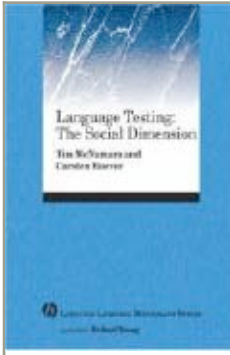


<i>Language Testing: The Social Dimension</i>		
<b>Author:</b>	Tim McNamara & Carsten Roever (2006)	
<b>Publisher:</b>	Oxford, UK: Blackwell Publishing	
<b>Pages</b>	<b>ISBN</b>	<b>Price</b>
Pp. xv + 292	978-1-4051-5543-4 (paper)	£ \$38.95 U.S.



McNamara and Roever's *Language Testing: The Social Dimension* is the first volume in the *Language Learning Monograph Series* to focus on language testing. It opens with a forward by the series editor, Richard Young, and consists of eight chapters that chart the social dimensions of language use in tests and test use in various situations. A brief introduction, chapter 1, orients readers.

Chapter 2 surveys the development of validity theory from the 1950s up until the present, putting Samuel Messick's influential contribution at the center of the discussion. Messick defined validity as "an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the *adequacy* and *appropriateness of inferences* and *actions* based on test scores or other modes of assessment" (1989, p. 13, italics in original). The stance adopted by McNamara and Roever is that while recent work has been beneficial in providing conceptual frameworks for understanding test design and score interpretation, Messick raised questions about the social consequences of testing that others after him often leave unaddressed. Those familiar with this critique from McNamara's recent writing (2006a; 2006b) will be interested in reading the longer version supplied here. The chapter subsequently considers validity as understood within language testing and concludes by turning to the critical language testing movement. Figures describing assessment design and score validation helpfully illustrate the far-reaching ideas in this chapter.

The remaining chapters build on this theoretical background. Chapter 3 provides authoritative coverage of approaches to assessing interaction in face-to-face settings and interlanguage pragmatics, two areas which bring to light the social dimensions of language proficiency. The major issues are reviewed and in-depth treatment given

to research methodologies and instruments, including conversation analysis and discourse completion tests. Most of this chapter is devoted to measures of pragmatics, presumably because they are less widely known. An interesting section on assessing pragmatic aptitude proposes measures that might tap individuals' ability to acquire pragmatic knowledge.

The authors begin chapter 4 with the historical background of investigations into bias and differential item functioning (DIF). They then go on to explain four methods of detecting DIF by examining studies of language testing employing each. These studies present an opportunity to address a recurring theme of this chapter, that of discerning when DIF constitutes bias. One of the merits of McNamara and Roever's book is that it allows us to see how social and psychometric understandings of problems in language testing research are interwoven. The close relationship between these two dimensions becomes apparent when the value judgments behind DIF analyses are exposed at the end of this chapter.

Chapter 5 concerns how fairness reviews and codes of ethics link testers to stakeholders and to the wider language testing community. Fairness reviews are used by large testing organizations such as the Educational Testing Service to attempt to eliminate bias before it occurs and ensure test content that will not be seen as controversial. The International Language Testing Association's *Code of Ethics* (2000) and *Draft Code of Practice* (2005) are designed to raise ethical awareness and to inform practice. The authors express skepticism toward this approach by questioning the extent to which such codes can be enforced.

In chapters 6 and 7, McNamara and Roever turn to issues of test use. They first analyze a broad range of historical and contemporary uses of language tests, from the shibboleth test used to determine identity (referred to in the Bible) to the Second Language Evaluation/Evaluation de langue second administered to Canadian civil servants, revealing how tests may at times function as "weapons within situations of inter-group competition and conflict" (p. 196). Chapter 6 also examines test procedures used to determine the refugee, immigration, citizenship, and professional status of individuals, both from measurement perspectives and in terms how such procedures support policy initiatives. The monograph's *raison d'être* is highlighted when the authors argue that traditional approaches confine our understanding of the social context of language testing within the discourse of psychometrics; and that social theory, especially Foucault's notion of tests as instruments of power, can complement research seeking to address the values and consequences inherent in test use.

Next, chapter 7, on language assessment in schools, looks at the values underlying policies seeking to redefine foreign language achievement in Japan and Europe. This is followed by a comprehensive account of assessment standards at schools in Australia. Several examples show not only how standards-based assessment is often implemented without regard for scholarly opinion but also how scholars themselves do not always share assumptions regarding such assessment. In the U.S. context, the authors' consideration of multiple views on the No Child Left Behind Act leads them to conclude that while the law may have unintended consequences, it also has the power to draw attention to ESL learners' specific needs.

What are the implications of this extensive and detailed discussion? In the final

chapter, McNamara and Roever make several proposals for future research on social factors in language testing, distinguishing impartially between those advancing psychometric and social theory approaches. They envision greater breadth and diversity which, in their own words, "will make the field more socially and intellectually responsive and less isolated from other areas of applied linguistics and the humanities" (p. 254). They also stress that the academic preparation of language testers should include both psychometric theory and critical perspectives on the role of tests in society.

The scope of this book is at once rewarding and challenging. It will certainly appeal to readers with previous exposure to the language testing literature. However, its well-chosen examples of tests, clear discussion of research techniques, and chapter-end summaries of the main points also make it accessible to those without a strong background in testing. The interdisciplinary nature of the volume should enable it to attract a range of professionals whose work involves assessing foreign or second language ability. It offers a unique perspective on the interface between assessment and a number of areas, including second language discourse, pragmatics, and language policy. Further, McNamara and Roever's book has enormous potential to assist teachers and graduate students who share their concern for the values and consequences associated with test use in becoming more conversant with the literature on language testing. I recommend this book for anyone wishing to broaden their understanding of language testing in general and its social dimensions in particular.

## References

International Language Testing Association. (2000, March). *Code of ethics for IITLA*. Retrieved October 28, 2007, from <http://www.iltaonline.com/code.pdf>.

International Language Testing Association. (2005, July). *ILTA: Draft code of practice: Version 3*. Retrieved October 28, 2007, from [http://www.iltaonline.com/CoP\\_3.1.htm](http://www.iltaonline.com/CoP_3.1.htm).

McNamara, T. (2006a). Validity in language testing: The challenge of Sam Messick's legacy. *Language Assessment Quarterly*, 3 (1), 31-51.

McNamara, T. (2006b). Validity and values: Inferences and generalizability in language testing. In M. Chalhoub-Deville, C. A. Chapelle, & P. Duff (Eds.), *Inference and generalizability in applied linguistics* (pp. 27-45). Amsterdam: John Benjamins.

Messick, S. (1989). Validity. In R.L. Linn (Ed.), *Educational measurement (3rd ed.)* (pp. 13-103). New York: American Council on Education and Macmillan.

Daniel O. Jackson  
J.F. Oberlin University, Tokyo  
<danielja@obirin.ac.jp>

© Copyright rests with authors. Please cite TESL-EJ appropriately.