
Testing Communicative Language or Testing Language Communicatively: what? how?¹

Alan Davies

Abstract

Writing in *The Mind Machine* of the so-called critical or sensitive period in a child's development, Colin Blakemore comments 'after that sensitive period, even the learning of a second language becomes an intellectual labour' (Blakemore, 1984: 190). Such an unfashionable approach to second language learning has remained the unshakeable psycholinguistic view throughout the march of the communicative juggernaut. Now at last applied linguists and language teachers are acknowledging that communicative competence was always more a methodology, an approach, a scene setting (a 'how') than a theory or even a different text (a 'what'). Interaction, we are recognising, is not causative, not instrumental: putting learners in groups or pairs does not magically lead to second language learning any more than it leads to the growth of new knowledge. Without intellectual development communication tends to sad routinisation, in the same way as always playing tennis with the same partner does. While communicative language teaching may not have fulfilled its revolutionary hopes, it was a genuine attempt to make the best more widely available by (paradoxically) emphasising the authenticity of the process of learning as itself a real language engagement. Its problem was that it needed excellent teachers with something approaching native-like proficiency in the language they were teaching. Apart from the sheer difficulty of recruiting teachers with these qualities, there has always been the doubt about how to assess communicative language learning (or should it be teaching? Is this the reason for the uncertainty as to

¹This paper was written to give as a talk at the British Council Seminar 'Communicative Language Testing Revisited', held in Lancaster 3-14 September 1995. As is explained in the paper, I had given a talk at the 1985 Seminar on Communicative Language Testing at the same place and under the same auspices. The style of the written talk is deliberately retained in the paper.

whether we are testing communicative language or testing language communicatively?)

The paper will consider some currently available language tests, suggesting that in these the notion of 'communicative' has been interpreted as audience specific (eg occupation specific) and production oriented (that is oral and written). The paper will also consider testing as an on-going activity (eg classroom-based testing), arguing that the communicative 'how' is more easily implementable when the 'what' is carried by the teaching, that is where there is a close interaction between teaching and testing.

1. Ten Years Ago

My talk at the first seminar on Communicative Language Testing was on 9 September 1985, ten years ago almost to the day. My notes tell me that I approached the topic sceptically and used the following key-words as sign-posts:

Fashion, Halos, Power, Practicality, Speculation, Testing, Teaching. I then described what seemed to me legitimate about the concept of communicative language testing. Let me summarise what I said.

Fashion powered change is acceptable but it is always necessary to examine the practical implications of any change and not allow the change itself, or the hype associated with it, to create its own permanent halo. Is it a real change or only a talked-about change? There is the danger that a functional emphasis can lead to formal stagnation, eg that an emphasis on communication can lead to a decrease in variety and range. The reason for this is that boundaries are always needed and the functional ones may be more restricting than the formal ones (eg an RP accent, a particular writing style).

Power: communicative competence takes in so much that it ceases to be (a) attainable (b) describable (c) testable. In other words it no longer explains because it explains everything.

Practicality: the demand for communicative language teaching has inevitably fuelled a demand for communicative language testing: what needs to be remembered is that the practicality of

communicative language teaching (whereby the teacher can ad hoc) is not found in communicative language testing.

Testing: the teaching imperative is indeed to fit the testing to the teaching which explains the frequent demand to make the testing fit the teaching. But this is not necessarily the case for testing. What theoretical argument is there for saying that teaching and testing must be isomorphic? Even in terms of content validity it is always possible to relate the teaching and the testing at a more abstract level.

The legitimacy of communicative language testing (CLT): here are some of the apologies used to recommend CLT: performance based; continuous assessment; diagnostic; qualitative not quantitative; speaking emphasis; situation-based.

Performance-based? yes, this has been a welcome movement.

Continuous assessment: the interest in formative assessment is welcome as long as summative assessment is not neglected.

Diagnostic: it is unclear how CLT can be used diagnostically because of its lack of obvious generality.

Qualitative: the notion that the CLT criterion is criterion-referenced (while more traditional language testing is norm-referenced) collapses into the previous paragraph.

Speaking emphasis: the requirement of testing speaking is to provide quantitative measures (through grids etc), that is to colonise and survey and stake-out speaking and thus to diminish the importance of fluency.

Situation-based: this starts as language in situation but may become only situation. Our doubt here is that what is being tested is no longer language.

2. Today

Those were my views ten years ago. Where am I now in relation to this critique? Fashion, halos, power, practicality, teaching v testing: these all still strike home. What CLT has done is twofold.

It has legitimised performance testing and it has made us think more carefully of the context of the language under-test. Of these the former is the easier to implement (and to some extent covers the latter). Both are attempts to move slightly towards a more direct form of testing, making tests as authentic as is sensible.

3. The Wider Effect of CLT

What has been the wider effect of CLT? Let me use part of my review of Brindley, Geoff (ed.) (1995) *Language Assessment in Action*.

While communicative language teaching may not have fulfilled its revolutionary hopes, it was a genuine attempt to make the best more widely available by (paradoxically) emphasising the authenticity of the process of learning as itself a real language engagement. What communicative language teaching was really about was getting teachers and learners to reflect on their behaviour in class and to view it as itself a learning experience, like reading a novel or doing an experiment, like in fact any good parenting, or good kindergarten teaching. Its problem was that it needed excellent teachers (while traditional teaching could always get by with good lecturers) and these teachers needed to have something approaching native-like proficiency in the language they were teaching.

Apart from the sheer difficulty of recruiting teachers with these qualities, there has all along been the doubt about how to assess communicative language learning (or should it be 'teaching': while you may be confident that your teaching is communicative you have no control at all over whether the learning, if any, is also communicative). Indeed the doubt about assessment has been so strong that the issue of communicative language testing has been pushed deep down in the too-hard basket. Where indeed are the communicative language tests: TOEFL? (I)ELTS? the UCLES suite? The ACTFL, ILR and ASLPR? TOEFL added speaking and writing, ELTS with its early emphasis on specific purpose modules was revised in a reduced form for IELTS and has now been reduced even further. 'Communicative' has meant audience specific (occupational, such as tour-guides or health professionals, academic such as life science postgraduates) and production oriented (that is speaking and writing). In both cases this has led to the development

of performance testing, which at its crudest means texts rather than rules. CLT is not so different from old-fashioned work sample tests.

Reflection on engagement in the language class has encouraged creative thinking about observation and process; it has also encouraged on the job, performance based integration of teaching and assessment. The eleven case studies described in *Language Assessment in Action*, edited by Geoff Brindley, provide vivid evidence of that integration in a range of Australian ESL teaching encounters. Brindley in his own chapter on competency-based assessment makes the point that this integration can provide a focus on 'language as a tool for communication rather than on language knowledge as an end in itself' (Brindley 1995: 158). What he does not say is that communicative language assessment must, like all assessment, provide clear information about learning success. It is not evident that it does: indeed the information it does provide may confuse because of mismatches of teachers' and learners' expectations.

Brindley makes the following case for the volume:

'the last few years have...seen greatly increased activity in the development of tests and assessment procedures for assessing, monitoring and reporting learners' proficiency, progress and achievement in ESL programs. These range from large-scale proficiency tests and reporting systems to informal monitoring procedures aimed at assisting teachers to keep track of individual classroom learning. The aim of this volume is to bring together a range of these testing and assessment initiatives and to document the issues, problems and dilemmas, which arise as practitioners and language testers attempt to devise systems, instruments and procedures to meet their particular assessment needs.' (op cit 1995: 1)

There are, writes Brindley, relatively few case study accounts of the way in which assessment tools have been constructed to meet the needs of particular groups. He hopes that this volume will fill the gap by providing some insights into the rationales and decision-making processes which have accompanied the development of tests and assessments in both institutional and classroom contexts.' (op cit: 1).

Does it fill the gap? The first three chapters bring together assessment needs of large-scale systems and theoretical approaches: McDowell and McKay appealing to Bachman (1990) and to Bachman & Palmer (1996), Mincham to Halliday (1985). Then, Corbel and McIntyre discuss very different uses of the ASLPR, Corbel describing his development of Exrater, a computer program incorporating an 'expert system' aimed at assisting language assessors to apply the ASLPR, and McIntyre reporting a comprehensive review he carried out of the ASLPR. Given the exposure ASLPR has had in Australia it is useful to have two such critiques: the fact that they both question the ASLPR's claimed validity is interesting. Clarkson and Jensen's chapter reports on their experience of developing a task-based instrument for assessing achievement of objectives in an English for Professional Employment course for adult immigrants; Grierson and Gunn both consider criterion-based assessment procedures while Cram and Wilkes describe their very different experiences with self-assessment in language programmes.

Bachman's *Fundamental Considerations in Language Testing* (1990) acts as a start-up vademecum for most contributors to this volume: the book appears in nine of the individual lists of references. A relevant question of course is to what extent the authors' conclusions on their assessment experiences supports that early appeal to Bachman. There is little evidence either way. True, Mincham concludes that 'focus on a predetermined set of criteria helped (the teachers) in becoming more aware of learners' individual needs' (op cit: 87). But Gunn takes a contrary view: 'we discovered only through practice how difficult it is to specify criteria for task performance in a clear and unambiguous way' (op cit: 261). My own reading of these contributions does not accord with the strong theoretical tilt Brindley gives in his 'Introduction' (op cit: 8). Indeed, by drawing into relief 'the growing number of test development projects which draw explicitly on current theoretical frameworks of communicative language ability, in particular those proposed by Bachman (1990) and Bachman and Palmer (forthcoming)' (op cit: 8), Brindley draws attention to the lip-service paid by language testing practitioners to theoretical models

Such a forced yoking takes us no nearer the holy grail of true proficiency: what it tells us is that if you set up a model of language proficiency people will say they are following it, whether they are

or not. The case studies reported in the volume under review could have done with less applied model and more applied linguistics. Nevertheless, they stand in their own right: they provide interesting and thoughtful accounts of the realities of engagement in language teaching and testing,

4. Performance Testing

I want now to look more closely at performance testing. In the University of Melbourne Language Testing Research Centre (LTRC), the tests (of English and of other languages or LOTEs) have tended to be 'performance tests'; and I wish to argue that it's in performance testing that CLT has actually made most progress. But I think it needs to be put in a historical context.

Tim McNamara (1996) writes of two traditions of performance testing:

1. the work sample approach, and
2. a complex cognitive approach focussing on (a) the quality of the execution of the performance and (b) the underlying state of language knowledge.

I take a more radical view.

Bachman (1990) appear to reject performance tests in favour of 'authentic tests of communicative language ability' (CLA) while McNamara (1994) favours performance tests, helpfully distinguishing between weak and strong varieties. I discuss these now and then examine two earlier positions on performance testing, those of Lado (1961) and Davies (1965). My proposal is that far from being a recent development, performance tests have long been in use, for example in the 50s and 60s as integrative tests, often labelled work (or job) sample tests. If my argument is conceded, then we may conclude that McNamara's performance tests, Bachman's communicative language ability tests and Lado's integrative comprehension tests may all be regarded as special cases of work sample tests.

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 but of the authentic features which underlie such texts.

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. Bachman on Authentic Language Tests

Bachman begins his discussion on 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' (Bachman, 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, rather it turns 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. The 'ability' part of the IA approach', claims Bachman, goes back to Lado, Carroll and Oller, it 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' (ibid 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 thus 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''(ibid: 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 while construct validity is necessary it is 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 these levels:

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 levels, each containing a descriptor of a mini performance. Little 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' (ibid: 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.' (ibid: 312). This seems an odd reason given the need for all tests 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 tasks, 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' (Bachman, *op cit*: 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. McNamara on Performance Testing

McNamara (1990) considers performance testing of crucial importance in specific purpose testing. Drawing an illuminating 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 as doctors 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 whether 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 such, questions of coverage have to do with varieties of context: in a medical test for example with doctor-patient and doctor-colleague interaction; case conference; telephone communication and so on. Such work sampling is likely to take for granted coverage of the skills and components frameworks, ie 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, 1996) 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 of performance. What this implies 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 does not make it 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 it is in my view 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. Lado's Position

I turn now to Lado's consideration of performance. I have already quoted Bachman on Lado. He 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'' (Bachman, 1990: 302).

So what did Lado have to say about performance tests?

The answer seems to be nothing: I can find no reference to the term 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.' (Lado, 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/her conscious attention to the over-all meaning s/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' (Lado, 1961: 206). This is not I suggest just the skills and components framework which Bachman relegates to 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' (Lado, 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 but not two apples short of an integrational picnic: 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' (op cit: 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' (op cit: 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.' (op cit 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' (Bachman, 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.' (Bachman, 1990: 333)

8. A More 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. 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.

A proficiency test (EPTB) constructed in the 1960s (Davies, 1965) may be cited as an example of an integrative work sample test:

'The proficiency tester will...be as concerned with language problems as is Lado; but they will be uni-language problems, towards which he will make two approaches: in the first, the 'learning' approach, the linguistic categories and skills will be sampled and tested...in the second, the 'performance' approach, typical and necessary situations are constructed.' (Davies, 1965: 53).

A battery of tests was designed to assess the proficiency in English of overseas students intending to enter tertiary institutions in the UK. It was decided 'there should be two areas to draw on, one the linguistic aspect, the other one the language-at-work aspect' (op cit: 59).

'What...does an overseas student have to do with his English? The answer is, of course, exactly what a British student has to do with his English, he must listen to his lectures and his tutors (and understand them). The work sample area's contribution to the battery thus suggested itself readily: there should be tests which involved comprehension of typical lecture material and comprehension of typical text-book material. But what is 'typical'? '...Ideally the work sample tests in a proficiency battery would relate directly to the needs of each individual student...(however) precise work sampling for a proficiency battery may well be an

impossibility. And so the proficiency tester has to fall back on 'typical' material '(op cit: 64-6).

The work sample tests actually used in the battery were general, non-specific tests. Of these one of the tests of reading comprehension proved to be one of the most valid tests in the battery.

9. A Comparison

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' (Lado, 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' (op cit: 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' (op cit: 25).

But in spite of his insistence on habits, Lado is also interested in what language is for, ie 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 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' (Bachman, 1990: 317).

What the interactional aspect seems to offer is a means of contextualising 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 eg 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 interactional 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 such an approach is more likely to be successful in an oral test, 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 and work-samples. 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 Universal Grammar-influenced tests. But that apart, when we actually look at Bachman's exemplars of his interactional 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 dilemma as Bachman's quest for interactional 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 never-ending search for a grammar of discourse or of communication.

10. Need for Systematic Description

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 (possibly Lado's heresy) and we lose the performance correlative; too much on capacities (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.

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). A big ask! But not impossible. For my part a combination of elements and work sample tasks still seems to me a common sense position to take up. So: two cheers for Lado!

11. Conclusion

My title for this paper has been:

'Testing communicative language or testing language communicatively: what? how?'

So what are my conclusions?

1. We test communicative language: we do not test language communicatively. In other words it is language that is communicative not the tests.

2. We always have tested communicative language—in the form of interviews, role plays, essays etc. The recent emphasis on CLT comes from a greater concern with the spoken and written skills and with the contextualisation of language.

3. There is of course another view, which is that we should be testing (or we are testing) language communicatively, that is that our testing methodology should itself mirror some aspect of the performance we wish to capture.

This no doubt underlies McNamara's view that performance testing is a complex cognitive approach focussing on (a) the quality of the execution of the performance and (b) the underlying state of language knowledge. My own view is that while this is necessary for testing the productive skills, it is both unnecessary and may be distracting for the receptive ones where what matters is that the task itself should be performance oriented.

References

- Bachman L.F. (1990) *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Bachman L.F. & A.S. Palmer (1996) *Language Testing in Practice*. Oxford: Oxford University Press.
- Blakemore C. (1988) *The Mind Machine*. Penguin/BBC: London.
- Brindley Geoff (ed.) (1995) *Language Assessment in Action Research Series 8*, Sydney: NCELTR Macquarie University.
- 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.
- Davies A. (1965) *Proficiency in English as a Second Language* Unpublished PhD thesis. University of Birmingham.
- 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: Longman.
- McNamara T.F. (1990) *Assessing the Second Language Proficiency of Health Professionals* Unpublished PhD thesis. The University of Melbourne.
- McNamara T.F. (1994) 'Models of performance in second language performance tests' paper read at *Language Testing Research Colloquium* Washington, March.
- McNamara T.F. (1996) *Measuring Second Language Performance* London: Longman.
- Oller J.W.Jr. (1979) *Language Tests at School: a pragmatic approach* London: Longman.
- Skehan P. (1993) 'Task based language testing' paper given as plenary at *Language Testing Research Colloquium* Cambridge, August.
- 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.
- Wesche M. (1987) 'Communicative testing in a second language' *Canadian Modern Language Review* 37,3: 551-571

Tests referred to:

ACTFL: American Council for Foreign Languages Rating Scales

ASLPR: Australian Second Language Proficiency Ratings

IELTS: (International) English Language Testing System

ILR: Inter-Agency Language Round Table Rating Scales

TOEFL: Test of English as a Foreign Language

UCLES: University of Cambridge Local Examinations Syndicate