Testing the Untestable in Language Education

Edited by
Amos Paran and Lies Sercu

Testing the Untestable in Language Education

Edited by
Amos Paran and Lies Sercu

Testing the Untestable in Language Education

Contributors

Phil Benson is a Professor in the English Department at the Hong Kong Institute of Education. He has published widely on the subject of autonomy, including the book *Teaching and Researching Autonomy in Language Learning* (Pearson, 2001) and a 'State-of-the-Art' paper in *Language Teaching* (2007). His current research interests include language learning histories, popular culture and language learning, and language teacher education.

Leni Dam took the first steps towards developing learner autonomy in language teaching and learning in her own classes at secondary school level in Denmark in the early seventies. Even then, a strong focus was students' evaluation of their own learning. She was later educational adviser at University College, Copenhagen, which provided excellent opportunities for implementing the principles of learner autonomy as well as evaluation into teacher education. From 1993 to 1999, she was co-convenor of the scientific commission of learner autonomy within AILA, and also carried out the research project, Language Acquisition in an Autonomous Learning Environment, together with Lienhard Legenhausen. She has published widely and given numerous lectures and workshops around the world. In 2004, she was awarded an honorary doctor's degree in pedagogy from Karlstad University, Sweden. In 2007, she stepped down from official duties and is now freelance.

Kaisu Korhonen holds a PhD degree in intercultural communication and a Master's degree in business English. She has a long experience in working as principal lecturer at a university of applied sciences in Finland. She has worked with engineering and business students as well as with professionals in various fields, providing intercultural communication and ESP training. She has published two study books for engineering students and a number of papers for conference and professional publications, discussing intercultural competence, its development with

the aid of intercultural communication training, and the effectiveness of such training. Her current interests include the Finnish style of communication.

Terry Lamb works in the School of Education, University of Sheffield. He has authored and edited numerous publications in the areas of learner autonomy, multilingualism and teacher development, and is editor of the *International Journal of Innovation in Language Learning and Teaching*. He has presented keynote papers in conferences around the world. Terry is former President of the Association for Language Learning and current President of FIPLV (Fédération Internationale des Professeurs de Langues Vivantes). In 2008, he was appointed Chair of the Development Partnership for the Diploma in Languages and International Communication by the Secretary of State for Education. He is also Chevalier des Palmes Académiques, an honour awarded by the French Government.

Lienhard Legenhausen received his PhD from Stuttgart University in 1974. He wrote his thesis on *Error Analysis and Error Evaluation* (Berlin: Cornelsen, 1975). A year later, he became a Professor of English and Applied Linguistics at Düsseldorf University. In 1992, he accepted an offer from Münster University where he was responsible for Second Language Acquisition Research and Language Pedagogy. His main research interests include technology-enhanced language learning and the analysis of learner languages. Together with Leni Dam, he started the research project, Language Acquisition in an Autonomous Learning Environment (LAALE). Since 2006, Lienhard Legenhausen has been an Emeritus Professor.

Constant Leung is Professor of Educational Linguistics at King's College London. He is Chair of the MA English Language Teaching and Applied Linguistics programme, and Director of MA Assessment in Education programme in the Department of Education and Professional Studies. His research interests include education in ethnically and linguistically diverse societies, language assessment, second/additional language curriculum development, language policy and teacher professional development. He has written and published widely on issues related to ethnic minority education, additional/second language curriculum, and language assessment nationally and internationally.

Anthony J. Liddicoat is Professor in Applied Linguistics at the Research Centre for Languages and Cultures in the School of Communication, International Studies and Languages at the University of South Australia. His research interests include: language and intercultural issues in education, conversation analysis, and language policy and planning. In recent years, his research has focussed on issues relating to the teaching and learning of culture through language study and his work has contributed to the development of Intercultural Language Teaching and Learning.

Hui-wei Lin received his PhD from University of Nottingham in 2005. He is now Director of the Centre for Learning and Teaching at Chia Nan University of Pharmacy and Science. A specific focus of his current research is how literature renders qualia through stylistic patterning. His research seeks to explore the possibilities for the assessment of learner achievement in literary education, with a concomitant purpose of initiating innovation of the traditional paradigm of literature teaching and literature testing in the ESL/EFL classroom.

Marylin Low is motivated by her concerns about the complexities of teaching and learning (in) multiple languages and literacies, especially in the diverse and rich cultural contexts of Pacific small island nations. Assessment-as-teaching opens to an abundance of knowing that teachers work hard to restrict in their attempt to 'get it right'. Her interest is in breaking open such restrictions and getting in on the real, living conversations that constitute teaching/learning/assessment in everyday classroom life. She is a Senior Specialist at Pacific Resources for Education and Learning, a non-profit organisation that serves the US affiliated states of the Pacific.

Bernard Mohan is Emeritus Professor, Language and Literacy Education Department, University of British Columbia, and Research Fellow, Department of Education and Professional Studies, King's College London. He has published widely on language as a medium of learning and the integration of language and content learning in second language learning, drawing on the theory and analysis of systemic functional linguistics. His longstanding research interest in linguistic aspects of assessment developed when he was a member of the Division of Measurement and Experimental Design, Department of Educational Psychology, University of Calgary. He is currently Chair, North American Systemic Functional Linguistics Association.

Angela Scarino is an Associate Professor of Applied Linguistics and the Director of the Research Centre for Languages and Cultures at the University of South Australia. She has had extensive experience in the conceptualisation, design, implementation and evaluation of curriculum and assessment frameworks for languages learning across all levels of education. She has conducted research, development and evaluation projects in the areas of languages curriculum, assessment and teacher development. Her experience includes working in various states and nationally in Australia, Hong Kong, Singapore, Malaysia and New Zealand.

Tammy Slater is an Assistant Professor of Applied Linguistics/TESL in the English Department at Iowa State University of Science and Technology. Her research, which draws upon analytic methods from systemic functional linguistics, seeks to understand the development of academic language through content-based and project-based teaching and learning, particularly as it informs the field of education for English language learners.

Jane Spiro is Principal Lecturer running International and TESOL MA programmes at Oxford Brookes University. She has previously worked at the College of St. Mark and St. John, Pécs University, Hungary, and Nottingham University, and has been consultant worldwide on projects in materials writing, test writing, curriculum development and creative writing. She was co-editor of the journal *Reading in a Foreign Language*, and judge for the first *Poetry in a Foreign Language* competition. She is the author of *Creative Poetry Writing* (Oxford University Press, 2004) and *Storybuilding* (Oxford University Press, 2007) and is a published poet and novelist with a research/teaching commitment to the connections between creative practitioner and educator. Her PhD on this subject was completed in 2008 and published in full online by the University of Bath.

Contents

Chapter 1

More than Language: The Additional Faces of Testing and Assessment in Language Learning and Teaching

AMOS PARAN

Testing and Assessment: A Ubiquitous Phenomenon

'Teaching involves assessment' (Rea-Dickins, 2004: 249). This simple, three-word sentence hides what is in fact a whole world, a world where 'young people in many countries... are now faced with an unprecedented number of exams and tests as they go through school and higher education' (Broadfoot, 2005: 125). It is a world which has been developed into a testing society (Broadfoot, 2005), where standardised testing is a major part of the assessment regime, which in some countries, e.g. the UK, starts as early as the age of seven (for an overview, see Leung & Scott, 2009).

Unsurprisingly, such a ubiquitous phenomenon as testing exerts an extremely powerful influence on its environment; it is now recognised that tests have powerful washback effects, what Cheng and Curtis (2004: 7) call, 'a set of relationships, planned and unplanned, positive and negative, between teaching and testing'. These effects extend throughout the educational system and, indeed, throughout society, becoming, as Shohamy (2007: 120) has argued, *de facto* instruments of language policy: 'since tests are often more powerful than any written policy document, they lead to the elimination and suppression of certain languages in societies... Tests can also be used as tools to privilege certain forms and levels of language knowledge... Thus, language tests, given their power and influence in societies, play a major role in the implementation and introduction of language policies'.

In language teaching, the field of testing and assessment has an additional effect: a number of important models of language competence,

1

such as Bachman's (1990) model or the Common European Framework of Reference (CEFR) (Council of Europe, 2001) originate in the need to specify language competences for testing and, in the case of the CEFR, finding a way 'to compare the objective and achievement standards of learners in different national (and local) contexts' (Morrow, 2004: 6; see also Alderson, 2004). Thus, our view of language learning and language competence is strongly influenced by our understanding of language testing and assessment.

An additional issue is reflected in the title of this book: the assumption that everything we teach in the language classroom can in fact be tested. Within language education, since often more is taught than only language (see below for a discussion of this point), the case can be made that not only language should be tested.

This volume brings together 12 chapters in which educators from around the globe grapple with issues that arise from these points. In this introductory chapter, I start by looking at some of the recent critiques of policy and practice in language testing, and at some of the responses to the current situation. I then present the four areas in language education on which the present volume focuses, and provide an overview of the different chapters. I end with a discussion of the themes emerging from the different chapters in the book.

Language Tests: A Narrowing Agenda?

In spite of the ubiquity of testing, there is nevertheless 'a widespread perception that the needs of teachers and learners are not currently well served by assessment practice and by assessment research' (McNamara, 2001: 340). The reason for this becomes clear when we consider the way language testing has developed over the last half-century. Spolsky (2008) charts the major trends in language testing, highlighting the dominance of the psychometric approach and the industrialisation of tests. McNamara and Roever (2006: 1) suggest that in language testing, 'psychometrics became the substrate discipline... and language was virtually poured into these preexisting psychometric forms'. Leung and Lewkowicz (2006: 212) voice a similar view, suggesting that 'the form that has been most prevalent in ELT all over the world in the past 50 or more years has been standardised, psychometrically oriented testing'. Leung and Lewkowicz (2006) attribute the tendency towards standardisation in a wish (or requirement) for fairness, but make it clear that a commitment to standardisation will come at the expense of acknowledging differences between test takers.

The drive for achieving standardisation, alignment and conformity has another important consequence: it will almost always come at the expense of broadness of vision. Wall (1997) provides a history of the worry about the narrowing of education as the result of tests, tracing it back to the beginning of the 19th century. In language testing, this trajectory in the history of tests is exemplified in Weir's (2003) fascinating account of the development of the Certificate of Proficiency in English (CPE) during the 20th century. What emerges from Weir's account is how the examination, through its numerous revisions, increasingly focused on language and on language only. With each revision and expansion of the construct of language proficiency, the examination shed aspects that did not reflect this construct, reflecting 'a gradual but critical change of the examination to one of language as against language, literature and culture' (Weir, 2003: 18). This process may well be underway in other countries as well – Eckes *et al.* (2005: 373) seem to imply a criticism of language tests that 'assess far more than language proficiency proper (e.g. they also tap knowledge of German literature, history, and civilisation)'; elsewhere they mention the importance of increasing reliability in the marking of essays on a literary theme in Slovenia. It seems logical to assume that this will be accompanied by a narrowing of the scope of the examination. Thus, 'the psychometric, the industrial and the scaling trends' (Spolsky, 2008: 450) continue to dominate, and they often entail a more focused – and hence narrower – approach to what is tested and how this is done. This development receives an added dimension in Broadfoot's (2005: 135) discussion of 'the inseparability of the affective and cognitive domains in learning', and the resulting implications for testing.

Together, the points that have been presented here – the ubiquity of tests, their powerful washback effect, which can ultimately dictate what is taught in the language classroom, and the narrowing vision of language tests – have meant that one consequence of large-scale language tests has been to circumscribe the content that is taught in the language classroom. The way in which language tests have developed has meant that language education, in many countries, is now concerned with teaching only language. True, developments in our view of language, of communication and of language learning mean that we have a more nuanced and more complex view of the language classroom than it would have been half a century ago. McNamara and Roever (2006), for example, provide an extensive discussion of assessing second language (L2) pragmatics, and Spolsky (2008: 450) does note 'the broadening of the content to include sociolinguistically influenced

aspects of language'. However, the focus is still on language, with an artificial separation between language and content.

At the same time, our understanding of language, language teaching and language learning has moved away from this separation, and includes resistance to a reductionist view of language education where all that is taught is what can be easily tested. This volume is therefore an attempt to bring together and record the endeavours of language educators in the global arena (the 15 contributors to this volume work in 10 different countries) to incorporate assessment and integrate it into four different areas of language and language development: intercultural competence; autonomy; literature; and content teaching.

Current Critiques of Language Testing

For some time now, views of language testing have recognised many of the problems that beset this field. There has been a focus on the social consequences of tests as well as on their use as instruments of power and control, as the passage from Shohamy (2007) quoted above suggests. The critique of language testing has focused on critical views of the uses of testing (Shohamy, 1996, 2005, 2007); there has been a move to use-oriented testing (Shohamy, 2005) and critiques of many of the uses of language tests (McNamara, 2005; McNamara & Roever, 2006). This is linked to ethical issues in testing (see, e.g. Lynch, 2001, for a discussion of ethical issues in different approaches to language testing).

One approach has been alternative assessment (Huerta-Macías, 2002; Fox, 2008), sometimes conceptualised as 'alternatives in assessment' (Brown & Hudson, 1998). Indeed, Shohamy (1996: 144) has suggested that we are now in an 'alternative era'. Birenbaum (1996) views alternative assessment as being grounded in an alternative approach to instruction, and as integrating assessment and instruction. She then lists a number of what she calls 'alternative assessment devices', such as 'authentic performance tasks, simulations, portfolios, journals' and many others (Birenbaum, 1996: 8). There is much talk of the 'ethical dimension in so far as (testing) affects people's lives' (Weir, 2005: 1).

However, the critique of testing and of language tests normally focuses on the teaching and learning activities that are being affected, rather than the content of what is being taught. Shohamy (1996: 150) points out that even Bachman's (1990) elaborate model of communicative competence 'does not account for the domain knowledge in performance testing'. In Weir's (2005: 212) discussion of washback, it is clear that what is at stake (and to some extent, rightly so) is success on the test, and the

washback effects considered are, for example, 'training teachers in the new content and methodology required for the test' and 'support in the forms of appropriate teaching materials'. Even where washback is conceived in a wider context than the traditional sense, what Weir (2005: 213) calls 'the wider impact of the test, its effects on other systems in the administrative and academic contexts of the tests, and on the attitudes and behaviour of the stakeholders in these', we are still talking about the *test* as the main factor.

More than Language

This volume is a response to what I have described above, and presents the reactions of educators to the challenges that testing raises, through a focus on additional aspects of language testing, either the content of what is being tested, or additional skills that are related to language learning. The educators in the chapters that follow all realise that in the language classroom, more can and should be taught than language. What is taught in addition to language can either be conceptualised as content (e.g. literature or science) or as transferable skill (e.g. intercultural competence, creative writing, literary competence, autonomy). What this volume implicitly does is to call for an end, at least in some contexts, to the attempt to isolate linguistic competence and test it without reference to other competencies and other areas of knowledge.

Cumming (2009: 91) discusses three approaches to the question of 'how integrally language tests, curricula and pedagogical practices should be aligned and what benefits or consequences may arise when they are'. One of the approaches he discusses is 'to adapt assessment policies and practices for particular populations, such as test accommodations for certain learner groups or setting performance standards for occupational purposes' (Cumming, 2009: 91). The chapters in this collection may be interpreted as taking this approach, in that each section looks at the way in which testing practices have been developed and adapted, though not for particular populations, but for particular modifications and viewpoints of the curriculum/construct.

However, it is not only in terms of format and procedures that alternative assessment is important. Fox (2008: 97) points out that 'alternative assessment represents a conception of language that is diametrically opposed to that of traditional tests'. In terms of this volume, this opposition is presented in terms of what is actually being tested, assessed or evaluated, in addition to language.

Mirroring its concern with the content being tested, this volume is structured around the four educational areas of interest: intercultural communication; autonomy; literature; and assessment of content and language. These areas are then examined for their interaction with issues of testing and assessment. The four areas were chosen because of their educational relevance to the majority of language learners, in the majority of contexts; indeed, the contributors deal with a variety of language learning contexts starting with kindergarten and primary education, through secondary schooling to university.

The four areas chosen are central to the educational endeavour. The intercultural dimension of language learning is increasingly coming to the fore in many situations, and has led to an expansion of communicative competence to include intercultural communicative competence (Byram, 1997). Autonomy is important in that it is linked to lifelong learning, which is increasingly an explicit goal of education (Broadfoot, 2005; Jones & Saville, 2009). The relevance of literature and its importance in education is acknowledged by most educational systems, where a knowledge of first language (L1) literature is a requisite; although the link between L2 learning and literature may not be as strong as it was in the past, the use of literature is more prevalent than is commonly thought, and in many cases it never left the language curriculum (see Gilroy & Parkinson, 1996; Maley, 2001; Paran, 2006). The final section deals with the linking of assessment of content and assessment of language in contexts where English as a second language (ESL) students are in mainstream classes, learning alongside English L1 students (see Mohan *et al.*, this volume; Leung & Lewkowicz, 2006).

Outline of the Four Sections

The four areas listed above provide the focus for each of the sections in this volume. The first section, Intercultural Competence, opens with a survey by Sercu (Chapter 2) that discusses the way in which the definition of the construct of intercultural competence has evolved, and looks at issues of reliability and validity in this context, including issues of defining levels of intercultural competence. Chapter 3, by Korhonen, then provides an example of an intercultural training programme and the testing it involved. Of particular interest is the way Korhonen used web-based learning with her students (with a connection here to issues of learner autonomy, the focus of Chapters 5–7). Liddicoat and Scarino (Chapter 4) then present two case studies in which they explore issues of data elicitation in two contexts (secondary school and university) and a number