroom (e.g. social themes or personal experiences), or reflected professional development topics (issues related to language teaching).

Similarly, the speaking tasks were of two types. The first required candidates to produce the kind of language that a teacher might need in the classroom, for example, telling a story or giving

instructions to a class of students. The later section, integrating listening and speaking tasks and relevant to professional development activities, required candidates to respond orally to a discussion on language teaching.

In the listening test it was considered important to include a variety of spoken models, both native and non-native (American, Australian, British and Indonesian), for three reasons:

- to reflect the range of types of English which candidates could potentially encounter;
- because it was not possible to predict exactly what varieties each test taker was familiar with, and including only one variety (e.g. Australian) could disadvantage candidates not familiar with this variety;
- there was potential for beneficial washback onto the curriculum by demonstrating that both native-speaker and non-native-speaker varieties are appropriate in English language education in Indonesia.

7. Test standards

In any criterion-referenced test it is necessary to define the standards against which candidates are to be assessed. In some tests this is explicit in that, where descriptive criteria (e.g. scales or other level-related statements) are used, reference is made at the highest levels to native-speaker-like behaviour. In others it is implicit in that where native speakers (most commonly teachers) are used as assessors it is assumed that they will assess candidates against their norms, i.e. those of the *educated* native speaker. It was, as discussed earlier, considered undesirable in this context to refer to a native-speaker ideal. This decision was consistent with the practical requirement that the test be fully administered in Indonesia, utilising local IKIP lecturers as assessors.

It was therefore necessary to consider how the local variety might be expected to differ from the native speaker ideal. Most local varieties of English have not been described; this is certainly the case with Indonesian English. Hence it was not possible to present a fully elaborated or comprehensive description of the model against

* * * * *

which candidates were to be assessed. The assumption is made when educated native speakers are used as assessors that they have in their heads a model of acceptable language use. In the same way, if proficient users of the local variety of English (i.e. IKIP lecturers) were employed in this context, then it seemed reasonable to assume that they had a good understanding of what constituted 'good' local usage, without it being described explicitly to them.

Once the test and the assessment criteria were produced, it was necessary to present both of these to the local raters, firstly to explain the theoretical basis of the test and secondly to train them in the use of the assessment criteria. This training consisted of the participants developing a local consensus on the interpretation of the assessment categories and levels described in the scales, based on discussion of actual performances, spoken and written, taken from the test trials. In order to establish this consensus it was considered necessary to 'raise their awareness' of relevant features of language, and in particular, those which might distinguish the non-native variety of English from the native variety.

A review of writings on non-native varieties of English (e.g. Platt, Weber and Ho 1984; Johnson 1990; Lowenberg 1993; Kachru 1986; Mohanan 1992; Clyne 1987; Kachru, Y. 1992; Reynolds 1992) indicated that there were four main areas where differences might be observed:

- morphosyntax
- lexical and idiomatic usage
- phonology
- discourse style

These four general areas related to the assessment categories in the following way:

- intelligibility and pronunciation (related to phonology)
- grammatical accuracy and sentence structure (morphosyntax)
- control, breadth and appropriateness of vocabulary (lexical and idiomatic usage)
- coherence and cohesion (discourse style)

-
- a general category of overall task fulfillment

8. Feedback

We were interested in these questions concerning the test's suitability:

- Would the test be seen as culturally appropriate?
- Would it be seen as a test appropriate for testing the English language proficiency of Indonesian school teachers?
- Would a local variety of English be perceived as acceptable:
 - a) as a spoken model in the listening test?
 - b) as the basis for the assessment criteria for both the written and spoken production of the test takers?

In order to assess the degree to which the test was appropriate and acceptable in these various ways, questionnaire responses were sought from the 200 candidates (trainee teachers in the final year of their studies, from four IKIPs in Java) who took part in the trials, as well from the 24 lecturers, from the same four IKIPs, who participated as raters. Additional feedback was also obtained from the lecturers during the rater training sessions.

A 4-point scale was used to elicit reactions to the suitability of the test in general for a number of purposes (Table 1).

How good do you think this was as a test of ability to use the kind of English needed for the following purposes:

	Mean ratings	
	test takers	raters
a. for everyday communication?	2.7	2.9
b. when preparing to teach English classes?	3.2	2.8
c. while teaching English in class?	3.3	3.0
d. for professional development activities/courses conducted in English?	3.4	3.1

(4 = very good, 3 = good, 2 = fair, 1 = poor)

Table 1. Summary of questionnaire responses from test takers and raters

The test was generally viewed as suitable for these purposes, although more so as a teacher-specific test (b, c, d) than as a general test of proficiency (a). It is interesting to note that the test takers as a group (mean ratings = 3.2 - 3.4) considered the test to be better suited to its purpose as a teacher-specific test than did the raters (mean ratings = 2.8 - 3.1). This may be the result of their more recent and immediate experience of classroom teaching, and very possibly bodes well for the introduction of a more communicative language teaching curriculum.

Typical responses from test takers were by and large very enthusiastic, including comments such as: "I learned from it"; "interesting"; "challenging"; "varied"; "well structured and graded"; "unfamiliar but good"; "a good model for a test and suitable for SE Asia", amongst numerous additional favourable comments. Similarly, the raters were keen to take away copies of the rating scales, and would have liked to keep copies of the test itself.

Problems identified revolved mainly around the unfamiliarity of the test format, rather than the content. This was to be expected, given that communicative testing techniques are not generally employed in this context. However, one anticipated benefit of the test design was that teachers would learn from it how to incorporate more communicative language assessment into their classroom activities (known as washback). In the light of the favourable reception to the test by both test takers and assessors there are grounds for optimism on this point.

As for the issue of what varieties to model in the Listening/Speaking Test, there were mixed reactions. Some lecturers questioned the appropriateness of using non-native (i.e. Indonesian) speakers in the listening test, as did some of the test takers. On the other hand, similar numbers of test takers either complained about the difficulty of understanding some of the native speakers or requested native speaker accents which were more familiar to them. Clearly, the issue of whether to use native or non-native speakers, and which native speakers, requires further debate. It would appear, however, that the concept of the non-native model of English is not yet fully accepted in Indonesia.

In the assessment criteria used in the test, as stated, there was no reference to the native speaker. The highest levels on the rating scales referred to the notion of the proficient speaker working in the classroom and included a level of tolerance of non-nativeness. This does not imply simply that expectations are lower (that the native-speaker is above the top level of the scale), rather it involves an acceptance of features (phonological, lexical, syntactic, discorsal) which would be considered wrong in relation to native-speaker norms, but nevertheless acceptable in Indonesian English. The lack of reference to native-speaker norms initially caused some surprise amongst raters; however, the rating scales and the model implied in them were accepted by the raters and they had no more difficulty in using them than would a comparable native speaker group.

9. Conclusion

In this paper we have described a test development project in which a primary aim was to ensure cultural and situational appropriateness in relation to the test's purpose and context of use. A subsidiary aim was to achieve some measure of washback onto the language teaching curriculum (both that within the teacher training institutions, but also that of the school classroom) by providing a model of 'good practice' in terms of communicative testing. These aims were achieved, we believe, through

- a) the judicious selection of tasks relevant to teachers of English in Indonesia. These tasks reflect two areas of teachers' required English usage: classroom-related usage, i.e. teaching and preparation for teaching, and training and professional development.
 - b) the selection of culturally appropriate content. Test materials (input texts) were drawn from local materials of the type that would be relevant to and used by teachers for either of the two purposes (teaching or professional development). These included professional journals, newspapers and magazines.
 - c) an emphasis on assessing test takers in relation to regional norms. In training the raters, local norms of English language usage and production were established as the model against which candidates were to be assessed. In addition, the native
- * * * * * * *

speaker was not set as the 'ideal', that is the highest level of test achievement.

- d) the use of local raters, i.e. non-native speakers of English (whose level of proficiency was nevertheless of a high standard) who are familiar with the local regional norms, and who are representatives of the local test users' community.

Ultimately, no solution we devise as test developers will be acceptable to all those involved in using the test, because acceptability relates intimately to individual values. However, the model we have put forward appears to address central factors presented in the first part of this paper, and to have been generally well received by both test takers and raters, in the context in which it is proposed for use. We have attempted in designing this test to incorporate a set of general principles that are appropriate for such a test of language proficiency, taking into account as wide a range of factors as was practicable given the constraints of the project.

10. References

- Clyne, M. (1987) Discourse structures and discourse expectations: Implications for Anglo-German academic communication in English. In Smith, L. E. (ed) *Discourse across cultures*. pp. 73-84. Prentice Hall: Hemel Hempstead, Herts.
- Davidson, F. (1993) Testing English across cultures: Summary and comments. *World Englishes* 12, 1: 113-125.
- Davies, A. (1991) *The native speaker in applied linguistics*. Edinburgh: Edinburgh University Press.
- Elder, C. (1994) Performance testing as benchmark for foreign language teacher education. *Babel* 29, 2: 9-19.
- Hamp-Lyons, L. (1983) Motivation for learning English as a world language: Integrative and instrumental. *World Language English* 2: 45-49.
- Johnson, R.K. (1990) International English. *World Englishes* 9, 3: 301-315.

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

Kachru, B. (1986) *The alchemy of English: The spread, functions and models of non-native Englishes*. Oxford: Pergamon.

Kachru, B. (1990) World Englishes and applied linguistics. *World Englishes* 9, 1: 3-20.

Kachru, Y. (1992) Speech acts in the other tongue: An integrated approach to cross-cultural research. *World Englishes* 11, 2/3: 235-240.

Lowenberg, P. H. (1993) Issues of validity in tests of English as a world language: Whose standards? *World Englishes* 12, 1: 95-106.

Mohanan, K.P. (1992) Describing the phonology of non-native varieties of a language. *World Englishes* 11, 2/3: 111-128.

McNamara, T. F. (1996) *Measuring second language performance*. London: Longman.

Platt, J., H Weber and M.L. Ho (1984) *The new Englishes*. London: Routledge and Kegan Paul.

Reynolds, D.W. (1992) Illocutionary acts across languages: Editorialising in Egyptian English. *World Englishes* 11, 2/3: 35-46.

Wesche, M. (1987) Second language performance testing: The Ontario Test of ESL as an example. *Language Testing* 4, 1: 28-47.

Widdowson, H.G. (1979) *Explorations in applied linguistics*. Oxford: Oxford University Press.

Widdowson, H.G. (1983) *Learning purpose and language use*. Oxford: Oxford University Press.

(Paper completed 1996)