Language Testing Reconsidered

Fox, Janna, Wesche, Mari, Bayliss, Doreen, Cheng, Liying

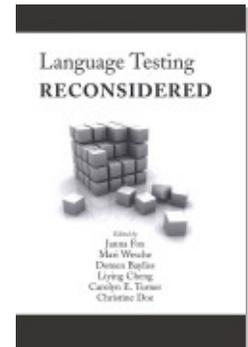➡ For additional information about this book

INTRODUCTION

Janna Fox
Carleton University

Language Testing Reconsidered[1] is a collection of selected papers by several respected colleagues in the field of language testing, who engaged in animated conversation, dialogue, and debate at the Language Testing Research Colloquium (LTRC) 2005, in Ottawa. Part of that conversation is captured here. As the title of this volume suggests, each of the contributors has reconsidered language testing with the benefit of hindsight — looking back at its history, reflecting on personal experiences that highlighted key challenges or issues of concern, taking stock of both the limits and potential of current practice, and looking ahead to future possibilities. Central themes are evident across the chapters in this volume:

- the ongoing challenge of construct definition in language testing and the need for interdisciplinary research, not only within the narrow (and often inward-looking) disciplines of language testing and second language acquisition but also across the broader fields of applied linguistics/language studies and education;

- the expanding repertoire of research methods in language test development and validation, and the recognition of the important role that qualitative approaches can play in increasing our understanding of what tests do and what they measure;

- the evolving and increasing influence of social theory in language testing, and in reconceptualizations of the issue of *context* in test interpretation; and,

- the implications of the use of tests as decision-making tools — their limitations and potential.

As a whole, the volume provides a comprehensive account of issues of continuing concern in language testing and assessment from the perspectives of researchers who have, over the years, contributed in significant ways to the overall direction of the field. It captures the history of language testing, highlights useful methodological approaches in test development and validation, and elaborates a research agenda for the future. As one of the reviewers of this volume noted, "This account could not have happened earlier. Our field was due for this."

1

The first section of the volume is reserved for a paper by Bernard Spolsky and bears a famous question as its title — a question that has long been associated with him — namely, "What does it mean to know a language?" Spolsky acknowledges in his chapter, "On Second Thoughts," that this remains his "fundamental research question." In fact, as evidenced in the discussions across the other chapters in this volume, this question continues to challenge most language testing researchers. Spolsky includes in his paper some of the remarks he made in accepting the International Language Testing Association's (ILTA's) Lifetime Achievement Award at LTRC 2005. This is a remarkable chapter, wherein Spolsky recounts some of the key events in the history of language testing from the perspective of and in relation to his own professional life as a language tester (albeit, an "accidental one," as he points out). History and historian are intertwined in this paper, whether it is Spolsky's articulate lament for our lack of "historical sense" in consideration of current language testing issues; his long-standing concern over the ethics and consequences of large-scale, high-stakes "industrialized" tests; his intriguing discussion of the development of the "noise test" (with colleague Bengt Sigurd) as part of his early quest for "the chimera" of "overall proficiency"; or his recollections of thought-provoking discussions with William Angoff during long, treacherous winter drives between New York and the Educational Testing Service (ETS) in Princeton. His unique insider's view of language testing provides experienced language testers with a rarely shared perspective on the field and is an important frame for the other chapters in the volume. For those who are new to language testing, Spolsky's paper provides an impressive introduction to key questions, developments, and unresolved challenges in the field.

The second section of the volume, "What are we measuring?," consists of three landmark papers by Charles Alderson, Lyle Bachman, and Alan Davies, respectively. Each chapter probes and problematizes what language tests purport to measure.

Alderson, playing the useful role of iconoclast, attacks our soft assumptions about diagnostic assessment in his paper, "The Challenge of (Diagnostic) Testing: Do We Know What We Are Measuring?" Alderson points out that although language testers frequently mention diagnostic assessment, there is little or no consensus on this "rarely defined" and "much neglected" concept. Alderson situates his discussion of diagnostic language assessment within his work on the Common European Framework of Reference for Languages (CEFR), recent experiences with European assessment projects (i.e., the DIALANG project and the Dutch CEFR Construct Project), current theories of language use and ability, and research regarding formative/classroom assessment. Throughout, he highlights what we *do* and *do not* know about language development — knowledge that is essential for useful and meaningful diagnostic assessment and/or feedback. In this paper, Alderson takes the initial steps

in his call for an aggressive research agenda to be undertaken by researchers across the fields of language testing, second language acquisition, and general education. He contends that such a comprehensive approach is long overdue and essential: if we are to clarify what it is to "diagnose" language development, we must first increase our understanding of it. This paper is a call to arms — similar to his call for research on *washback* (which was both a ubiquitous and largely under-examined concept before Alderson and his colleagues examined it). As such, Alderson's paper will certainly garner the attention of the field.

Lyle Bachman provides a comprehensive and thoughtful overview of the testing of oral language in addressing the question "What is the construct?" His paper, "What Is the Construct? The Dialectic of Abilities and Contexts in Defining Constructs in Language Assessment," begins with a review of language testing past, and a discussion of the differing approaches to construct definition that have been dominant within the field over the past 45 years. This historical perspective allows him to examine how the field has evolved in its definitions of abilities, performance, context, and task, and the interactions between them. He highlights the *dialectic* between what has been called *trait/ability-focused* and *task/context-focused* perspectives and examines how these perspectives have influenced the definition of constructs in language testing. Although socially informed *interactional* approaches may suggest a new synthesis to some, Bachman problematizes this notion by identifying important unresolved issues that proponents of a social view have failed to address. In the end, he discusses the implications of language testing past and present for future language assessment research and practice. This paper provides a richly detailed review of language testing and an incisive and well-articulated refutation of current arguments for a social turn in construct definition. It will stimulate ongoing discussion within the field for years to come.

The third chapter in this section is also a provocative one, which promises to engage its readership in considerable debate. In "Testing Academic Language Proficiency: 40+ Years in Language Testing," Alan Davies considers the construct of academic language proficiency as it has been operationalized in tests that were developed over the past four to five decades within the United Kingdom (and later Australia). He examines three attempts to test the English for Academic Purposes (EAP) construct:

1. the British Council's English Proficiency Test Battery (EPTB), developed in the late 1950s;

2. the innovative and experimental English Language Testing Service (ELTS) of the 1980s; and

3. the powerful, high-stakes International Language Testing System (IELTS), of the present day.

In his historical analysis, he links test development to changes in the definition of what language is and highlights sampling as a key issue in representing language use for academic purposes. Davies argues that there is little evidence to support the notion of a tenable EAP construct that is relevant to all students entering English-medium universities regardless of their disciplinary backgrounds or interests. He cites research evidence arising from studies of the three British tests considered in the chapter. Each defines academic language proficiency differently, but each accounts for approximately the same amount of variance (10%) in predictive validity studies. Thus, he argues, a construct of academic language proficiency may not differ in any appreciable (or measurable) way from less situated constructs of language use, precisely because tests — whether they claim to measure global/general proficiency or specific/academic proficiency — are unable to evoke the unique, highly situated relationships, interactions, and engagements that are characteristic of students undertaking academic work. His analysis and this paper directly challenge work on specific purpose language testing and will provoke ongoing debate.

The third section of the volume includes three chapters that examine new "Points of Departure." Each of the contributors to this section, Andrew Cohen, Anne Lazaraton, Lynda Taylor, and Tim McNamara, provides a state-of-the-art review of key approaches and new considerations in research supporting language test development and validation.

Cohen's paper, "The Coming of Age of Research on Test-Taking Strategies," begins this section of the volume with a comprehensive overview of the increasing importance of test-taker strategy research. He systematically examines the movement in language testing research from an exclusive focus on items and scores to an increasing focus on test-taker responses to tests, with particular attention directed at test-taker strategy use. Cohen discusses in some detail the conceptual frameworks that have been used over the years to classify types of strategies used and/or reported by test takers. He examines the contributions of verbal report research in helping us to better understand what tests actually measure and lauds this research for its direct contributions to arguments for validity. His review incorporates many concrete examples of test-taker strategy research and will be enormously beneficial to those new to language testing. It will also be of great benefit to those who wish to rethink and/or re-examine research on test-taker strategy use over the past thirty years in relation to the increased role that test-taker input is playing in test development and validation research.

Given the current, extensive use of qualitative approaches in language test development and research, it may be difficult for some to appreciate how limited a role qualitative approaches once played in language testing research. It is clear, however, in the chapter by Anne Lazaraton and Lynda Taylor, "Qualitative Research Methods in Language Test Development," that the grow-

4

ing presence of qualitative research methods in language testing research advances our understanding of both assessment products and the processes that contribute to them. In their chapter, Lazaraton and Taylor review key studies in speaking and writing assessment, which have utilized methods selected from what they refer to as the *qualitative toolkit*. Through practical example, systematic explanation, and discussion, they provide researchers with a comprehensive overview of how qualitative research tools have been used in the past for both test development and validation. Most importantly, they provide researchers new to these methods with not only an essential background regarding the range of options available for qualitative research but also practical, hands-on explanations of their use.

Tim McNamara's paper, "Language Testing: A Question of Context," completes this section of the volume. True to the volume's title, McNamara reconsiders his experience as language tester and language testing researcher in a thoughtful examination of context — what many consider the most challenging issue in language testing today (see Bachman's chapter, in this volume, for additional discussion of the issue of context). Unlike Bachman, however, McNamara examines context from the perspective of contemporary social theory. He uses the development and validation of the Occupational English Test (OET) as an example in exploring the paper's central premise, namely, "that all tests are tests of identity," because, according to McNamara, "all language tests are about identification." He contrasts traditional conceptualizations, which have attempted to characterize *context* in language proficiency testing but have ultimately failed to resolve the issue with those informed by social theory. In this paper, McNamara applies *subjectivity theory*, as developed in the work of Michel Foucault, in a reconsideration of the key properties of the OET in *context*. McNamara challenges the field of language testing to further explore social theory as an increasingly important explanatory resource for future language testing research.

McNamara's application of social theory in examining the OET provides a richly elaborated backdrop for the final chapter (and section) in the volume, Elana Shohamy's "Tests as Power Tools: Looking Back, Looking Forward." In section four, "Antecedents and Prospects," Shohamy situates her discussion of the power of tests within a personal narrative, beginning with reflections on her experiences as a test taker and describing how those experiences led her into language testing as a graduate student, at a time when language testing was just beginning to emerge as a field. Shohamy's passion and conviction as a language test *reformer* is evident in her account of the movements in her own research agenda from a focus on test development (methods and bias), to a focus on washback (how tests influence teaching and learning), to her current focus on social and political consequences of tests and their power. Her driving motivation in understanding tests and testing has always been "to possibly

5

create better tests," but her understanding of what might make a test better has evolved from micro-consideration of the individual test taker, teacher, tester, and so on, to macro-considerations of the role of tests in education and society.

Like Spolsky, at the beginning of *Language Testing Reconsidered*, Shohamy relates her personal history as a language tester and researcher to developments in the field of language testing. She defines key issues that need to be examined "for a deeper and more comprehensive understanding of language testing" and elaborates a research agenda in the form of specific research questions that will address those issues. As such, she pulls together the themes that are developed across the chapters in this volume: the challenges of construct definition; the expansion of research methods in exploring what tests do and what they measure; the usefulness of socio-cultural/socio-historical theory in providing important new perspectives; and the implications of using tests as decision-making tools. In *Language Testing Reconsidered*, as Pearl Buck once observed, we face our future with our past.[2]

Language Testing Reconsidered has greatly benefitted from the generous engagement of its authors; from the feedback of respected colleagues in the field of language testing, who participated in the blind review process; and from the editors, who carefully examined each manuscript and provided suggestions to improve the overall quality of the book. *Language Testing Reconsidered* could not have been published, however, without the backing of the Social Sciences and Humanities Research Council of Canada (SSHRC), the Second Language Institute of the University of Ottawa, and the School of Linguistics and Applied Language Studies of Carleton University. Proceeds from *Language Testing Reconsidered* will support the work of the International Language Testing Association (ILTA), the organization of professional language testers and language testing researchers, who are committed to improving the quality of tests and who gather each year to discuss their research at the Language Testing Research Colloquium.

Janna Fox
Ottawa, 2007

Note

[1] The editors wish to acknowledge *Language Aptitude Reconsidered* (*Language in Education: Theory and Practice, 74*) (1990), edited by Thomas S. Parry and Charles W. Stansfield (Englewood Cliffs, NJ: Prentice Hall), which, as its title suggests, like *Language Testing Reconsidered*, stimulates further thought and research regarding language assessment.

[2] Pearl S. Buck, winner of the 1938 Nobel Prize for literature, once commented: "One faces one's future with one's past."