

ISSUES IN COMMUNICATIVE LANGUAGE ASSESSMENT

Richard Blight

ABSTRACT

Contemporary issues pertaining to communicative learning theory as practised in language assessment are considered and evaluated. Issues discussed include: the meaning and significance of the *communicative approach* and underlying *communicative principles*, concerns associated with incorporating *authenticity* into test design, the establishment of *performance criteria*, difficulties with *situation* and the processes of *measurement* and *extrapolation*, the representativeness of *real-life* encounters, the impact of *non-language variables* on test performance, and the complex and conflicting requirements of three critical goals - *validity*, *reliability*, and *efficiency*. The theoretical discussion is then considered in practical implementation through close analysis of a current in-service language test. While some difficulties are evident, particularly with regard to efficiently standardizing subjective evaluations, the test is found to be *authentic* and *communicative* in orientation.

COMMUNICATIVE LANGUAGE ASSESSMENT

Communicative language assessment is currently a primary goal of the assessment program in a majority of English Language Teaching (ELT) institutions. It is concerned with the testing of *communicative language*, and is best contrasted against forms of *non-communicative* testing which are also common in many ELT programs. But just what is meant by

communicative language? Since “communication” is generally regarded as the primary purpose for which we use language, it is at once hard to conceive of language being used for “non-communicative” purposes, particularly since even isolated individual learning activities can be viewed as contributing to an underlying communicative goal. So it is at once apparent, from a viewpoint of ordinary language use, that the logical delineation we are seeking is not straightforward.

It is helpful at this point to take an historical view, and to understand the way in which the term “communicative” has come to be used in a current language-learning context. Applied Linguistics theory describes *communicative language learning* as having evolved against the background of a number of other movements in language testing, each of which have been found to be limited by various shortcomings. The “other” movements are generally labelled as: “Grammatical” (Canale & Swain 1980: 2); “Structural” (Spolsky 1985: 181); “Psychometric-Structuralist” (Weir 1990: 2); “Psycholinguistic-Sociolinguistic” (Weir 1990: 3); “General Language Proficiency”; and what was at the time called “communicative” (Canale & Swain 1980: 2), but is now identified as the “Functional/Notional” (Spolsky 1985: 182) movement.

So the term *communicative language learning* is now used in comparison to a number of previous theoretical approaches, and signals a change of emphasis in underlying theory. We can hence infer that the current approach is more “communicative” than previous approaches, which were somewhat deficient in this regard. But to understand specifically what is meant by “communicative”, we need to review the earlier approaches and to consider their various shortcomings. The “Grammatical/Structural” approach was: “organized on the basis of...grammatical forms (i. e.

phonological forms, morphological forms, syntactic patterns, lexical items) and emphasizes the ways in which these forms may be combined to form grammatical sentences" (Canale & Swain 1980: 2). Spolsky further describes it as being based on the assumption that "knowledge of a language is best described... by describing its structures... [which is] a competence model, that is a claim dealing with underlying knowledge", with testing based on discrete point tests (1985: 181). Weir also discusses inherent shortcomings in the theory: "in terms of the construct validity... crucial properties of language are lost when its elements are separated... the whole is greater than the sum of its parts... testing a candidate's linguistic competence is a necessary, but not sufficient, component of a test battery" (1990: 2).

Weir describes the "psycholinguistic-sociolinguistic" movement as growing out of a need to overcome the "atomistic assumptions of the discrete point tests", and as being concerned with "global integrative tests such as cloze and dictation [which] went beyond the measurement of a limited part of language competence" (1990: 3). But while this direction provided a significant step forward in the testing of integrative language skills, it still did not adequately provide a sufficient model for language testing. Weir continues to discuss Oller's derivation of a "General Language Proficiency" model, criticizing that "his concept of overall proficiency has inevitably merged into a hypothesis of an underlying unitary competence", whereas, in fact "there is a growing body of evidence favouring a divisibility hypothesis" (1990: 5). Spolsky also discusses the "General Proficiency" issue, arguing that: "to say that linguistic and communicative competence are divisible does not necessarily rule out the claim that there is a core of common knowledge of a language underlying the specific

abilities of a speaker" (1985 : 185). The "General Proficiency" orientation does, I think, provide too broad a viewpoint to be useful for the development of linguistic theory, but one which is valuable to the needs of the more general community (eg. for providing standardized proficiency rating scales).

The "Notional/Functional" approach provides a significant step in the communicative direction, being based on communicative functions (e. g. apologizing, describing, inviting, promising), and emphasizing "the ways in which particular grammatical forms may be used to express these functions appropriately" (Canale & Swain 1980 : 2). Spolsky describes the underlying rationale : "the nature of language knowledge is best captured by detailing the various uses to which the language can be put : it aims therefore to list exhaustively the various possible functions of language, including all the notions that can be expressed in it" (1985 : 182). He also considers Canale & Swain's model of communicative competence, and discusses the need for a theoretical model to extend well beyond areas of grammatical competence to include other "rule-governed language systems... pragmatics, discourse rules, rules of sociolinguistic appropriateness, and rules for verbal and non-verbal conversational strategies" (1985 : 184). Spolsky's comments, and the early work of Canale & Swain, lead us directly into an understanding of the current *communicative approach*.

With the requirement for language teaching and testing methodologies to now cover an expanded domain, the *communicative approach* is based on a number of *communicative principles*. Weir (1990 : 7) discusses the significant requirement for testing to be performance based, so as to provide a more direct measure of communicative ability. Performance tests are regarded as demonstrating underlying knowledge of grammatical struc-

tures and integrative principles, and as well of the broader range of rule-governed systems newly incorporated into the model. Significance is also attached to language learning and testing being undertaken within an authentic environment, where *authenticity* is concerned with providing as realistic a context as possible. But there are still some serious problems associated with implementation of the communicative principles, and these shall be discussed.

Early communicative theory sought to identify those characteristics that can be associated with *typical* communicative situations. Lists of criteria for performance conditions (*specifications*) are then drawn up and used as the basis for a practical framework for implementation of language testing. Morrow, for example, discusses those "features of language use which do not seem to be measured in conventional tests" in terms of the requirement for language: to be interaction-based, to be unpredictable, to be contextually appropriate, to have a meaningful purpose, to demonstrate performance, to be authentic, and to satisfy behavioural outcomes (1981: 16). Canale (1983: 3) develops a theoretical framework for communication processes in terms of a requirement for: social interaction, unpredictability and creativity, discourse/socio-cultural contexts, limiting conditions (psychology, memory, fatigue, distractions), purpose, authenticity, and outcomes. Canale's framework, together with his differentiation of the components of communicative competence (grammatical, sociolinguistic, discourse, and strategic competence), built a foundation for current communicative theory, and was revised and extended in Bachman (1990) into a comprehensive model of communicative language ability.

Skehan discusses the "pivotal importance" of Bachman's model, commenting that it: "is both theoretically well-founded and comprehensive. . .

has been subject to empirical data validation. . . contains within itself, as it were, a concern for the competence-performance relationship” (1991 : 9). He comments on the value of the model at two critical levels, that of the “strong version. . . of proficiency - it attempts to tell it the way it is, by being as consistent as possible with modern linguistic theory and existing data”, and the “weak interpretation. . . the model provides a systematizing framework for the conduct of research in language testing which will increase the chances that testers will make cumulative progress in the future” (1991 : 14-16). In my opinion also, Bachman’s model of communicative language ability represents the most comprehensive and valuable theoretical approach yet developed to serve as a basis for communicative language testing.

But there are still a number of issues to be considered in relation to the value and significance of communicative testing, and these are of a complex and multi-dimensional nature. Central are concerns associated with incorporating *authenticity* into test design. Alderson discusses two key problems, the *nature* of testing, and the need for meaningful *extrapolation* :

any test is in danger of affecting performance if the testee is aware that he is being tested. To that extent, it is impossible for a test to be *authentic* in the sense of mirroring reality. Of course, tests are themselves authentic situations, and anything that happens in a testing situation, must be authentic in its own terms : the problem comes when one tries to relate the testing situation to some other communicative situation. . . the pursuit of authenticity in our language tests is the pursuit of a chimera : it is simply unobtainable because they are language tests (1981 : 57)

Bachman suggests two possible approaches in light of this difficulty. The *real-life* approach incorporates the highest level of authenticity, but is significantly limited by: "failure to distinguish ability from behaviour", and "inadequate basis for validation" (1990: 308-309). The second approach, that of *interactional ability (IA)*, provides a more useful framework: "the IA approach views authenticity as residing in the interaction between the test taker, the test task, and the testing context. The primary consideration, according to this view... is to construct or select test tasks that reflect our knowledge of the nature of language abilities and language use" (1990: 322). So that emphasis is returned to the requirement for carefully considered test design and construction in order to achieve meaningful results. In a later work, Bachman (1991) provides a useful *authenticity grid* that sets dimensions of *situational authenticity* vs. *interactional authenticity*, and this could be used to approximately map the quality of authenticity present in specific testing procedures. Of course, there remain also concerns with other significant areas, including the determination of communicative criteria and situational specifications, and the processes of measurement and extrapolation.

The question of *sampling* is also raised by Alderson (1981: 57-61), who is concerned with just how representative the design scenario is of the full range of possible actual scenarios. Also, even within the design scenario, how accurately can native speakers' behaviour be modelled for the purpose of testing? Alternatively, if the test occurs within a real-life context, how representative is the actual encounter of the range of possible encounters? So we return to the significant difficulties associated with the need to extrapolate from a narrow performance situation, since we ideally wish to measure communicative language ability within a broad range of situations.

Weir, like Bachman, responds to these issues by returning responsibility to the processes of test design and development which are specific to each test situation: “we also have to ensure that the sample of communicative language ability in our tests is as representative as possible. . . if we are to extrapolate from our test data and make statements about communicative language ability in real life situations, great care needs to be taken over the texts and tasks we employ in our tests” (1990 : 11).

Also significant is how to deal with those many non-language variables which may have a significant effect on test performance results. Such variables include personality, knowledge of subject matter, opinions, past education, and various environmental and psychological conditions (memory, fatigue, nervousness). Bachman provides a model for “explaining performance on language tests” which usefully allows for variation based on *personal characteristics*, *random factors*, and *test method facets*, and suggests a hypothesis for calculating the variance in language scores in terms of these variables (1990 : 350). Unfortunately, it is in practice somewhat difficult to accurately apply numerical quantities against variables such as *personal attributes*, so it’s difficult to see how this model could be implemented. Weir does not go into detail in this area, and simply comments that: “direct testing requires an integrated performance from the candidate involving communication under realistic linguistic, situational, cultural and affective constraints. Candidates have to perform both receptively and productively in relevant contexts” (1990 : 12).

There is also the ongoing requirement to balance the frequently opposing demands of *validity*, *reliability*, and *efficiency* in test situations. There are many complex issues to be considered in this regard. Morrow (1981 : 13) discusses the *reliability–validity tension*, and the need to provide for

validity in five key areas: face, content, predictive, concurrent, and construct validity. Weir suggests a two-stage approach which achieves an acceptable compromise between conflicting demands: "the first stage involves the development of a direct test that is maximally valid and reliable, and hence inefficient. The second stage calls for the development of efficient, hence indirect, tests of high validity" (1990: 18). He concludes:

retreat from direct evaluation of performance may be acceptable, provided relationships or even correlations between data from competence testing and predicted behaviour have been established... The crucial stage in any test development occurs when the specification is translated into a test realisation. The test that results should exhibit the qualities of validity, efficiency and reliability... these qualities need to be determined both qualitatively a priori and empirically a posteriori (1990: 21)

So that while there remain a number of concerns pertaining to issues in *communicative learning theory* and *communicative language testing*, it is valuable to provide a short summary of the current perspective on the *communicative approach*. The *communicative approach* is a social view of language learning which is primarily concerned with using language to communicate appropriately within various social contexts. It is neatly summarised by Burns & Joyce in terms of the following basic features: conceptualising language within social contexts (rather than as a system of grammatical patterns); a concern with both language form and function; selection of content based on student needs; a concern with the four macro

skills; tolerance of learner errors as an inevitable part of language learning; encouragement of learner independence; and finally: the view of the teacher as a facilitator of learning (1997: 44).

LANGUAGE TEST ANALYSIS

Theoretical issues discussed in the preceding section are now considered in their practical implementation through close analysis of a current in-service language test in terms of significant criteria. The test (see Appendix, below) to be considered in this section serves as an assessment task for intermediate level students working towards a *Certificate III in Spoken and Written English (CSWE III)*. The CSWE is a nationally accredited Certificate program developed and employed in Australia for English language students (see, for example: Wilson 1999) enrolled on the *Adult Migrant English Program (AMEP)*. The AMEP has been the primary English language instruction program delivered to newly arrived migrants to Australia from the mid-1970s to the present day (Manidis & Prescott 1994: 3), and itself applies according to, and is regulated by, various governmental immigration and education policies.

The CSWE is composed of a number of distinct *Competency* tests, each of which is designed to test a specific language skill. To achieve the Certificate, students need to successfully perform the Competency tasks according to the standards established in the specified *Performance Criteria*. In particular, students are required to achieve several competencies from each of the four skills areas, so that they need to demonstrate a sufficient level of speaking, listening, reading, and writing competency to be awarded the certificate. Teachers conducting the CSWE are all specifically

trained in administration of the various Performance Criteria according to national education policy, which is itself regulated through administration of the program funding to each language institution conducting CSWE training and testing. The current test serves as a sample task for achievement of one of the required speaking competencies on the CSWE III, which is listed as: "*Competency 5 : Can obtain information through a telephone enquiry*".

The test shall now be discussed according to a number of significant issues raised in the preceding section of this paper. Possibly the first question that should be asked about a language test is: *What is the purpose of the test?* The current test is an *Achievement Test* that aims to demonstrate satisfactory achievement of a task according to the test specifications. The specifications provide various criteria against which the performance of each candidate is measured. The CSWE III Manual provides test specifications for *Competency 5 : Can obtain information through a telephone enquiry (Community Access)* according to the following criteria: *Elements, Performance Criteria, Range Statements, Evidence Guide, and Phonology*, as follows:

ELEMENTS - Discourse Structure.

- i. can use appropriate strategies to negotiate a telephone enquiry
- ii. can request information
- iii. can provide information

ELEMENTS - Grammar and Vocabulary.

- iv. can use appropriate vocabulary
- v. can use appropriate grammatical structures

PERFORMANCE CRITERIA.

- i. uses appropriate strategies as required (e. g. opens, identifies self, states purpose, closes, seeks clarification, asks for repetition, gives feedback)
- ii. requests information as required
- iii. provides information as required
- iv. uses appropriate vocabulary
- v. uses grammatical structures appropriately (e. g. uses modals, interrogative structures)

RANGE STATEMENTS.

- familiar and personally relevant
- simulated or authentic
- telephone
- fluent interlocutor
- approximately 4 pieces of information
- whether authentic or simulated, request should be for information which is new to the learner
- may include a few grammatical errors, but errors do not interfere with meaning

EVIDENCE GUIDE - Sample Tasks.

Learners undertake role-play or authentic telephone enquiry to seek information on topic of interest/relevance, for example :

- inoculation of children
- local community services
- naturalisation procedures

- school enrolment
- community college information
- after-school care

PHONOLOGY.

Specific performance criteria related to phonology have not been included.

However, it is assumed that :

- articulation of some phonemes and clusters as well as intonation, stress, and rhythm of longer phrases and clauses may occasionally be inaccurate/unconventional
- teaching programs will pay attention to phonological features in longer utterances and to developing learner repertoires of self-correction and repair strategies

It is evident from the detailed specifications prescribed for satisfactory performance of the competency that the test developers had a certain range of acceptable performance standards in mind when they devised the test. However, my own personal experience, as well as the shared views of numerous teaching colleagues, in measuring student performance against the prescribed standards suggests that despite the level of detail provided, there remain significant areas of complexity and ambiguity which are not adequately covered by the test specifications. The test developers of course would maintain that this is the function of the mandatory ongoing test training program, to provide a forum for such issues to be discussed and mutually resolved by a team of trained professionals each having different levels of knowledge and experience concerning the application of specific areas of various Competencies. In any case, since performances on

the test are measured against the prescribed test standards, it follows that students can be specifically trained to perform against those standards, and therefore that the *purpose* of the test is to directly measure *achievement* of material taught and practised during class sessions. Hence the Certificate itself, wherein students must satisfactorily complete a specified number and combination of Competencies from within a broader range of possible Competencies, should be viewed not as a form of *Proficiency Test*, which would be: “based on a specification of what candidates have to be able to do in the language in order to be considered proficient. . . (where) ‘proficient’ means having sufficient command of the language for a particular purpose” (Hughes 1989: 9). Rather, the CSWE should be viewed as *demonstrated achievement* of a number of prescribed Competencies (or specific language skills), according to the requisite performance ranges specified by the test developers.

Another significant question to be asked of a language test is whether it involves *direct* or *indirect* testing, or a combination of both forms. This issue is important since while forms of indirect testing are generally more practicable from a perspective of classroom administration, the accuracy of indirect tests need to be considered carefully, particularly in relation to degrees of representativeness of the prescribed activities and standards of performance. The *Evidence Guide* for the current test prescribes that the test may be conducted either as a classroom role-play (a *simulated* performance), or as an actual telephone enquiry (an *authentic* performance). However, in both cases, since the candidate is required to directly obtain information on a telephone enquiry through an activity in which speaking performance is measured, it follows that the test is a *direct* test of language performance, rather than an *indirect* test of underlying skills. Hence it is

not necessary to be concerned with the relationship between the desired test goal and the actual test performance, as discussed by Hughes (1989 : 16), since there is a direct mapping between these variables.

The test is also *integrative*, rather than *discrete point*, since the test requires the candidate to combine many language elements in order to successfully accomplish the task. A discrete point test, in contrast, would specifically test discrete elements of language performance, item by item. In order to achieve this task, it is evident from the test specifications that a student would need to integrate several complex areas of performance, including: using appropriate negotiation strategies to undertake a telephone inquiry, requesting information in relation to specific questions, and providing information in response to interlocutor requests. And further, these discourse functions would need to be undertaken while employing appropriate vocabulary and grammatical structures. Notably, as specified in the Range Statements, grammatical errors are permissible on the test but essential performance criteria are evaluated in terms of whether (and to what degree) errors interfere with meaning in the actual communication process. So the student's integrated performance is being measured against specified criteria during the test.

The test needs also to be evaluated in terms of *objectivity* and *subjectivity*. Tests which are objective require no judgement on the part of the marker during the scoring process, while subjective tests require degrees of judgement to be employed during the grading process. Objective tests are *reliable* in their scoring, whereas the degree of *reliability* of subjective tests needs to be considered carefully. In the case of the current test, while detailed test specifications are provided with the intention of guiding the evaluation process, important judgements are still required from the

marker when evaluating student performances. So the test is clearly subjective in nature, and as a consequence it is necessary to consider the types of judgement called for and the degree of subjectivity employed in the test. Notably, the test specifications frequently employ the term: "appropriate" (*appropriate strategies, appropriate vocabulary, appropriate grammatical structures*), so that the marker needs to relate observed performances against a theoretical range of satisfactory performances. And the process of *relating* what is an *appropriate* standard of performance for the CSWE III level is itself a subjective and somewhat circular judgement; while test developers may maintain that the information provided in the Manual, combined with the ongoing mandatory training program, adequately establish a sufficient level of *inter-rater reliability*, it remains possible for a single performance to be judged differently by two markers, even in terms as significant as achievement or non-achievement of the competency. Moreover, the guidelines provided in the test specifications that are intended to facilitate the judgement process in fact tend to themselves draw on further subjective judgements. So the marker, while attempting to determine the appropriacy of the student's responses, is also called upon to determine the degree to which observed errors actually interfere with meaning, or to what degree the articulation or spoken rhythm is unconventional. So the inbuilt subjectivity of this test actually limits the success of the test itself. And although an adequate regulatory system has perhaps been developed to provide for the substantial judgement and reliability issues associated with this test, it can be argued strongly that there will be errors associated with processes of *relating* observed against theoretical performance, and of *interpreting* the desired standard. So it is arguable that the test can only be employed with satisfactory reliability by an

experienced practitioner who has specific professional training beyond the level of standard classroom teachers.

The test is *criterion-referenced* rather than *norm-referenced*, since performance is measured against specific criterion of what a candidate "is capable of doing in the language" (Hughes 1989: 17), rather than directly against the performance of other candidates. Indeed, the system of criterion referencing employed is a significant strength of the CSWE, since as a consequence, students are focussed on individual performance of set tasks, and further, it is possible for all the candidates taking the test to successfully achieve the competency. The detailed test specifications hence facilitate the progress of the students, since relevant performance criteria are clearly indicated, as well as the performance conditions which apply (*see Range Statements*, above). So in order for students to successfully achieve Competency 5, they are required: to use appropriate strategies for telephone calls and for making inquiries, to request and provide information, and to use appropriate vocabulary and grammatical structures.

Finally, there is the question about whether the test sets out to measure, and actually measures, *communicative abilities*, and hence whether the test can accurately be described as a *communicative language test*. This question can also be understood in terms of whether the test promotes *communicative language* use by undertaking *communicative language activities*. Firstly, it is apparent that the basic competency structure ("Can obtain information through a telephone inquiry") itself constitutes a communicative act that is commonly practised as a real-life activity. If the test is undertaken by having students directly make a phone call to a public service, then they are also instigating an *authentic* act of communication. If the test is instead *simulated* in a classroom environment (also acceptable,

see *Range Statements* above), authenticity is lost but the test would still be considered as communicative provided the simulation sets up a realistic communication process. This could be achieved, for example, by having students matched with native speakers who were instructed to respond appropriately to the student's language and also to behave in a manner befitting real-life performance of the task. And while asking native speakers to make judgements about "appropriate responses" and "expected behaviour" within a situational context is itself an unreliable practice, a realistic communication process would most probably ensue.

The intention of the test to measure communicative language ability is also evident in the communicative orientation promoted in the test specifications. The Performance Criteria deliberately indicate communicative stages necessary to ensure a communicative interaction: "opens, identifies self, states purpose, closes, seeks clarification, asks for repetition, gives feedback". So that successful performance of the criteria will necessarily correspond with a realistic communication process being undertaken. Hence it is clear that the test does measure "communicative language abilities", and that the test can appropriately be described as a "communicative language test". And since, as discussed in Hughes (1989: 19), measuring the ability to perform a realistic act of communication is central to the purpose for learning a language, the current test can be regarded as being valuable to the students' process of learning.

APPENDIX

SAMPLE LANGUAGE TEST. Please note the test provided below is a current in-service test performed at the Adelaide Institute of Technical and Further Education (English Language and Literacy Services).

Certificate III in Spoken and Written English

Community Access Program

Competency 5 : Can obtain information through a telephone enquiry

INSTRUCTIONS.

Look at the task below.

Prepare a set of questions to obtain the information you need.

Make the call and write down the answers.

Complete the self-assessment sheet.

TASK.

You would like to have some driving lessons.

Ring a Driving School (use the Yellow Pages) and enquire about :

- the cost
- the length of the lessons
- the type of cars available
- the test
- the length of time generally needed
- whether or not instruction in your first language is available

BIBLIOGRAPHY

- Alderson, J. C. 1981. Report of the discussion on communicative language testing. In J. C. Alderson & A. Hughes (eds.), *Issues in Language Testing*, 55-65. ELT Documents 111. London: British Council.
- Bachman, L. F. 1990. *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Bachman, L. F. 1991. What does language testing have to offer? *TESOL Quarterly*, 25(4), 671-704.
- Bachman, L. F. & A. D. Cohen (eds.). 1998. *Interfaces Between Second Language Acquisition and Language Testing Research*. Cambridge: Cambridge University Press.
- Bachman, L. F. & A. S. Palmer. 1996. *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.
- Brindley, G. 1989. *Assessing Achievement in the Learner-Centred Curriculum*. Sydney: NCELTR, Macquarie University.
- Brindley, G. 1991. Defining language ability: the criteria for criteria. In S. Avivan (ed.), *Developments in Language Testing*, 139-164. Singapore: Regional Language Centre.
- Brindley, G. (ed.). 1995. *Language Assessment in Action*. NCELTR Research Series (8). Sydney: NCELTR, Macquarie University.
- Brown, H. D. 1994. *Principles of Language Learning and Teaching*. Englewood Cliffs: Prentice Hall Regents.
- Burns, A. & H. Joyce. 1997. *Focus on Speaking*. Sydney: NCELTR, Macquarie University.
- Canale, M. 1983. From communicative competence to communicative language pedagogy. In J. C. Richards & R. Schmidt (eds.), *Language and Communication*, 2-47. London: Longman.
- Canale, M. & M. Swain. 1980. Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1-47.
- Cohen, A. D. 1994. *Assessing Language Ability in the Classroom*. Boston: Heinle & Heinle.
- Cumming, A. & R. Berwick (eds.). 1996. *Validation in Language Testing*. Clevedon: Multilingual Matters.
- Henning, G. 1987. *A Guide to Language Testing*. Cambridge, Massachusetts: Newbury House Publishers.

- Hughes, A. 1989. *Testing for Language Teachers*. Cambridge: Cambridge University Press.
- Manidis, M. & P. Prescott. 1994. *Assessing Oral Language Proficiency: a Handbook for Teachers in the Adult Migrant English Program*. Sydney: NCELTR, Macquarie University.
- Morrow, K. 1981. Communicative language testing: evolution or revolution? In J. C. Alderson & A. Hughes (eds.), *Issues in Language Testing*, 9-25. ELT Documents 111. London: British Council.
- Shohamy, E. 1997. Second Language Assessment. In G. R. Tucker & D. Corson, *Second Language Education*, 141-149. Encyclopedia of Language and Education, Vol 4. Dordrecht, Netherlands: Kluwer Academic Publishers.
- Skehan, P. 1991. Progress in language testing: the 1990s. In J. C. Alderson & B. North (eds.), *Language Testing in the 1990s*, 3-21. London: Macmillan.
- Spolsky, B. 1985. What does it mean to know how to use a language? An essay on the theoretical basis of language testing. *Language Testing*, 2(2), 180-191.
- Underhill, N. 1987. *Testing Spoken Language: A Handbook of Oral Testing Techniques*. Cambridge: Cambridge University Press.
- Van Lier, L. 1996. *Interaction in the Language Curriculum: Awareness, Autonomy & Authenticity*. New York: Longman.
- Weir, C. J. 1990. *Communicative Language Testing*. London: Prentice Hall.
- Weir, C. J. 1993. *Understanding and Developing Language Tests*. London: Prentice Hall.
- Widdowson, H. G. 1989. Knowledge of language and ability for use. *Applied Linguistics*, 10(2), 128-137.
- Wilson, L. 1999. *Certificate III in Spoken and Written English: Workbook*. Sydney: NSWAMES.