

Testing the Untestable in Language Education

Amos Paran and Lies Sercu
(editors)

Multilingual Matters (2010)
ISBN: 978-1-84769-265-8

By Frank Monaghan

I must admit that when I first picked up this book with its 13 chapters and 266 pages on assessment, I wondered whether its title should have been *Reading the Unreadable*, but it really is true that you shouldn't judge a book by its cover.

The book explores four aspects of language assessment and asks each time whether students can or even should be tested on each dimension. The four areas are: intercultural competence, learner autonomy, literature, and language and content.

The book opens with a very helpful overview chapter by Amos Paran in which he reflects on the problem that the move towards standardisation, alignment and conformity in testing (which may be driven by a desire to ensure fairness) almost always comes 'at the expense of broadness of vision' (p. 3). An interesting example of this is Weir's account of the development of the Certificate of Proficiency in English over the course of the twentieth century as successive versions have focussed more and more narrowly on 'language' as opposed to language, literature and culture, when it began. Paran argues that as it is self-evident that a language classroom teaches about more than language only then it is wrong to attempt to 'isolate linguistic competence and test it without reference to other competencies and areas of knowledge.' (p. 5).

The remainder of the book consists of four sections, with three chapters on each of the four dimensions. The first chapter in each section offers a more general overview of the topic and the other two report on research-carried out in the domain. For reasons of space, I'm limiting my discussion here to the overview chapters.

Intercultural Competence

Chapter 2, *Assessing Intercultural Competence: More Questions than Answers*, by Lies Sercu, provides an overview of the construct of intercultural competence which has developed alongside a shift in education from an acquisition metaphor (where education is seen as the accretion of discrete – and therefore testable – chunks of knowledge) to a participation metaphor, (where education is seen as a means to participate in and contribute to society at large). As our societies are now characterised by increasing diversity so has intercultural competence taken on greater and greater significance. Sercu identifies the following personality traits as conducive to intercultural competence: empathy, respect, interest in cultures, flexibility, tolerance, open-mindedness, initiative, sociability and positive self-image, which are less easily testable than, say, the correct use of the simple past tense. Sercu goes on to pose the question as to whether any instrument exists for testing intercultural competence holistically and the answer is no, but there are various tests (used largely in the business world) that do attempt to test the knowledge dimension of intercultural competence (e.g. simulated interactions in specific cultural settings, describing a photo of a culture-specific situation, etc.). There are also, of course, ways of assessing attitudes using scaled responses to statements such as: 'I don't see why I should cover up my body when I travel to another country.' Sercu identifies a couple of key issues in this area, firstly, 'Does education want to be prescriptive about the intercultural attitudes learners should develop and can learners be punished for not having particular desired personality traits, such as 'interest in cultures' or 'positive self image' which have been identified as characteristic of the effective intercultural person?' (p. 28). Secondly, 'Another problem ... is the suggestion that training will culminate in full competence, which once acquired will remain acquired... a person might become less willing to empathise with or act tactfully towards members of particular cultures after many frustrating and disappointing experiences in communicating with them.' (p. 30). Despite the difficulties inherent in the endeavour, Sercu comes out boldly in favour of the development of 'a limited number of reliable and valid assessment tools for different age and ability groups as well as curricula, which can then exert mutual positive backwash effects from which teachers, learners and society will benefit.' (p. 31)

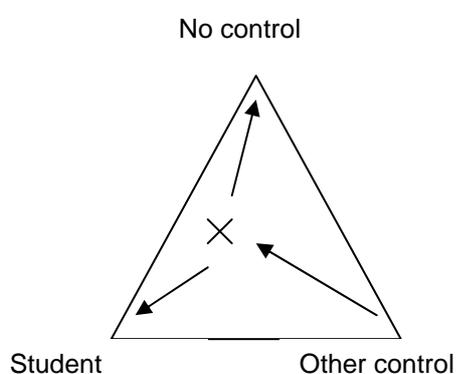
Our age has seen not only the proliferation of diversity, migration, and inter-connectedness, but also of colonialism, terrorism, and division. The

development of greater intercultural competence is not only an educationally and societally desirable goal, but increasingly a matter of survival.

Autonomy

Chapter 5, *Measuring Autonomy: Should We Put Our Ability to The Test?*, by Phil Benson, starts from a laudable wish to 'explore the risks that might be involved in delving too deeply into questions of the measurement of autonomy in educational climates in which the 'unmeasurable' often seems to lack value.' (p.78)

He begins with a discussion of how complex a construct autonomy is. Beginning with its origins in political philosophy, where it 'primarily refers to the self-determination of the affairs of individuals, groups and polities' (p. 78) Clearly, this is a wider definition than is likely to ever be realised in most classrooms (Summerhill aside) and what it more generally comes down to in the literature is the notion of students exerting some measure of control over their learning. Even this is circumspect as 'control' can mean many different things in terms of language learning – control over task selection, pace, degree of participation, length of time devoted to study, and some or all of these factors are very likely to be constrained by other people (teachers, parents) and other things (locations, text books, etc.) Benson concludes that learner autonomy can be recognised as 'behaviours in which control over one or more dimensions of the learning process is manifest' (p. 79) and he suggests that there are three poles of attraction between which it is possible to locate a learner's degree of autonomy along multiple dimensions: student control, other control, no control. He envisages a series of triangles such as the one below covering dimensions such as control over study planning, over attention input, etc.



A student might be observed to display behaviours that would locate them at different places within each triangle depending on the degree of control they exhibited

Measuring autonomy is recognised as being highly problematic, involving at various types of control: management of day-to-day learning, cognitive processes, and the actual content of what is being learned. (p. 82) There is then also the question of whether autonomy is better captured as a capacity rather than an observed set of behaviours – it is possible to appear to have control whilst not actually having it, and vice versa. He refers to the concept of the 'mask of autonomous behaviour' developed by Breen and Mann, in which learners can adopt a set of behaviours because they know this is what their teacher/examiner wants to see. A further issue is that as autonomy is a developmental process, there are likely to be steps forward followed by steps back along the way, i.e. it is not necessarily a smooth process and points of disruption to the flow, which may appear to suggest a lack of autonomy, may actually be evidence of the emergence of the next phase of it.

Benson goes on to look at a number of studies that have attempted to measure learner autonomy in language learning. Rowsell and Libben (1994) conducted a study at a Canadian university to identify differences between high and low achieving students (based on self-ratings) who kept diaries of their learning. The researchers examined the diaries for evidence of 'autonomously controlled tasks' (ACTs). These were of two types, *pedagogical* (e.g. repeating a task, adding a task, choosing not to do a task, etc.) and *functional* (trying to achieve a communicative use of language even when learning alone – e.g. by talking to a teddy bear in the target language) and found that higher achievers reported more 'functional' ACTs, than lower achievers. Simmons and Wheeler (1995) studied a group of migrant learners in Australia who were given control of what and how they studied and found that they were indeed able to take responsibility for their own learning and syllabus design and that whilst discussions were initially dominated by more fluent speakers, the less fluent gradually took on more decision-making roles as the course developed. In a third study, Rivers (2001) looked at self-directed learning behaviours amongst a group of US university students and again positive outcomes. Lai (2001) looked at how students both exhibited autonomy at the micro level of particular tasks but also at their overall sense of self-direction over the course of a term. This was based on an analysis of

journal entries. Perhaps the most interesting study, as most directly designed for its classroom application, was Sinclair's (1999), carried out at Temasek Polytechnic in Singapore. She has a distrust of judging autonomy on the basis of observable behaviours and so sought to address the problem by assessing metacognitive awareness, the 'capacity for making informed decisions about language learning' (p. 89). She poses the following questions (p. 89):

Can the student:

- Provide a rationale for her/his choice of learning activities and materials?
- Describe the strategies he/she used?
- Provide an evaluation of the strategies used?
- Identify his/her strengths and weaknesses?
- Describe his/her plans for learning?
- Describe alternative strategies that he/she could have used?

Sinclair then translates these into a series of questions that teachers can use as prompts for discussions with their students, e.g. 'What else could you have done/could you do?' She suggests three levels of awareness might be assessed based on such discussions, 'largely aware', 'becoming aware' and 'largely unaware'. She also suggests that the way students talk about their learning can also provide evidence about their differing levels of metacognitive awareness, 'anecdotal evidence, introspection, metaphor, epiphanies, questions and metalanguage, for example, appear to be characteristic of the 'becoming aware' stage, while the 'largely aware' stage is characterised mainly by description of alternatives.' (p. 89) Benson concludes that whilst it may be possible to assess aspects of autonomy, that does not mean it is necessarily desirable or efficacious to do so and may even be counter-productive in terms of the washback effect: 'we might ask what the likely consequences of 'autonomy testing' are. Students are naturally concerned with the consequences of passing and failing tests. Would they quickly learn what they are required to do in order to achieve high grades? Would their teachers find that they were no longer engaged in fostering autonomy but in helping their students put on Breen and Mann's 'mask of autonomous behaviour' or, perhaps, in devising ever more ingenious ways of forcing their students to remove the mask?' (p. 95)

Benson suggests that there is no need to assess student autonomy and that our interest in doing so is comparatively recent and arises from more general discourses that promote measuring anything that moves. He wonders if in even thinking of autonomy as a measurable construct we are simply internalising such discourses.

In Chapter 3, *Between Scylla and Charybdis: The Dilemmas of Testing Language and Literature*, Amos Paran explores the uneasy choices that face teachers of literature when deciding whether to assess their students' skills in literature or language. He sets out six dilemmas for testing literature in foreign language teaching:

Dilemma 1: To test or not to test

He begins with the inherent contradiction between the goals of testing (external, gate-keeping, public accountability functions) and those of teaching literature (internal, personal development, private appreciation).and points out, 'Once we introduce testing into the equation, it becomes apparent that the values implicit in the act of teaching literature on the one hand, and those values implicitly inculcated into our pupils through the examination system on the other, are in fact at odds with one another.' (p. 145)

Dilemma 2; Testing language or testing literature?

Whilst the chapter is written from the perspective of foreign language teaching, it is interesting to consider whether teachers of literature in mainstream English classes have given much thought to the idea that their EAL students are also learning the *language* through it and explicitly exploit it with that purpose in mind. This is, of course, true for any subject and is also applicable to the L1 English speakers – are chemistry teachers teaching science or the language of science? However, there is a particular tension that Paran sees between the aesthetic, affective, goals of teaching literature that might be less clearly evident to other subjects (though it would be surely possible to make a similar case for music and, more abstractly, perhaps, mathematics). Notwithstanding the somewhat elitist positioning of literature, the point remains valid – is the aim to assess the subject and its concepts or the language through which it is mediated?

Dilemma 3: Testing Knowledge or testing Skills?

In this section Paran discusses whether we are teaching and testing knowledge *of* and *about* literature or whether the interest is in developing

literary competence. If we are interested in the former, then students can easily ‘get away’ without actually reading literary texts themselves; they can read cribs and summaries if all they are going to be asked about, for example, is for a description of a character and their development. This has been described as the ‘flight from the text’ (Short and Candlin, 1986: cited here page 150)

Dilemma 4: testing private appreciation of literature or testing public knowledge about/of literature?

Going on from the previous section, this dilemma relates to whether students are being taught *efferent* reading, i.e. where the focus is on public knowledge, shared meanings (from the Latin *effere* – to carry away) or *aesthetic* reading, where the focus is on the learner’s individual and private experience and pleasure. He cites, for example, the York Notes on *Pride and Prejudice*, which include a section on ‘How to study a novel’, which recommends, ‘start by reading quickly for pleasure, then read it slowly and carefully’ as if a slow and careful reading were a chore and not a source of pleasure in itself! And I guess we have all met English Literature graduates who bemoan the fact that their studies have rendered them incapable of reading solely for pleasure. The simple fact of the matter is that most tests will opt for the *efferent* because that is easy to devise questions and marking schemes for. How would you devise criteria for personal response?

Dilemma 5: Authentic/genuine tasks or pedagogic tasks?

When we read outside the classroom we generally respond to it by talking to others about what we have read, recommending the book or explaining why we enjoyed it, or didn’t. Should we be building more of this type of assessment into classroom discussions and writing about literature?

Dilemma 6: should we require metalanguage?

Paran immediately confesses that this isn’t a real dilemma at all as the answer is an unequivocal ‘yes’ as ‘if we are teaching skills or encouraging a response, then we need also to teach the vocabulary to talk about that response.’ (p. 153) This does then also imply that the use of metalanguage is also assessed, or at least included in the marking criteria.

In response to the dilemmas he comes up with a series of principles for testing and assessment (pp 153-156):

- Include both public knowledge and private appreciation
- Use a variety of tasks
- Include choice
- Provide texts
- Include portfolio assessment
- Ensure that any criteria are transparent
- Minimise the weighting of language
- Construct multi-part tests

He concludes by reminding us that a major reason we teach literature is because it is a good thing in and of itself, ‘because it will enrich our students’ lives; because an appreciation of literature is an appreciation of many of the spiritual and intellectual aspects of life’ (p. 161) and so if we encourage an assessment of it that leads students and their teachers to ‘flee from the text’ then we are likely to alienate students from it. He argues that, to prevent this negative washback effect, ‘it is important to collect cumulative evidence of the process of ongoing engagement with literature rather than collecting summative evidence of knowledge about literature’. As he pointedly states, when you run into ex-students they rarely say things like ‘I still remember that lesson on embedded clauses’.

Language and content

In Chapter 11, *Assessing Language and Content: A functional Perspective*, Mohan, Leung and Slater discuss the integrated assessment of language and content (IALC) in a second language context.

They point out that ‘in an increasing number of education systems, an integrated approach to language and content *instruction* for second language learners is mandated policy. However, in a striking inconsistency, policy for integrated language and content *assessment* is essentially absent.’ (p. 217)

They point out how teachers often struggle with IALC, typically awarding some marks for language and some for content separately, as if the language used did not, to a very large degree, determine the meanings students are able to construct.

They propose a functional approach to IALC based on Hallidayan systemic functional grammar, arguing that, ‘If language is the primary evidence for

learning, then assessment is primarily assessment of text or discourse, and of how wording constitutes meaning in text.’ (p. 221). They cite a number of researchers who have analysed the registers of subjects and devised ‘knowledge paths’ that chart the development of them through increasingly complex conceptual hierarchies. Veel, for example has done this for science, Coffin for history and Christie and Derewianka have produced a cross-curricular model in their *School Discourse* (2008). We have the knowledge!

Assessment, they argue, should be linguistically principled, ‘it should be explicit about the language resources learners need to perform tasks in different disciplines; and it should provide specific criteria that recognize the difference between different tasks.’ (p 224) They recognize the importance of establishing why systemic functional linguistics (SFL) is different from other ‘grammars’ in its underlying assumptions and why it, therefore, offers the most appropriate tool for the job (Fig 11.3, p. 229):

Systemic Functional Linguistics	Traditional grammar
Language as a resource for making meaning	Language as a set of rules
Language form related to meaning	Form unrelated to meaning
Text makes meaning using language resources in context	Written or spoken text as a display of sentence grammar forms
Relates language system to both text and values	Values language rules (competence) rather than text (performance)
Language learning as extending resources for making meaning in context	Language learning as acquiring correct forms
Evaluate text as making meaning with resources in context	Evaluate correctness of form; judge meaning independently from form

They draw on formative assessment to provide a basis for teachers to ‘take the initiative when larger scale classroom assessment has been found wanting’ and make a strong recommendation that

teachers be given the resources to pursue it in the context of an SFL informed approach.

SFL has underpinned the work done by LILAC and was (unacknowledged) evident in the former National Strategies for literacy and the ill-fated *Assessing Pupil Progress* (APP) that has been driven underground with the arrival of the 1950s educational philosophy of the current political regime. The fact that, despite promises to publish it, the work that was done on APP for EAL students by the now defunct QCA still remains hidden in a DfE vault somewhere is an indicator that this approach is not currently welcome by the powers that be, but as with the LINC materials of the 1980s, teachers’ good sense about what works will always triumph over politicians notions of what they delude themselves to be true. Perhaps, as I speak, a samizdat DVD is being produced...

SFL has often been regarded as ‘difficult’ but as more and more guides to it become available, this will surely pass. The evidence anyway, as Mohan, Leung and Slater argue, is that many teachers already operate a form of ‘intuitive functional formal assessment’ (p. 237) simply because they know that the meaning and the words are inextricably linked and should not be assessed and addressed separately.

I have only been able to review a third of the book here, in effect, but I hope I have set out some of its principle concerns and arguments and can only commend the rest of it to the intrepid reader.

It is, ultimately, the final section on content and language that will be of most interest to those involved in EAL of course, and interesting that this is the one that seems most positive and confident not only about our ability to ‘test the untestable’ but also about the fact that we should be aiming to do so and suggesting a method (SFL) for doing so.

