The competing claims of accuracy and fluency in the construction of performance tests of language proficiency: two cheers for Robert Lado![1]

Alan Davies

Abstract

The difficulty of distinguishing between accuracy and fluency, both theoretically and empirically, is considered and possible solutions proposed and discarded (Bialystok's automaticity/knowledge model and the native speaker model). A more hopeful solution is to be found through the operationalising of language testing. Accuracy is related to discrete point tests and fluency to integrative tests. The current emphasis on performance testing (and particularly on tasks) is shown to reflect a greater concern with fluency than with accuracy. But the importance given by Bachman to the underlying or component abilities indicates an equal concern with accuracy. The need to balance the two appears as strong in the 90s in the work of Bachman as in the late 50s in the work of Lado.

1. Introduction

What I shall argue in this paper is:

1.1. Accuracy and fluency cannot seriously be distinguished, since, for example, being fluent pragmatically means getting it right in terms of accuracy and being accurate in on-line speech processing means being fluent, that is, not being dysfluent.

1.2. Accuracy and fluency are both attributes of the native speaker and it may be helpful to consider them separately.

1.3. In practice, language testing has always been concerned with both accuracy and fluency (or, in more traditional testing terms, with discrete point and integrative items).

---

[1]A version of this article was read at the Certification Conference held at the University of Siena in May 1996.

2. Defining accuracy and fluency

Accuracy and fluency are so-called pre-theoretical terms, totemic appeals, where they are not claims on knowledge of the grammar (accuracy) and on unpremeditated speech (fluency). The only more theoretical use appears to be in clinical linguistics where the term 'dysfluency' is current, relating to problems of on-line processing of speech, eg 'the repetition of syllables, part words or phrases within an utterance, revisions within an utterance (false starts) or significant pauses within an utterance' (Fletcher 1985: 102). Otherwise, fluency is used in connection with what it is the native speaker is (ideally) thought to be capable of, as in Spolsky's Condition 12: 'Automaticity condition: ability to use language varies in automaticity, this is shown by the fluency with which a person speaks' (Spolsky 1989: 49).

Richards, Platt and Platt (1992) define fluency as follows: 'the features which give speech the qualities of being natural and normal, including native-like use of pausing, rhythm, intonation, stress, rate of speaking and use of interjections and interruptions.'.

In second and foreign language teaching, fluency describes a level of proficiency in communication which includes:

a.   the ability to produce written and/or spoken language with ease,

b.   the ability to speak with a good but not necessarily perfect command of intonation, vocabulary and grammar,

c.   the ability to communicate ideas effectively

d.   the ability to produce continuous speech without causing comprehension difficulties or a breakdown of communication.

It is sometimes contrasted with accuracy, which refers to the ability to produce grammatically correct sentences but may not include the ability to speak or write fluently (1992: 141,2).

Brumfit's contribution to the debate has been two-fold. First, he has argued (as Johnson 1996 notes with approval) that communicative language teaching has recognised the need to provide students with

practice of doing more than one thing at the same time, while accuracy is doing only one thing at a time. The distinction therefore has to do with the combining approach of fluency. Second, he has made a strong case for making language teaching fluency-based because it 'may be closer to the apparent learner syllabus of the natural learner in a total immersion situation, in that the naive learner operates more on an oral basis of fluent and inaccurate language than on a careful building up analytically of accurate items according to a descriptive model' (Brumfit 1979:188).

And Skehan has summarised part of the work on the good language learner as follows: 'in the earlier stages of language learning good language learners may emphasize fluency over accuracy' (Skehan 1989: 77).

It is salutary therefore to remind ourselves that Brumfit admits that 'in one sense, as will be appreciated, the contrast between accuracy and fluency is largely metaphorical. Classrooms are always concerned with both' (Brumfit 1979: 189). But testing is not necessarily concerned with both together and may wish to consider them separately.

However, where fluency is reified into a test format (eg the former FSI Oral Interview), experimental results of the scales 'that pertain to accent, grammar, vocabulary, fluency, and comprehension... these separate ratings apparently do not contribute different types of variance and... they appear to add little to what could be obtained by simply assigning an overall rating of oral language proficiency' (Oller 1979: 321).

Oller's much-quoted conclusion appears to be supported elsewhere in the literature by, for instance, the experimental findings of de Jong and van Ginkel, who also conclude that separation is not feasible: 'only after having arrived at a certain level of accuracy can a subject expect to develop into a fluent—and comprehensible—speaker' (1992: 204).

And Rea Dickins, while arguing for a rethinking of how to test communicative grammar, implicitly accepts the case for bringing accuracy and fluency together (Rea Dickins 1991: 112–131).

3. Accuracy, fluency and the native speaker

3.1. Language testing must consider both accuracy and fluency on the grounds that they are both thought to be aspects of the native speaker. Fluency, in particular, is often identified with native-speakerness, as Bialystok (1982) argues. Bialystok's processing model of second language learning contains, as its second component, the control or retrieval procedures. She maintains that there is a difference between first and second language use in efficiency, or, as she prefers, automaticity. It is in automaticity that Bialystok finds the basis of fluency since to her fluency is distinct from knowledge, which we locate in accuracy.

Bialystok's model helps make clear that fluency and accuracy are usefully seen not as attributes of learning or of the learner but as a means of investigating and determining automaticity and knowledge. Which explains why a measure such as the FSI Oral Scale has scales labelled Fluency and Accuracy, not Knowledge and Automaticity. Again, in a part of the Richards, Platt and Platt (op cit) definition, they define fluency in terms of native-like use. My concern in this section of the paper is whether the native speaker is a useful model or criterion with regard to fluency for our understanding of the attainment of the second language learner. (I shall confine my comments here largely to the fluency aspect; a similar discussion could be held of accuracy.)

3.2. Coppieters in a much-quoted paper reports a grammatical judgement experiment (Coppieters 1987). He took a group of 27 non-native adult speakers of French who had 'so thoroughly mastered French that it was no longer clearly possible to distinguish them from native speakers by mistakes which they made, or by the restricted nature of their choice of words and constructions' (1987: 544). For baseline data, he took 20 native speakers of French, matched with the experimental group as far as possible. He used 107 sentences illustrating various aspects of French and asked his subjects individually for acceptability judgements. His results indicated that the two groups belonged to two different populations, with no overlap between them, even at the extremes. He continues: 'it is also clear that the variation between native speakers and non-native speakers cannot simply be subsumed as a special case of the variation among native speakers: that is non-native speakers have been found to lie outside the boundaries of native speaker variation'

(1987: 545). Native speakers, reports Coppieters, 'did not need the help of an explicit context. No matter how skilful non-native speakers might be at deriving the appropriate interpretation of a sentence in context, their inability to do so in the absence of an explicit context indicates a fundamental difference between their knowledge of the language and that of native speakers' (1987: 566,7).

Birdsong (1992) takes issue with researchers such as Long (1990) who appear to make an absolute distinction between the native speaker and the non-native speaker, viz that 'ultimate attainment' for the non-native speaker can never be equal to native speaker competence. Birdsong reexamines the Coppieters (1987) experiment with learners of French and reports also on his own parallel study. What he concludes is that, while as a group his French language learners and the French native speaker subjects differed significantly, the large amount of overlap suggests that 'this general lack of difference is taken as..evidence that ultimate attainment by non natives can coincide with that of natives' (1992: 739). Of course, those who overlap are, as Birdsong admits, 'exceptional learners'; but the implication here is that 'our attention should turn to the issue of trainability: what can be discovered from exceptional learners that could be applied to improve other learners' chances of attaining native norms' (1992: 742).

3.3. It appears then that the native speaker may not be helpful in measurement terms with regard to fluency for our understanding of the attainment of the second language learner. If the native speaker remains elusive as a criterion (except in some ideal way) what I want to propose is that we use a more operational approach via language test data to help describe the native speaker. After all, we still need to demonstrate how it is we describe the attainment of these exceptional learners.

Various ways of demonstrating this are discussed in Davies (1997) and will not be addressed here.

4. Testing accuracy and fluency

In practice, language testing has always been concerned with both accuracy and fluency.

I connect accuracy and fluency with the traditional distinction between discrete point and integrative tests (Lado 1961). My approach is by way of the recent concern with so-called performance tests, themselves an extension of communicative tests for specific purposes (English for Medicine, Japanese for Tour Guides etc) with their particular emphasis on the testing of oral interaction.

I argue that this development is not really innovative in terms of what to test: here the approach is quite traditional. If there is innovation it is in terms of how to test: a relevant concern is with the variability among raters of oral production tests and with the procedures now available for taking this variability into account, procedures analagous to the methods used for removing non-discriminating items from objective tests. What we see here is the deliberate objectivising of subjective testing.

I also argue that where performance tests emphasise fluency they put themselves in just the same dilemma as earlier integrative tests with regard to sampling. There is inevitably a lack of generality about all integrative/ fluency/work-sample tests, which no doubt is why performance tests have concentrated mainly on specific purposes. As a result, test reliability for performance tests of fluency is in question. On the other hand, while generality for tests of accuracy (which we also find, as we shall see, in performance tests) is much easier to claim, they raise (as Rea Dickins reminds us, op cit) a serious question of validity.

Views of performance tests vary: one is that they are only possible where there is a relatively homogeneous clientele with known and relatively specific language use needs. Their rationale is to replicate those aspects of context which can be shown to influence language performance in a systematic way and in so doing to establish greater predictive validity.

The extent to which performance tests can approximate real life settings is disputed, some writers arguing for a continuum from direct to indirect tests, offering ways of approximating as closely to real life as possible in the test situation and introducing a category of semi-direct tests. Others have argued that the actual test encounter is authentic in itself, and still others for the importance of construct validity in test construction and for a more precise analysis of the critical features of communicative language use. According to this

view, performance testing becomes the testing not of authentic texts (ie fluency) but of the authentic features which underlie such texts (ie accuracy). ·

We may summarise by suggesting that there are two distinct views of performance tests, one, the LSP/work-sample view and two, the testing of authentic features underlying authentic texts.

5. Performance testing

Bachman (1990) begins his discussion of performance testing under the general heading of authentic language tests, pointing out that 'the search for authenticity continues to be a major consideration in language testing, and tests described variously as 'direct', 'performance', 'functional', communicative' and 'authentic' have been developed and discussed in recent years' (1990: 301). He distinguishes between the 'real-life' (RL) approach and the interactional/ability (IA) approach to defining authenticity. The *RL approach*, he maintains,

> *'considers the extent to which test performance replicates some specified non-test language performance. This approach thus seeks to develop tests that mirror the 'reality' of non-test language use, and its prime concerns are: (1) the appearance or perception of the test and how this may affect test performance and test use (so-called 'face validity'), and (2) the accuracy with which test performance predicts future non-test performance (predictive validity). This approach does not, in effect, distinguish between language ability and the context in which this ability is observed, since non-test language performance constitutes both the criterion for authenticity and the definition of proficiency.' (1990: 301–2).*

In other words, it does not make the necessary distinction between the test and the criterion, turning the criterion into the predictor. Bachman judiciously observes that the RL approach has been helpfully dominant in the testing of oral proficiency in the last period and that it underlies the ACTFL/ILR oral interview.

The IA approach 'is in keeping with both the mainstream approach to measuring language as a mental ability and the current view of communicative language use (302). The 'ability' part of the IA

approach', claims Bachman, goes back to Lado, Carroll and Oller; i t informs the TOEFL concept and other large-scale institutional testing. The 'interactional' part of the IA approach shares its view of interaction with communicative language teaching.

This, says Bachman, is the 'distinguishing characteristic of communicative language use—the interaction between the language user, the context, and the discourse. It thus attempts to design tests that will involve the test taker in the appropriate expression and interpretation of illocutionary acts... (the) primary concern is with demonstrating the extent to which test performance reflects language abilities or with construct validity' (302–3). In other words, performance for Bachman resides in the test not in the behaviour to be predicted, not in criterion behaviour. A performance test for Bachman then becomes an opportunity for the testee to perform the 'features of language use that are relevant to both the interpretations and uses to be made of test scores' (317), which the test constructor has sampled for the purposes of the test. Bachman accepts that the 'ability' component of his IA approach is not new; it may be found in 'the skills and components frameworks of Lado (1961) and Carroll (1961) and in Oller's (1981) "pragmatic expectancy grammar"' (1990: 302). What is new is the interactional component.

In commenting specifically on performance tests, Bachman maintains that:

> *The crux of the problem lies in identifying performance, or behaviour, with trait or ability, and this is most apparent in the term 'direct test''' ... Language tests, like all mental measures, are indirect indicators of the abilities in which we are interested*   *(ibid: 309).*

Bachman argues strongly for the primacy of construct validity, discarding face, content and predictive validities. The problem with his argument, however, is that construct validity is necessary but not sufficient. Content validity cannot be avoided, if only to ensure coverage in the test of the skills and components which underlie communicative language use. Bachman's solution, which is to define language proficiency in terms of component abilities, makes precisely this point. He dismisses language performance definitions of proficiency because they allow no basis for distinguishing

separate abilities. And yet his own examples of component abilities are in essence mini-performance definitions, eg Use of Cultural References, which has the Scale:

No evidence of ability to use cultural references

Some evidence of ability to use cultural references appropriately

Full control of appropriate cultural references.

In other words, the outcome of Bachman's model of IA tests is a set of scales with sub-levels, each containing a descriptor of a mini performance. Less is said about the tasks provided for eliciting students' responses which trigger determination of a band score except that they should be varied and interesting or motivating.

One possible source of such tasks, the work or job sample, is dismissed on the grounds that all such selection will be inadequate sampling-wise. Jones (1979) is regarded as optimistic for his proposal that a 'test must be representative of typical job-related situations, and short enough to be practical' (Bachman 1990: 311). For Bachman, performance tests are test performances which cover the skills and components frameworks underlying communicative language use.

Work-sample tests cannot, apparently, do this because they do not adequately sample non-test language since '"real-life" language is extremely complex'(1990:312). This seems an odd reason, given the need for all tests somehow to predict precisely the ability to control this extremely complex phenomenon. I am given to conclude that Bachman has failed to maintain the distinction necessary between the scales he advocates, which provide the measure of interaction, and the task for the testee. He quotes Swain (1985) and Wesche (1987) with approval:

> *... the content of communicative language tests should also be motivating, substantive, integrated, and interactive... selection of appropriate topics... opinions or controversial ideas... reading passages, audio-taped lecture, dictation and structured composition... on a common theme related to either science and technology or social sciences* (1990: 320).

In my sense of the term, these are indeed work-sample type tests so far as the stimulus to the student goes: what remains is how s/he shows a response and how that response is judged. What Bachman appears then to mean by a test performance is limited to a specification of the judgements (of mini-performances) made by the judges.

6. The strong-weak distinction

McNamara (1990) considers performance testing of crucial importance in specific purpose testing. Drawing an insightful distinction between strong and weak performance tests, he gives as an example of a strong performance test a clinical medicine test in English for experienced doctors and other health-professionals seeking registration to work in their professions in an English-speaking country. Success on such a test would be judged in terms of both medical knowledge and English proficiency. The same test used as a weak performance test would judge success only in terms of skills in English language. In weak performance tests, getting the right subject answer is less important than making the right language choice.

McNamara's weak/strong distinction is in practice difficult to maintain, for two reasons: first, because in its strong form it is unclear that medical knowledge and English proficiency can be separated. Second, since a clinical medicine test is likely to contain work sample material, making the right language choice cannot be judged in terms of language alone; of necessity, recourse must be made to context. In the one case, knowledge needs language to encode it; in the other, language needs knowledge or content to give it meaning. On this basis, all performance tests of the medical type referred to here are more or less strong. At the extremes, the strong performance test is a medical test; the weak performance test is a test of uncontextualised English which has no claim on specific purpose.

Because performance tests have represented a marriage between direct tests and languages for specific purposes (LSP) tests, like all direct tests they suffer from problems of coverage, of generalisability and reliability, as I noted earlier. As such, questions of coverage have to do with varieties of context: in a medical test, for example, with doctor-patient, doctor-colleague; case conference; telephone communication and so on. Such work

sampling is likely to take for granted coverage of the skills and components frameworks, that is, of the language features which are needed for the communicative language ability in question. In such LSP test development, the issues are largely practical, ones of context sampling.

More recently, McNamara (1994) has argued for a more theoretical approach to performance testing. It should be noted that such an argument may represent a rejection of the LSP approach to testing: McNamara may here be reflecting the general trend, which appears to be returning to a more unitary view of language ability, without prejudice to a position on language variety. McNamara is here, I suggest, close to the Bachman position but pushes the argument on to a call for a theory of the capacities involved in language use, which presumably can lead to a systematic analysis of the skills and components frameworks, not of language but it seems of performance. What this seems to imply is not just a grammar of the language in use, not even just a grammar of the discourse in use, but a grammar of communication in human interaction.

This may be what Bachman had in mind: certainly, his advocacy of interactional ability tests implies a need for specification of what the elements are that are involved. But he never seems to make clear whether he is referring to language elements or interactional elements. McNamara takes the process to its logical conclusion. It is helpful, to have it spelled out in this way, but is, in my judgement, a vain hope.

McNamara looks to ethnomethodology and conversation analysis to provide a description of the capacities involved in language use. Richly insightful though these discourse studies may be, their forte seems to be thought-provoking rather than descriptive. If we are to make use of their insights for test research and construction purposes, we will need descriptions (like a grammar) of the elements of interaction. Without that, what we will inevitably use in a test stimulus is either a piece of conversation (that is, a kind of work sample) or a language element (eg a question).

7. Skills and components

I turn now to Lado's consideration of performance. Bachman, it will be recalled, tells us that the ability component of his IA approach

may be found in 'the skills and components frameworks of Lado (1961) and Carroll (1961) and in Oller's (1981) "pragmatic expectancy grammar"'(1990: 302).

I can find no reference to performance tests in Lado (1961). But let me turn to what Lado thought language tests should concentrate on. In a well-known sentence he wrote: 'testing control of the problems is testing control of the language. Problems are those units and patterns that do not have a counterpart in the native language or that have counterparts with structurally different distribution or meaning.' (1961:24). Although Lado's view of testing is commonly associated with discrete point tests, which indeed he does describe fully, he is in fact much more catholic: his theory of language testing assumes that language is a system of habits of communication which permit the communicant to give his conscious attention to the over-all meaning he is conveying or perceiving.

Lado's view of language (as of language learning) is that it consists of 'control of the signalling elements of the language in communication situations' (1961: 206). This is not, I suggest, just the skills and components framework which Bachman relegates as the concern of the ability side of his IA model. The 'in communication situations', I suggest, takes us on some way into the Interactional aspect.

But Lado goes even further. In his integrated tests of speaking, one task suggested for the upper level is that of Sustained Speech :

> ... *you have met a young German in Europe who seems to you to have the makings of an outstanding American citizen. You resolve to try to convince him that he should emigrate to the USA. Tell him about the US so that you may help him decide whether he would like to come, etc (1961: 244; remember this is 1961!).*

The method of scoring advocated is that of rating scales, not of course as sophisticated as Bachman's or the ACTFL. In the rating scales he advocates, there are 'references to fluency, vocabulary, pronunciation and enunciation, and grammatical correctness which show that as soon as an examiner begins to listen carefully for speaking ability he notices differences in the various elements of speech which he can grade more accurately than the over-all

impression of speaking or the desirable but elusive criterion of intelligibility' (1961: 241).

As for his views on work sampling, he does say that the same techniques advocated for testing sentences and short texts 'can be used to test comprehension of a specific book or article' (1961: 238). But in general he is not in favour of work sampling:

> *... to approach the integrated skills in terms of situations rather than language brings in extraneous factors such as the selection of the subject matter, sampling of problems, what constitutes handling a situation etc and even if some of these factors can somehow be controlled we would find that the sampling of language problems would remain inadequate. It is more economical and will result in more complete sampling to work from the language problems and then to seek situations in which particular problems can be tested* (1961: 205).

This is not a world away from Bachman's desire to provide 'a description of language abilities and characteristics of language use tasks'(1990: 332). Except that what is called language in 1961 becomes language ability and language use tasks 30 years later. While I admire Bachman's vision, I have to say that I find him far more programmatic than Lado and at the same time envisioning the scope of language testing as ranging too widely across linguistics and applied linguistics:

> *the answers... are not to be found in further speculation and debate... [but]... in continued construct validation research. We need to employ both logical analysis and empirical investigation, including a variety of observation techniques, from ethnographic observation to controlled elicitation... only through such a broad-based program of research will we continue to expand our understanding of the factors that affect performance on language tests, and, more important, of the language abilities that are involved in communicative language use* (1990: 333).

8. A pragmatic approach

But can we, should we, wait on science in this way? Let me suggest a more pragmatic approach. This is by way of work-sample tests

which have been carefully constructed to test selected language problems (such indeed as the Proficiency Test for Language Teachers: Italian). There is no need to fall into the Bachman trap of confusion between ability and behaviour. Work sample tests meet the basic language testing requirements: they test language ability, in context and with adequate coverage. While avoiding 'real life', such tests have claims as performance tests. The problem for test construction is therefore practical rather than theoretical, a question of appropriate task sampling.

At one time, it would have been politically correct to denounce Lado as a behaviourist. Developments in cognitive science (notably in connectionism) make such a denouncement these days less of a knee-jerk. But denouncement or not, Lado certainly accepted a behavioural interpretation of language learning: 'the complex process of communication through language as we know it', he claimed, 'is made possible by the system of habits which operate largely without our awareness' (1961: 13). Not surprisingly, his view of language testing was of a piece: 'the theory of language testing assumes that language is a system of habits of communication' (1961: 22). These habits are based on the elements of language, separately and in combination. They are: 'sounds, intonation, stress, morphemes, words and arrangements of words having meanings that are linguistic and cultural' (1961: 25).

But in spite of his insistence on habits, Lado is also interested in what language is for, that is, performance (even though he does not use the term). He deals (in Lado 1961) with the questions of the integrated skills, of over-all control of the language and with how to test cross-cultural understanding. But his main concern is with the testing of the elements and with providing for adequate element sampling. And in my view he is right. Because of course this is precisely where both Bachman and McNamara end up.

Bachman, you will remember, distinguishes between the ability part of the IA approach and the interactional part: the ability part is to be activated through the 'skills and components frameworks of Lado' (1990: 317).

What the integrational aspect seems to offer is a means of contextualizing the Lado-type elements—but not too much, otherwise proficiency tips over into achievement. We have

suggested that Bachman's integrational effort leads to the provision of mini-performance levels, or even to the delineation of test items. As we have seen, Lado himself does this in different ways: nakedly, by testing the elements alone; through integrated tests of, for example, reading comprehension; and through the testing of the integrated, skills, the higher values and cross-cultural understanding.

Lado talks about language and situation: we can call these elements and work-samples (or tasks); and then he tells us how to measure control of these, through advice on items and judgements. For his part, Bachman has abilities and tasks (since his tests need content which should, as we have seen, be motivating, substantive etc) and attempts to bridge them with his integrational component: but this, I suggest is no more than a method of scoring—advice to judges, essential of course but at the same time equivalent to Lado's objective test provision. No doubt it is more obviously potentially communicative when it has to do with the oral component, which is the main example given in Bachman 1990.

Our conclusion must be that there is little difference between Lado and Bachman in terms of what to test; in my terminology, they are both agreed that it must be elements (accuracy) and work-samples (fluency). Where they differ is in how to test—for Bachman much more important, perhaps because undervalued: it is not without interest that an emphasis on the what leads to variety and LSP tests. Emphasis on the how leads to unitary and UG tests. But that apart, when we actually look at Bachman's exemplars of his integrational component, what we find is, as we have seen, not really a manual of how to do it, but rather a refining of the what into what we have called mini-performance levels.

McNamara has recently put his emphasis on the performance capacities involved in language use. Does this lead us into the same ball park as Bachman's quest for integrational component abilities? Well yes, it seems to me that it does, with the reservations I have expressed so far, that such a quest flags a search for the snark of a grammar of discourse or of communication.

Are the ideas of these three scholars so different? Yes and no. Yes, because where Lado takes as his concern language, McNamara has moved on to communication; where Lado is primarily interested in

the what of testing, Bachman concerns himself with the how. But in practice? In practice, they are all equally concerned with what language is made up of, whether they are called elements or capacities or component abilities.

The art of language testing comes from better descriptions of language features and from more valid sampling of critical contexts of language in use. Too much emphasis on features, in our terminology, accuracy, (possibly Lado's heresy) and we lose the performance correlative; too much on capacities, in our terminology, fluency, (where both Bachman and McNamara seem to be heading) and we lose coverage. Skehan (1993) has recently argued interestingly for a third approach, which he calls 'an information processing perspective to task-based testing... which will allow future target-language use situations to be analysed not only for the underlying language abilities that they require, but also for the performance conditions that are involved' (Skehan 1993: 20). I warm to this idea of a bridge but wonder how far it is possible to provide a taxonomy or a grammar of performance conditions. For there is, I suggest, no escape from the imperative to provide a systematic description of the elements underlying performance, whether we call them abilities, capacities or performance conditions.

9. Conclusion

Lado warns us against choosing situations first and then testing their language; better, he says. 'to work from the language problems and then seek situations in which particular problems can be tested'(1961: 205). In other words, select for accuracy and then determine fluency, the approach Rea Dickins (op cit) advocated 30 years later. For my part, a combination of elements (accuracy) and work-sample tasks (fluency) still seems to me a commonsense position to take up. So let's hear it! Two cheers for Lado!

10. References

Alderson J.C. and B. North (eds) 1991 *Language Testing in the 1990s: the communicative legacy.* London: Macmillan, Modern English Publications in association with the British Council: 136–149 (published simultaneously as Volume 1/1 of *Review of English Language Teaching*)

Bachman L. 1990 *Fundamental Considerations in Language Testing.* Oxford: Oxford University Press

Bialystok E. 1982 'On the relationship between knowing and using linguistic forms' *Applied Linguistics* 3: 181–206

Birdsong 1992 'Ultimate attainment in second language acquisition' *Language* 68: 706–755

Brumfit C. J. 1979 '"Communicative" language teaching: an educational perspective' in Brumfit C. J. and K. Johnson (eds) op cit: 183–191

Brumfit C. J. and K. Johnson (eds) 1979 *The Communicative Approach to Language Teaching.* Oxford: Oxford University Press

Carroll J. B. 1961 'Fundamental considerations in testing for English language proficiency' in *Testing the English Proficiency of Foreign Students.* Washington DC: Center for Applied Linguistics: 30–40

Coppieters R. 1987 'Competence differences between native and near-native speakers' *Language* 63: 544–573

Coulmas F. (ed.) 1981 *A Festschrift for Native Speaker.* Mouton: The Hague

Davies A. 1991 *The Native Speaker in Applied Linguistics.* Edinburgh: Edinburgh University Press

Davies A. 'What second language learners can tell us about the native speaker: identifying and describing exceptions.' *Melbourne Papers in Language Testing* 5,2: 1–27

de Jong, J.H.A.L. and L.W. van Ginkel 1992 'Dimensions in oral foreign language proficienc' in Verhoeven & de Jong (eds) op cit: 187–205

Elder, C. 1994 'Performance testing as benchmark for foreign language teacher education.' *Babel.* Journal of the Federation of Modern Language Teachers' Associations 29,2: 9–19. Also

published in *Melbourne Papers in Language Testing* (1994) 3,1: 1–26.

Fletcher, P. 1985 *A Child's Learning of English*. Oxford: Basil Blackwell

Johnson, K. 1996 *Language Teaching and Skill Learning*. Oxford: Basil Blackwell

Jones, R.L. 1979 'Performance testing of second language proficiency' in Briere E.J. and F.B. Hinofotis (eds) *Concepts in Language Testing: some recent studies*. Washington DC: TESOL: 50–57

Lado, R. 1961 *Language Testing*. London: Longmans

Long, M. 1990 'Maturational Constraints on Language Development' *SSLA* 12/3: 251–285

Lowenberg, P.H. 1995 'What do ESL tests test? Issues of cross-cultural norms'. In Gill, S.K. (ed) *Intelec '94: International and English Language Education Conference: National and International Challenges and Responses*. Kuala Lumpur Language Centre Universiti Kebangsaan Malaysia: 57–65

McNamara, T.F. 1990 *Assessing the Second Language Proficiency of Health Professionals*. Unpublished PhD thesis University of Melbourne

McNamara, T.F. 1994 'Models of performance in second language performance tests'. Paper read at LTRC Washington March 1994

Oller, J.W. Jr. 1979 *Language Tests at School: a pragmatic approach*. London: Longman

Rea-Dickins, P.M. 1991 'What makes a grammar test communicative?' In Alderson, J.C. and B. North (eds) op cit: 112–131

Richards, J.C., J, Platt & H. Platt 1992 *Longman Dictionary of Language Teaching and Applied Linguistics* 2nd edition. London: Longman

Skehan, P. 1989 *Individual   Differences   in   Second-Language Learning*. London: Edward Arnold

Skehan, P. 1993 'Task based language testing'. Paper given as plenary at LTRC Cambridge August 1993

Spolsky, B. 1989 *Conditions for Second Language Learning*. Oxford: Oxford University Press

Swain, M. 1985 'Large-scale communicative language testing: a case study'. In Lee Y.P., A.C.Y. Fok, R. Lord and G. Low (eds) *New Directions in Language Testing*. Oxford: Pergamon Press: 35–46

Verhoeven, L. and J.H.A.L. de Jong (eds) 1992 *The Construct of Language Proficiency*. Amsterdam: John Benjamins

Wesche, M. 1987 'Communicative testing in a second language'. *Canadian Modern Language Review* 37/3: 551–571