

## **The competency movement, Applied Linguistics and Language Testing: some reflections and suggestions for a possible research agenda**

**T. J. Quinn**  
**The University of Melbourne**

### **1. The “déjà vu” theme: perspectives on competency-based approaches from the point of view of the recent history of applied linguistics**

#### **1.1. Introduction**

The competency movement in Australian education circles, in the guise of many acronyms such as CBE (competency-based education), CBT (competency-based training), CBA (competency-based assessment) and CBET (competency-based education and training), is undoubtedly enjoying vogue status in the nineties<sup>1</sup>. Few people seem to be aware that the ideas associated with competency-based education are not new or recent: they have a relatively long (and chequered) history in American education, from which Australian educators could be expected to learn some salutary lessons and heed some warnings. Perhaps applied linguists should be the ones to voice these warnings. Scholars observing the competency movement from the perspectives of applied linguistics and language testing are likely to be aware of the pitfalls of a competency-based approach, since there have been several significant “moments” or movements in the recent history of language education that have foreshadowed important issues likely to surface in the evolution of competency-based approaches in Australia.

---

<sup>1</sup> In the interests of clarity, an attempt will be made to avoid acronyms in this paper.

Before addressing the specifics of this matter - the impact of competency-based approaches on language education in the recent past - it is appropriate to go a little further back and point to the origins of the competency movement in general American education.

1.2. The historical origins of competency-based education in the United States

Competency-based education in the United States had its roots in two movements that would attract few contemporary adherents<sup>2</sup>: behaviorist psychology and “scientific management” derived from time-and-motion studies. The early behaviorist psychologists - Thorndike (1913) and Watson (1924) - elaborated a theory of learning that focussed exclusively on what was overtly observable and measurable, to the exclusion of any supposedly non-observable matters such as “skill” or “knowledge”: only what could be specified in very exact behavioral terms could lay claim to educational and scientific legitimacy. This line of thinking reached something of a high point (or low point, depending on one’s point of view) when it was applied specifically to language and language learning by Skinner in one of his major works (Skinner 1957)<sup>3</sup>.

American educators who felt comfortable with behaviorist learning theories claimed to find support in what were called scientific approaches to industrial management: these were based on time-and-motion studies (principally in the steel industry and the automotive assembly factories): demonstrable efficiency, productivity, value for money and maximum worker output were the bywords of these approaches (known variously as Taylorism, Fordism or Americanism) and they appealed to curriculum writers who believed that such concepts were also fundamental to education. The leading promoter of these ideas as the basis for

---

<sup>2</sup> A valuable summary of these historical origins of the competency movement can be found in Tumposky 1984.

<sup>3</sup> It was, of course, Chomsky’s famous critical review of Skinner’s *Verbal Behavior* (Chomsky 1959) that could be said to have inaugurated the modern era in linguistics and language studies.

curriculum planning was Bobbitt (1924), whose emphasis on the measurement of defined performance standards and on long lists of specific skills and sub-skills would look very familiar to contemporary enthusiasts for competency-based training. In the field of teacher education, Ralph Tyler, a prominent faculty member of the College of Education at The Ohio State University in the thirties, was influential in imparting similar ideas to a generation of young teachers in training (Tyler 1931).

It is beyond the scope of this paper to trace in any detail the further development of these ideas and their impact on general education in the United States. Their influence has waxed and waned. In recent times, competency-related ideas seem to have been revived as part of the response of conservative governments (particularly the Reagan and Bush administrations) to situations like the unemployment crisis and the demands of economic rationalists for mandated accountability in education (U.S. Department of Labor, 1992). Of particular relevance to this paper is the appearance of these ideas specifically in refugee resettlement and education. Competency-based ESL in the context of refugee resettlement has been subjected to a thorough critique (albeit from a political rather than an applied linguistics perspective<sup>4</sup>) by such writers as Tollefson (1986, 1989 and 1990), Auerbach (1986, 1987), and Auerbach and Burgess (1985).

### 1.3. Competency-based education and the recent history of language education

What is of more direct interest for the purposes of the present paper is the applied linguistics perspective that was foreshadowed above, ie, the "moments" in the comparatively recent history of second language education when competency-related ideas have

---

<sup>4</sup> This is not to suggest that the political perspective is unimportant: in fact, as Pennycook (1989) convincingly argues, it may well be that a politically contextualised analysis of language teaching and learning is the crucial starting-point for understanding what really goes on in language programs; it is simply that this political perspective is not the central focus of the present paper.

had a direct impact on various aspects of language teaching and learning. One can point to at least three such moments or episodes, although the second is probably better analysed as an aspect of the first. The three to be considered are these:

- the audio-lingual movement
- the behavioral objectives movement in foreign language education
- the ESP (= English for Specific Purposes) movement.

The first two of these can be considered in a fairly summary fashion, since the same point relevant to competency-based approaches - that any attempt to “decompose” language skill into atomised elements is a distorting and trivialising process of reductionism - emerges from both. The same perspective is also evident in critical studies of ESP.

#### 1.3.1 The audio-lingual movement

Analyses of the failure of audio-lingualism are fairly commonplace (cf., for instance, chapter 2 of Ellis 1990). Despite the widespread official endorsement of audio-lingual methodology in the US during the sixties, its influence rapidly declined when the central concepts of the approach came under sustained attack from three quarters: from psychologists who questioned the behaviorist psychology (eg, Rivers 1964); from educational researchers who conducted empirical investigations of the effectiveness of audio-lingual instruction (Smith 1970 being the most notable of these); and, most importantly, from the mentalist perspective on human language capacities so cogently argued by Chomsky (1959).

For the purposes of the present paper, it is the last of these that is most relevant. The Chomskyan analysis of the nature of human language changed for ever and irrevocably our understanding of what it means to “know” a language: knowledge of a human language can never be adequately portrayed as the mastery of a fixed repertoire of utterances; creativity and the capacity for innovation are of the essence, and are always manifested in any

language learning situation, first or second. These are fundamental propositions that no serious student of language phenomena would question. In the light of these considerations, it is surprising to examine some of the language course materials derived from competency-based principles, and to see just how behaviorist they actually look. Competencies are expressed in lists of "elements" and "performance criteria" that belong to a view of language seen as a series of situation-based repertoires.

### 1.3.2 The remnants of audio-lingualism: performance objectives

Educational movements seldom die quickly or fade away quietly, even when their supposed theoretical underpinnings are shown to be baseless. So it was with the audio-lingual approach to second language teaching and learning. Under the "triple whammy" impact of Rivers, Chomsky and the Pennsylvania Project (Smith 1970), audio-lingualism might have been expected to disappear rapidly, but this was not so, and traces of the movement lived on in a movement in foreign language education that outlasted audio-lingualism: the specification of behavioral objectives (with accompanying discussions about suitable criterion-referenced tests to measure the attainment of the objectives) as part of individualised instruction. This was a phenomenon of the sixties and seventies. As with audiolingualism, it represented a "decomposing" view of language into supposed constituent parts. It has now largely disappeared from foreign language education, under the impact both of principled theoretical critiques and of rejection by practising teachers<sup>5</sup>. But it now seems to be resurfacing in competency-based language programs.

---

<sup>5</sup> The leading exponent of discrete point language testing and of "atomised" component teaching was Robert Lado (1961 and 1964). Behavioural objectives and individualised instruction were the subject of regular columns in the journal *Foreign Language Annals* in the early seventies. Typical statements of enthusiastic support can be found in Steiner 1970 and 1975, while the opposing (and ultimately dominant) view is expressed in writers like Grittner (1972) and Valdman (1975).

### 1.3.3 The ESP movement

It has been argued so far that the recent history of language education and applied linguistics raises issues that are relevant to an assessment of the contemporary competency movement. The argument becomes even clearer when one considers the case of English for Specific Purposes (ESP). The ESP movement can be seen as an early form of a competency-based orientation, driven by the same laudable objectives, but, like competency-based language programs, foundering on the quandary of just how specific and detailed one can be about language performance without distorting reality.

The ESP movement appears to have enjoyed exactly the same sort of powerful attraction that the competency movement does, as it draws on the dissatisfaction that committed teachers feel when they see no clear goals and achievements being incorporated into their language courses: surely, they believe, it is much better to specify exactly what you are going to achieve in a particular course, so that everyone knows what is going on and what is expected to happen; this must be better than the vague aimlessness that has sometimes tended to characterise "General English" courses.

Despite this immediate and undeniable appeal, the ESP movement has not had an easy path towards general acceptance in the ESL/EFL profession. Serious problems lurk beneath the surface appeal<sup>6</sup>. Not surprisingly, therefore, it has been the subject of two major critical analyses (Widdowson 1983 and Skehan 1984), both of which are relevant to the present focus on the competency movement. Both sound cautious notes of warning that need to be heard by people developing programs of competency-based education, training and assessment.

Both these critical analyses derive from the same central preoccupation with the fundamental view of human language

---

<sup>6</sup> A straightforward analysis of some of these problems (eg, the practical problems of "needs analysis," the neglect of curriculum and methodology in ESP courses, and the problems of teachers struggling with unfamiliar scientific and technical subject matter) is provided in Quinn 1985.

mastery and performance that underpins particular language programs. Both see a considerable danger in excessive emphasis on the "fixed repertoire" view of human language, or on the predictability of human language performance. Both consider this view a total distortion if it reduces a complex skill like language performance to nothing more than pre-rehearsed moves, and if it does not also allow for the essentially unpredictable and creative dimension that is an essential part of authentic human language performance. Thus, Widdowson, although he is undoubtedly influenced by the key ideas of the Chomskyan paradigm, adapts the Chomskyan terminology and redefines it in a fundamentally humanist way by using a crucial distinction between *competence* and *capacity*. Competence is the given of any language system, the underlying rules, conventions and structural patterns (including the socio-linguistic and communicative ones), to which any speaker of the language is in some sense bound to conform. But human language users do more than just conform to pre-existing rules of behavior, as if language use simply meant responding to linguistic and socio-linguistic control; they are, to a considerable extent, in control of their destiny, and exploit the rules at their disposal for their own ends. In addition to communicative competence, they also possess communicative *capacity*, which refers to the creative dimension, the ability to handle the unexpected and the unpredictable by building on and utilising the resources of competence; capacity is thus "... the ability to create meanings by exploiting the potential inherent in the language for continual modification in response to change" (Widdowson 1983: 8). It is this dimension of *capacity* that Widdowson sees as being at risk in ESP courses, or in other approaches such as the competency-based ones, that see human language performance as a well-defined domain of predictable tasks or even as "an infinite set of unique and widely varied speech events" (Bachman 1991: 690). The crucial question for competency-based language programs is thus whether the emphasis on training for the predictable and expected is so dominant that it distorts and precludes the *capacity* dimension. To the extent that such distortion occurs, Widdowson would see the approach as being *under-theorised*, ie, lacking an adequate theory of language and language use, a point that will be taken up again later in this paper.

The critical analysis of the ESP movement from the perspective of *language testing* in Skehan 1984 proceeds from a similar emphasis on the underlying assumed theory of language. Skehan, like Widdowson, rejects as inadequate the view of language performance as the mastery of a well-defined domain of tasks. To focus on these is to emphasise the most trivial aspects and to miss the important ones. Skehan develops this theme from the point of view of language testing: it is not possible to develop criterion-referenced tests for ESP programs because our theoretical tools are not sufficiently powerful to enable us to define the criteria without distortion of the essential character of human language performance; our attempts to understand even the most fundamental aspects of human language, language function, language acquisition, language behavior and language proficiency are still quite embryonic. If this is so, the likelihood of successful criterion-referenced language tests fades dramatically: "... any language performance that is worthy of interest will be complex and multidimensional. Because of this it will be impossible to state what the criterion is for any except a small number of very tightly-defined contexts" (Skehan 1984, p. 216). It is important to note that the language testing perspective highlights so dramatically the theoretical weakness of language programs derived from a "fixed repertoire" view of human language performance.

#### 1.3.4 Summary

The argument so far has been that the recent history of applied linguistics, second language education and language testing suggests lessons that may need to be learned by those who, under the influence of the competency approach, are designing competency-based language programs. It has been argued that one fundamental, recurring theme emerges from the collapse of behaviorist-based language teaching (ie, the audio-lingual movement) and its descendants (especially performance objectives), as well as from critical analyses of the ESP movement, and that recurring theme is this: a view of human language performance that emphasises only the "fixed repertoire" dimension is a distortion, and one that lends itself to a trivialising and reductionist process of focussing exclusively on what can readily be proposed as the constituent components of speech acts



and situations. It has been argued that language programs based on this sort of view are clearly under-theorised in the light of what we know about the nature of human language performance. We now proceed to try to relate this analysis to the situation of competency-based programs of education, training and assessment in contemporary Australia, and to consider how the research capacities of the NLLIA<sup>7</sup> - particularly in the area of language testing - might best be used as a means of helping the developers of competency-based language programs to avoid some of the potential deficiencies that are apparent.

It should be noted in passing that some supporters of competency-based approaches try specifically to eschew the behaviorist label with which they are readily tagged. Thus Thomson (1991):

Many of the criticisms of competency-based education and training seem to be founded in the belief that competency-based programs consist of interminable lists of skills to be mastered by the students or trainees. The critics apparently assume that competency-based programs are the old behavioral objectives programs revisited. The fact is that most people associated with competency-based training believe the behavioral objectives approach has been sufficiently discredited to make it an inappropriate model for their work.

The behaviorists require that all objectives of a program be prescribed and tested. The unfortunate consequence of this has been the breaking down of programs into more and more discreet (*sic*) skills and a consequent trivialisation of the educational or training program which attempts to deliver those skills. (p. 1)

There is, of course, no argument or evidence provided in this passage: it is simply an unsubstantiated assertion along the lines

---

<sup>7</sup> National Languages and Literacy Institute of Australia

of "You may think we look and act like behaviorists, but we're not really ..." But the assertion is grounded in an interesting paradox: the origins of the competency movement in modern industrial training springs from a rejection of the Taylorist and Fordist positions, as we saw earlier. Yet it may be argued that the movement, at least in its educational aspect, has not freed itself from its basis in behaviorism. That is, the question remains open: it is still up to the critical observer to assess whether or not competency-based programs, particularly language ones, fall into the same trap as the behaviorist approach of distorting and trivialising the nature of human language performance by excessive fragmentation into component skills.

## 2. Competency-based education in contemporary Australia - preliminary considerations

### 2.1. The political and social context

Two recent reports, the Finn Report and the Mayer Committee Report, have introduced a competency-based framework for curriculum, teaching and assessment in Australia. It is clear that the competency movement is a powerful force in the political and educational life of contemporary Australia, no doubt because the immediate attraction of its basic orientation is hard to resist: it looks so much like common sense<sup>8</sup>. It is hard to quibble about an emphasis on the *outcomes* of training and education, on what the learner can actually *do* as a result of the training program. Accountability, performance objectives, measurable outcomes and work-related skills are the coin of the realm in the education market-place. A competency-based approach that emphasises such considerations appears to be broadly endorsed by all Australian governments, both federal and State, as part of a broader economic reform agenda. The assumptions seem to be that

---

<sup>8</sup> Of course, as Tollefson points out, one can give a quite different interpretation to the appeal to common sense: "Assumptions that become widely accepted as common sense tend to sustain existing power relationships" (Tollefson 1990: 545).

- improving the skills and training of the workforce is a crucial step towards making Australian industry internationally competitive
- education and training systems should be responsive to economic and labour market needs and forces
- a competency-based approach to education and training will produce improvements of the kind demanded by industry
- language programs of all kinds - and specifically adult basic literacy and English training for migrants - are not fundamentally different from any other programs in training and education, and should therefore be subject to the same competency-based design imperatives.

Interestingly, this movement is inspired by an anti-Taylorist view of work. In this view, the features of productivity of the successful economies of Japan and Germany are contrasted with the Taylorist practices of the economies of the United Kingdom and the United States, which are in a position of competitive decline. The integration of conception and execution, a multi-skilled labour force, a focus on productivity, innovation and quality are the new watchwords. The seven Key Competencies<sup>9</sup> of Mayer are thus general, rather than specific as in the Taylorist tradition. In fact, they are hardly competencies at all, in the sense of discrete, specified skills, but what testing specialists would call rather general domains, areas within which specification will take place. The deliberate generality of the Key Competencies seems to have been misunderstood or overlooked in their application within particular contexts, and particularly within assessment practice, where older traditions of performance assessment seem to have taken over the intended purpose of the movement. In the Mayer Report itself, more specific competencies enter the discussion only in the discussion of performance levels and performance criteria, which begin to

---

<sup>9</sup> Collecting and analysing ideas and information; Expressing ideas and information; Planning and organising activities; Working with others and in teams; Solving problems; Using mathematics; Using Technology.

resemble the familiar skill statements of the previous atomistic tradition, but even here their relative generality is significant. The translation of the general framework within specific industry and educational contexts has led to on the one hand considerable frustration on the part of those charged with the task, because of the very generality of the Key Competencies, but on the other to a substantial distortion of the original intention of the statement of Key Competencies in that the necessary generality seems frequently to move out of focus.

## 2.2. The neglect of assessment issues

It is important to note, as a preliminary consideration, that the documentation related to competency-based education approaches often tends to leave assessment questions till last. There is often an indication that the testing and assessment aspects will be fixed up later. Thus, one of the most highly developed competency-based curriculum documents, the New South Wales AMES *Certificate in Spoken and Written English* (Manidis and Jones 1992), indicates that details of assessment measures are to be developed at a later stage:

A bank of assessment tasks is being developed to underpin the assessment of each competency ... A number of guidelines for assessment will be developed. The guidelines will contain the bank of assessment tasks for each competency ... The development of the standardised assessment tasks will further improve inter-rater reliability in the organisation<sup>10</sup>.

The information flier for an important competency-inspired project, the *National Framework of Adult English Language*,

---

<sup>10</sup> In the interests of fairness, it should be noted that the promised details on assessment matters were, indeed, at an advanced stage of preparation at the time of writing, and were due for publication early in 1993. In this respect, the New South Wales AMES was commendably different from several other similar curriculum development situations where the assessment details are promised, but never delivered.

*Literacy and Numeracy Competencies for Adult Basic Education and English as a Second Language Curriculum Purposes* also leaves assessment considerations for a later stage:

The Framework will provide clear criteria for assessment and accreditation implicit in the notion of competency and demonstrate an awareness of assessment consequences of competency descriptions and competency based curricula. Later stages of the project will involve addressing the specific issues of validating the assessments of these competencies ...

The Mayer documentation on key competencies is similarly light on details about assessment matters. The version of the *Report* being circulated late in 1992 (Australian Education Council 1992) has a section on "Assessment and Reporting" (pages 30-37) which is little more than a wish-list of desiderata, giving little information on crucial matters of detail. This is an aspect of the report that has also attracted unfavourable comment from the Australian Literacy Federation:

There is also a concern that such principles are of little relevance without details on implementation. The proposal offers no indication of how assessment will be implemented. What will it look like? What will the 'result' mean/look like? Will the results only be used for the purposes stated in the document? Will reporting on a criterion referenced test be satisfactory for employers and students alike? (Australian Literacy Foundation, nd, p. 6)

None of this is intended to suggest anything devious or conspiratorial on the part of the authors of any of these statements. In a sense, there is a certain logic in leaving the consideration of assessment questions until a late stage of the developmental process. On the other hand, anyone who approaches these questions from a measurement and assessment perspective - and particularly from a language testing perspective - knows that crucial issues of principle are often not addressed until assessment questions are faced; if this process is left until late, there is often a temptation to see earlier decisions as being

“set in concrete,” even though the measurement perspective would strongly suggest change.

### 2.3. An impoverished concept of validity

As a final preliminary consideration, it should be noted that many of the competency-based documents seem to espouse a quite impoverished version of ideas that are fundamental in the conceptual world of specialists in assessment and language testing. An instance of this is found in the so-called “validation” study of the Key Competencies of the Mayer Report. The conceptualisation of “validity” and the methodology used to establish it in the “Preliminary Industry Validation Report” and the “Project Brief - Validation of Key Competencies in Industry” are not such as to attract the admiration of anyone serious about these matters. “Industry validation” seems to mean finding out if a rather small group of people from a limited number of industrial contexts have any strong objections to the Key Competencies when they are explained to them, and whether they can suggest any improvements. The methodology comprised two components: discussion forums and structured interviews. There were, in all, three discussion forums (two in Victoria and one in New South Wales) with an unspecified number of participants. No information is provided about the number of structured interviews, nor about the number of participants. On the basis of this amount of empirical investigation, nine pages of “findings” are presented. The Project Brief document, circulated for tender purposes in mid-January 1993 for the project to be completed by 5 April 1993, suggests that this further “validation” study will be no less superficial and unconvincing than the preliminary study. Perusal of these documents suggests a quite trivialised conception of validation<sup>11</sup>. Perhaps the term ‘validation’ is being used rather loosely here, and it is inappropriate to insist on validation procedures familiar from good assessment practice when the focus of validation is a curriculum framework. However, the tender brief includes validation of performance

---

<sup>11</sup> A useful general and reasonably accessible discussion of validation in language testing contexts can be found in Bachman (1990).

levels (and, by implication, performance criteria), which is clearly an assessment matter. The Preliminary Industry Validation Report certainly includes extensive discussion of these matters. Interestingly, the greatest dissatisfaction with the Key Competencies Report is expressed in relation to the Performance Levels and their descriptors.

### 3. The language testing perspective

It is now appropriate to establish the context for a research agenda by setting out a series of issues or conceptual areas where it seems that the competency movement might benefit from interaction with scholars whose primary perspective is language testing. These are key ideas that underpin much contemporary research in language testing, or at least form a focus for research, but which seem to have had little impact on the thinking of those who are working on language programs conceived in a competency-based context, and the related assessment concerns. There is a certain amount of overlap between some of these issues, but they are treated separately to highlight particular points.

#### 3.1. The lack of a strong theory of language

In the remarks on ESP earlier in this paper, the views of Skehan (1984) were quoted, as an illustration of a view of language testing as restrained by the limitations of available general theories of human language and human language performance. In Skehan's view, language tests can only be as strong and effective as the language theory that underpins them; as we do not have a sufficiently strong theory to account for the complexity and "multidimensionality" of normal human language performance, then we do not have the means to develop effective and valid criterion-referenced tests of language performance for any but a very small number of contexts. It is the limitations of the linguistic theory that defines the limits of possibility of the language tests. In a similar vein, Bachman (1991) begins his analysis of language test development with an exposition of a

theory of *language ability*. Finally, Widdowson's 1983 treatment of ESP is principally about the general concepts of language and language performance that underpin the practice of ESP.

In the light of these facts, it is surprising to observe that the major statements of the competency movement and the documentation on performance assessment appear to be largely untouched by strong theories of language and language use (or, indeed, of *any* theories of language). To a large extent, the sorts of issues that would be "daily bread" for people working anywhere in the spectrum of applied linguistics receive scant attention in the vast literature of the competency movement. The analysis of a recent important conference will illustrate this. In December, 1992, NCVER (= National Centre for Vocational Education Research) held a prestigious 5-day international conference in Melbourne on the theme of "What future for technical and vocational education and training?" The conference attracted large numbers of both Australian and international participants. There were 57 conference papers (published in two volumes, one of 328 pages, the other of 274 pages) and five keynote addresses (published in a 67-page volume). Among all these words, there were but two papers that could be construed as having even a passing interest in applied linguistics matters: one was not available for publication, and the other consisted of one page. The concerns and expertise of applied linguists in Australia do not seem to have had an impact on the world of competency-based education.

To the extent that there is any theory of language evident in competency documents (as for instance in the NSW ESL courses of Manidis and Jones 1992), it appears to be largely the theory of systemic linguistics associated with the name of M. A. K. Halliday. (Indeed, the New South Wales documentation regularly repeats, in mantra-like fashion, that the material "... is based on a theory of language which systematically relates language to the contexts in which it is used": Manidis and Jones 1992, p. ii.).

The influence of Hallidayan linguistic theory probably has both positive and negative aspects. The attraction of such an approach is that it is very much a theory capable of *application*: Halliday is



a linguist for whom the term “applied linguistics” carries no negative connotations whatever; indeed, he sees the “applications” as being an essential part of the theory, and, throughout his career, has actively involved himself in the endeavours of his many students working to apply his ideas in various educational contexts.

On the other hand, there are probably three negative aspects. The actual *status* and *validity* of Hallidayan concepts still remain somewhat problematic: what we know about how language forms and social or work-related functions are linked is still somewhat embryonic. Systemic-functional grammar of the Hallidayan type can hardly be said to have passed into the linguistic mainstream: it is not widely taught outside a small number of centres (mainly in Australia) staffed by Halliday’s former students. This means that it has not been subjected to extensive discussion, challenge and peer evaluation by linguists of other persuasions. This also points to the second negative aspect: an approach as “applied” as this must ultimately prove its effectiveness in real-life application. One has to wait and see whether, in the view of practitioners like language teachers and ESL teachers, the approach actually *works*. In this regard, one can only say that it is still very early days, and that the jury is still out. Finally, the history of the development of the Hallidayan approach is relevant. In many ways, Halliday’s theories can be interpreted as a vigorous reaction against the arid formalism and anti-social individualism of Chomsky’s autonomous syntax. One might well say that Halliday’s greatest achievement was to change permanently the *agenda* of linguistic inquiry: as a result of his work, few linguistic scholars would now dare to ignore totally the social dimension of language use in a theory of human language. On the other hand, Hallidayan models may well ignore or under-estimate the essential Chomskyan insight: that human language knowledge and performance are essentially generative and creative, and not inherently constrained by the limitations of social convention.

This is very relevant to the competency debate, as will be further discussed below. It may well be that some aspects of human behavior, skills and knowledge are so conventionalised that they

lend themselves readily to competency descriptions and somewhat ritualised performance testing. But there is a strong case for saying that human language, because of its inherent creativity, cannot be so characterised without serious, reductionist distortion. Complex skills like language must be more than an amalgamation of purportedly separable parts. As in virtually all areas of competency analyses of complex skills, one needs strong evidence that the whole is nothing more than the sum of the parts.

### 3.2. Theories of language proficiency

If there is a gulf between strong linguistic theories and proponents of the competency approach, there is some evidence of an even greater gulf between the concerns of theorists of *language testing* and the preoccupations of *language teachers* engaged in competency-based assessment of the outcomes of language training. For the former, the very notion of “language proficiency” is problematic and elusive, and all tests of language proficiency have their considerable strengths and weaknesses. Language proficiency is certainly seen as multi-componential, and the results of language tests of any kind (performance, criterion-referenced, norm-referenced ...) are seen as highly sensitive to a wide range of factors such as test method, nature of task, culturally determined test taker characteristics, test taker strategies etc.

For the practitioners, on the other hand, it probably all seems pretty straightforward. The concept of “language proficiency” is not widely discussed, and test *scores* do not seem to be of much interest. What teachers (and presumably the funding authorities preoccupied with demonstrable accountability) want to know is whether learners, as a result of their language training, can *do* with language the specific things that are proposed as the goal competencies of the language course. It is hard to argue with this desire. Where the problems start, however, is with the apparently easy assumption that you can answer the key question (“Can the learner do with language the specific things set out in the competency descriptions?”) quite simply by setting up tasks or situations that appear to simulate the real-life competency

demands, and just deciding on the basis of your observation of the learner's performance whether or not s/he can satisfactorily do what s/he is supposed to be able to do. Certain concerns crucial to the language tester receive little attention: concerns about task equivalence and task-related variability, about rater training and moderation, about the validity and relevance of the detailed elements (both overt and covert) that teachers use in reaching their assessments, and about the evidence suggesting that *no* classroom-based simulation can accurately reflect real life - these are matters that language testing specialists see as urgent and crucial, but somehow that sense of urgency has not been conveyed to language teachers.

### 3.3. Competence and performance

The kind of distinction first made by Chomsky between language *competence* (underlying ability) and language *performance* (the overt manifestation of that ability) has been taken up by many subsequent writers, both in language-related fields and in other areas of human achievement. It was noted above that Widdowson (1983) uses the slightly different terminology of *competence* and *capacity*. Some writers in the competency movement make appeal to the distinction, because it raises a quite crucial problem about competency-based assessment, viz., "whether or not competence can be inferred from the performance evidence that has been collected" (Thomson 1991: 7).

This is probably the most fundamental problem that must be addressed in evaluating the viability of performance-based language assessments of the kind recommended by supporters of the competency approach. It has, of course, been widely discussed in academic treatments of performance testing, most notably in the work of Messick (Messick 1992). The issue is whether one is interested solely in the surface performance, or both in this and in the presumed underlying competence (knowledge and skill), and, if the latter, whether the surface performance is a clear indication of the underlying competence. Thus Thomson:

The decision to recognise a performance as satisfactory and infer competence is the basis for the success of the [competency-based] system. If these decisions are wrong, the system will almost certainly collapse ... Do we gain our evidence by assessing products or processes or the knowledge which underpins these? Do we want to assess all three? And how much evidence is needed? For example, if we choose to assess by observing somebody do something, then how many observations must we make to infer competence? ... Those who argue for more assessment of knowledge make the point that just because somebody can do something it doesn't mean they understand what they are doing ... Assessors charged with the responsibility of verifying that a person is competent must be sure (beyond reasonable doubt) that the necessary level of underpinning knowledge accompanies a performance (Thomson 1991: 7-8).

These are extremely serious issues in the context of assessing *language* competence. They are the issues that challenge the whole relevance of the competency-based training movement to language teaching. As language educators, is it language competence that we are interested in? Or only in language performance, what the learner can actually do with and through language when we ask him/her to do certain things? If we are interested in underlying competence, are we convinced that our performance tests do adequately reveal underlying skills and knowledge? Do our performance tests imply and promote a view of language learning as the learning of isolated building blocks and situation-specific behaviors (Auerbach 1986: 419), with the assumption that there is a well-defined domain of target language tasks outside the classroom, and that these tasks can be adequately sampled and simulated in a classroom assessment environment (Bachman 1991: 691)? or a view that real life is an infinite set of unique and widely varied speech events, that what matters most is the learner's underlying capacity to deal creatively with these, and that an indication of this underlying capacity may call for much more sophisticated and subtle assessment techniques than the mere observance of performance? The

\* \* \* \* \*

choices are stark, and even daunting, when one considers that we are dealing with issues that will determine the fundamental life chances of the language learners in our classes.

In the scholarly literature on testing, the issues adumbrated above would probably be referred to as the issues of *domain coverage* and *generalizability*, and the best recent discussion of these is in Messick 1992. He makes a valuable distinction between those situations (rare, in his view) where the evaluation of a performance is the actual *target* of the assessment (eg, in a ballroom dancing contest), and those situations where the evaluation of the performance is the *vehicle* of the assessment, ie, where you really want to assess the knowledge base underlying the performance: there is thus a fundamental question about whether we are assessing performance only, or some deeper, more abstract underlying capacity such as language competence or skill. The problem is to demonstrate convincingly that evaluation of performance effectively taps what really matters, which is the underlying capacity. Messick makes the point that, if the evaluation of performance features is what is stressed in the assessment, then "... inferences are not to be made about the competencies or other attributes of the performers, that is, inferences from observed behavior to constructs such as knowledge and skill underlying that behavior" (1992: 8). He goes on to summarise the basic dilemma:

In the performance assessment of competencies or other constructs - that is where the performance is the vehicle not the target of assessment - replicability and generalizability can no longer be ignored ... Problems arise when measurement practitioners try to have it both ways. That is, they focus on particular products or performances as if these were the targets of assessment, treating issues of domain coverage and generalizability with *belle indifférence* ... This might be defensible if the products or performances that are viewed as targets of assessment are actually targets of instruction. But we must ask ourselves how many educational objectives worthy of time and effort can be captured in a single task or a small set of tasks (or products or performances) (1992: 8-9).

It is probable that various areas of human learning, endeavour and achievement could be represented at various points on a continuum ranging from those where the judgment of competence on the basis of performance is straightforward and transparent, to those where the judgement of underlying knowledge and skill requires much more subtlety than the simple observation of performance. If there is any validity in this claim of such a continuum, then it would seem reasonable to assume that human language competence is at the "complex" end of the continuum. Messick would go on to claim that language assessments should therefore include far more subtle and varied approaches than have so far been considered within the competency movement. This point will be explored further below, in the discussion of validity matters.

### 3.4. Validity issues

There are two aspects of validity that seem to be neglected in discussions of competency-based assessment: the familiar concept of *construct* validity, and the somewhat less familiar idea of *consequential* validity. (Once again, Messick's 1992 paper is the best recent discussion of these.) The discussion of the latter will be considerably shorter, not because it is any less important, but simply because the issues are straightforward, which is not the case with issues of construct validity.

The establishment of construct validity seeks to demonstrate that the evaluation process is indeed assessing something (a construct) the existence of which best explains the data emerging from the testing process, ie, that such data are best explained by positing the existence of this entity or construct, rather than any other. Thus, in the New South Wales AMES documents, it is assumed that observing certain specific *elements* in a learner's language performance will reveal something about his/her language competence, knowledge and skill. The validity evidence seems to be the nature of the elements themselves, and the explicitness of the *criteria* which accompany the competency statements: they are assumed to be so explicit as to be transparent. Yet, from the point of view of construct validity, one must ask certain questions: what proves that these "elements" do,

indeed, constitute the defined competency, and that they are *exclusive*, ie, that they do not include elements that belong to other competencies, and *inclusive*, ie, that these are the only elements that constitute the defined competency, and that no other relevant elements have been omitted? The only validating evidence envisaged in such documentation seems to be some form of consensus among practising teachers and assessors. A measurement specialist would have grave doubts about this, and would seek much more convincing and technical evidence, along the lines suggested by Messick:

... although competence must be inferred from observations of performances or behaviors (or from their outcomes or products), this inference is not often straightforward, particularly inferences about lack of competence from poor performance. Indeed, this is the core problem of construct validity, namely, how to establish, via a theoretical integration of convergent and discriminant evidence, that an observed behavioral consistency (as well as relationships of that consistency to other variables) can be accounted for by a particular construct interpretation rather than by plausible rival interpretations ... (Messick 1992: 16)

The validity standard implicit in authenticity of assessment as a measurement concept is the familiar one of construct representation or minimal construct under representation. The validity standard implicit in directness of assessment is minimal construct-irrelevant method variance. Together, they signal the need for convergent and discriminant evidence that the test is neither unduly narrow because of missing construct variance nor unduly broad because of added method variance (Messick 1992: 37).

Messick uses this line of argument to call for much more subtle, sophisticated and elaborate forms of testing than the simple observance of behavior. The essential thrust of the argument centres around the concept of the *contextualisation* of the assessment task. A strong emphasis on contextualisation of task

and language is reminiscent of the basic Hallidayan position. Systemicists would probably approve of the Resnick and Resnick claim that "... knowledge and skill cannot be detached from their contexts of practice and use ... That means, in turn, that we cannot validly assess a competence in a context very different from the context in which it is practiced and used" (Resnick and Resnick 1991: 43). Messick, however, rejects this, as it "... appears to bring us once again to a behavioristic proliferation of skill constructs that are qualitatively different in different contexts of practice and use" (Messick 1992: 24). The alternative of *decontextualisation* (consisting of abstract knowledge-based tests like old-fashioned language tests) is rejected as lacking face validity, but an approach called *cross-contextual* measurement is proposed as the most satisfactory alternative. This is based on "multiple and varied contextualized tasks" (ibid., 25) and "modest correlations across task types or contexts (24). However, he voices some caution about "...how large a sample of varied tasks or contexts is needed to generalize with any confidence across genres or to a broader construct domain" (24), and goes on to argue that "... a multiplicity of problem contexts should be employed to facilitate cross-contextual assessment and the appraisal of generalizability or the lack thereof" (37).

The Messick approach to the construct validity of competency-based assessment constitutes a rich research agenda waiting to be applied to language assessments. It appears to be far more sophisticated and appealing than the fairly embryonic approaches to assessment that have been proposed so far in Australian competency-based documents.

The idea of *consequential validity* is also one that has been most eloquently stressed by Messick. In this regard, he is probably unusual among measurement specialists. The basic idea is one of social responsibility: the architects of any program of testing and assessment must accept responsibility for the way others use the testing approaches developed by the specialists. If their tests and assessments and evaluation approaches are such that they can be used by others as gate-keeping devices (eg, to exclude migrants from further education or re-training), even though this was not the intention of the developers, then the latter must still accept their fair share of responsibility. In other words, part of the

\* \* \* \* \*



assessment expert's responsibility is to foresee how the testing and assessment program might be used by others, and to build in measures that will forestall abuse and distortion. Messick's own words are once more worthy of extended quotation:

The consequential basis of test validity includes evidence and rationales for evaluating the intended and unintended consequences of test interpretation and use in both the short- and long-term. Particularly prominent is the evaluation of any adverse consequences for individuals and groups that are associated with bias in test scoring and interpretation or with unfairness in test use ...

... evidence of intended and unintended consequences of test interpretation and use should be evaluated as an integral part of the validation process ...

... some aspects of all testing, even performance testing, may have adverse educational consequences. And if both positive and negative aspects, whether intended or unintended, are not meaningfully addressed in the validation process, then the concept of validity loses its force as a social value (Messick 1992: 34-38).

Such a concept of consequential validity represents a far-reaching challenge for assessment experts to see an important and neglected dimension of their work: a realisation that language tests are about people and their life opportunities, and not just about data and statistical manipulations.

#### 4. Towards a research agenda of specific projects

##### 4.1. Preliminary

In the light of the analysis that has been presented so far, it is now possible to point to some major areas that appear to emerge as research questions for language testing specialists interested in offering their expertise to educators working in the competency movement, and to language educators in particular. Most of these potential topics are simply mentioned in the most general terms, as the details of specific research approaches is beyond the scope of this paper.

##### 4.2. Specific areas of research

There would appear to be four major areas: validity studies of several kinds; reliability studies; studies of test taker characteristics; and studies of “washback” effect.

###### 4.2.1 Validity

Probably the most important (and also the most difficult) kind of validity studies needing to be undertaken are those dealing with *construct validity*. To establish the reality of the language competency constructs being considered these would need to be quite technical studies designed along the lines indicated by Messick in an important passage already quoted above:

The validity standard implicit in authenticity of assessment as a measurement concept is the familiar one of construct representation or minimal construct under representation. The validity standard implicit in directness of assessment is minimal construct-irrelevant method variance. Together, they signal the need for convergent and discriminant evidence that the test is neither unduly narrow because of missing construct variance nor unduly broad because of added method variance (Messick 1992: 37).

In simpler terms, one needs to establish what the evidence is that particular tasks exemplify or instantiate or demonstrate particular elements of competencies, and what the evidence is that the stated elements adequately define the competence.

In addition, studies of *predictive validity* appear to be particularly important. Competency-based training programs, after all, make strong predictive claims, in that they purport to be able to predict the likely future language performance of the learner. These claims need to be tested in a series of well-designed follow-up studies.

Finally, in the light of the concept of *consequential validity*, descriptive studies need to be carried out to establish just how the results of performance testing are being used.

#### 4.2.2 Reliability

Many of the classical reliability approaches need to be applied to language assessments made as part of the competency-based approach. Thus, the reliability effects of such factors as the frequency of assessment, the length of the task assessment process, the range and number of tasks included, the possibility of establishing task equivalence, the number and role status (ie, participant in the interaction as well as assessor) of the assessors and similar factors all need researching. In addition, the whole area of rater training (frequency, type, effectiveness) throws up a complex of reliability questions. The topic of rater training assumes a central importance in the field of performance testing, and there are already suggestions that it is being neglected, as a recent competency writer points out:

Among the practices normally followed in education to increase reliability are:

- establishment of agreed-upon criteria;
- repeated measurement; and
- substantial training and practice in assessment for judges.

... The weakest of the areas is substantial training and practice in assessment. While descriptions of competency-based training all include reference to

appropriate training, it is this area that has (historically and across the whole range of Australian education and training) been weakest. It is here that those implementing competency-based training need to be most diligent and demanding to ensure that reasonable practices are followed (Foyster 1990: 44-45).

#### 4.2.3 Test taker effects

The topic of test taker characteristics and how these relate to test performance and the possibility of cultural, gender or class bias is currently a popular one in language testing literature (see Bachman 1991), and is very relevant to language performance tests in competency-based language programs. There is a need to address questions such as the following: how do various types of learners interact with different types of tasks? what kinds of strategies and processes do test takers employ in a performance assessment? do different strategies make a difference, and are they relevant to the performance assessment? in the case of unsuccessful performance, is it possible to judge whether the failure is due to lack of language skill or lack of metacognitive strategies, and, if so, what can be done about this? can test takers introspect and self-report on the strategies they use in tackling a performance assessment?

#### 4.2.4 Washback

It is generally assumed that language tests have an impact on language teaching<sup>12</sup>, and it needs to be established whether performance assessments have any negative (or, indeed, positive) effect on language curricula. It may well be the case that they have little effect on classroom practice. In principle, the performance specifications of competency-based language programs are largely descriptions of *goals*, and, as became very clear in the early days of the ESP movement, specifying a goal or destination tells you nothing about how to get there. The

---

<sup>12</sup> Bachman, however, citing Alderson, suggests that this assumption may be largely unfounded (Bachman 1991: 679).

specification of course goals in the early days of ESP had little effect on curriculum and methodology (see Quinn 1985 for an analysis of this phenomenon), and it will be important to establish whether the competency specifications in relevant language programs have any effect on what actually takes place in language classrooms. There is always the possibility that a principal classroom activity will be the simple rehearsing of the competency elements. This would be a far cry from good classroom practice, and could reduce language learning to an arid rehearsal of test performance. Such matters should be the subject of solid classroom-based research.

#### 5. What can language testing researchers contribute?

It seems fairly safe to assert that the whole field of assessment in the context of competency-based language training programs is massively under-researched. (Perhaps this is not surprising, given that the whole competency movement is relatively new.) There is a rich field of research awaiting cultivation by experts in the language testing area of applied linguistics. However, a note of considerable caution is in order. Many of the conceptual weaknesses and research questions discussed above may look highly esoteric to ESL teachers being carried along on the waves of official endorsement and enthusiasm accorded to competency-based training. There is a real danger of researchers being marginalised and their research being seen as self-indulgent nit-picking. Practising professionals in the language teaching field, like Candide, need to cultivate their garden, despite the metaphysical and philosophical agonisings of the language testing experts. The danger of being dismissed as irrelevant could be avoided by ensuring a central focus on three points in the "research message" conveyed by language testing experts to language teachers, and by working to promote one crucial professional orientation among language testing scholars. The central focus of the language testing research message should be on these matters:

- A message of caution on the use of tests and test results; a warning that these are not God-given certainties, but best

approximations that must therefore be used with great caution; a specific admonition of the dangers of basing irrevocable decisions about people's lives and futures on test results or performance assessments that are, at best, only partial revelations of the truth.

- A clear exposition of how the strategy of trade-offs works as a reasonable approach to good practice in language testing: a demonstration that repeated performance assessments increase reliability, and that therefore a result based on a single assessment (because there are not sufficient resources for multiple assessments) trades off reliability for economy; that regular rater training increases reliability, but that this may have to be traded off against making excessive demands on teachers' non-teaching time; that single task performance assessments are likely to have low validity, and that, if this must be sacrificed, perhaps it should be traded off against greater efforts to achieve high reliability.
- A commitment by language testing specialists to the development of sound, "best possible" and useable assessment instruments and strategies that will help practitioners meet the demands placed on them (by students, administrators, funding authorities etc.) without a total disregard for the complex realities that emerge in language testing theory.

Finally, it is suggested that a desirable fundamental professional orientation should be a commitment by the community of language testing scholars to give high priority to and make time for teacher development workshops and the in-service training of language teaching practitioners. This is neither an easy nor a particularly attractive stance, given the very different preoccupations of researchers and teachers, and the other demands made on the time of academic researchers, but it should be appropriate at least for those language testing scholars who work within the framework of the NLLIA. There is abundant expertise in the Institute relevant to the concerns of teachers working in competency-based frameworks, and they would benefit from clear and strong input by language testing scholars. But the input should come from testing professionals whose role

vis-à-vis the language teacher is that of interested and committed critical friend who takes the time to offer genuine help and constructive criticism.

## REFERENCES

- Auerbach, E. R. 1986. "Competency-based ESL." *TESOL Quarterly*, 20, 3.
- Auerbach, E. R. 1987. "The author responds." *TESOL Quarterly*, 21,3.
- Auerbach, E. R. and Burgess, D. 1985. "The hidden curriculum of survival ESL." *TESOL Quarterly*, 19, 3.
- Australian Education Council and Ministers for Vocational Education, Employment and Training. 1992. *Putting General Education to Work. The Key Competencies Report*.
- Australian Literacy Foundation. No date. "Response from the Australian Literacy Federation to the Mayer Committee *Employment-related Key Competencies: A Proposal for Consultation*." Unpublished mimeo.
- 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: 671-704.
- Bobbitt, F. 1924. *How to make a curriculum*. Boston: Houghton Mifflin.
- Chomsky, N. 1959. "Review of *Verbal Behavior* by B. F. Skinner." *Language*, 35: 26-58.
- Ellis, R. 1990. *Instructed second language acquisition*. Oxford: Basil Blackwell.
- Foyster, J. 1990. *Getting to Grips with Competency-Based Training and Assessment*. Adelaide: TAFE National Centre for Research and Development.
- Grittner, F. 1972. "Behavioral objectives, Skinnerian rats, and Trojan horses." *Foreign Language Annals*, 6: 52-60.
- Manidis, M. and Jones, M. 1992. *Certificate in Spoken and Written English - Description of Competencies: Stages 1, 2 and 3*. Sydney: NSW Adult Migrant English Service.
- Messick, Samuel. 1992. "The interplay of evidence and consequences in the validation of performance

- assessments." (Research Report RR-92-39.) Princeton, NJ: Educational Testing Service, July 1992.
- Pennycook, A. 1989. "The concept of method, interested knowledge, and the politics of language teaching." *TESOL Quarterly*, 23, 4: 589-618.
- Quinn, T. J. 1985. "Functional approaches in language pedagogy." In R. B. Kaplan, ed., *Annual review of Applied Linguistics Vol. 5*, 1984. London: Cambridge University Press.
- Resnick, L. B., and Resnick, D. P. 1991. "Assessing the thinking curriculum: New tools for educational reform." In B. R. Gifford and M. C. O'Connor, eds., *Changing assessments: Alternative views of aptitude, achievement and instruction*. Boston: Kluwer. Pp. 37-75.
- Rivers, W. 1964. *The psychologist and the foreign language teacher*. Chicago: Chicago University Press.
- Skehan, Peter. 1984. "Issues in the testing of English for specific purposes." *Language Testing*, 1, 2: 202-220.
- Skinner, B. F. 1957. *Verbal Behavior*. New York: Appleton-Century-Crofts.
- Smith, P., Jr. 1970. *A comparison of the cognitive and audiolingual approaches to foreign language instruction. The Pennsylvania Foreign Language Project*. Philadelphia: Center for Curriculum Development.
- Steiner, Florence. 1970. "Performance objectives in the teaching of foreign languages." *Foreign Language Annals*, 3, 4: 579-591.
- Steiner, Florence. 1975. *Performing with Objectives*. Rowley, Mass.: Newbury House.
- Thomson, P. 1991. *Competency-based training: Some development and assessment issues for policy makers*. Adelaide: TAFE National Centre for Research and Development.
- Thorndike, E. L. 1913. *Educational psychology: the psychology of learning*. New York: Teachers College.
- Tollefson, J. W. 1986. "Functional competencies in the U.S. refugee program: theoretical and practical problems." *TESOL Quarterly*, 20: 649-664.
- Tollefson, J. W. 1989. *Alien winds: the reeducation of America's Indochinese refugees*. New York: Praeger.



- 
- Tollefson, J. W. 1990. "The economics and ideology of overseas refugee education. (Response to Ranard and Gilzow.)" *TESOL Quarterly*, 24: 543-555.
- Tumposky, N. R. 1984. "Behavioral objectives, the cult of efficiency, and foreign language learning: are they compatible?" *TESOL Quarterly*, 18, 2: 295-310.
- Tyler, R. W. 1931. "A generalized technique for conducting achievement tests." *Educational Research Bulletin*, 8: 199-208.
- U.S. Department of Labor (1992) *Learning a living: a blueprint for high performance. A SCANS report for America 2000*. Washington, DC: Secretary's Commission on Achieving Necessary Skills.
- Valdman, Albert. 1975. "On the specification of performance objectives in individualised foreign language instruction." *Modern Language Journal*, 59: 353-360.
- Watson, J. B. 1924. *Behaviorism*. New York: Norton.
- Widdowson, H. G. 1983. *Language purpose and language use*. London: Oxford University Press.