

D. P. Barnwell. *A History of Foreign Language Testing in the United States from its beginnings to the present.*

Tempe, Arizona: Bilingual Press. 1996. Pp. 216.

Barnwell's *History* was published in 1996, a year after Bernard Spolsky's *Measured Words* (Spolsky, 1995). Spolsky documents the institutional history of language testing in the United States, with particular reference to the origins and development of TOEFL and the role in this development of ETS. Barnwell concentrates on the main actors over this period – some well known, Wood, Henmon, Carroll, Agard, Dunkel, Lado, Oller, Valette; some less so – Heuser, Handschin, Vander Beke. Decade by decade, Barnwell characterizes the main language testing activities and finds constant reinventing of the wheel:

Few ask whether problems presented in today's foreign language testing might be illuminated by the recognition that they were faced, if not solved at an earlier date. (: vii)

The main language testing issues recur, he argues, because there is in the profession little or no historical sense. I think that Barnwell is wrong here. Yes, there is constant return to the same issues, for the good reason that issues such as reliability, validity, fairness, scales, the meaning of correlation are not trivial, they have no final solution, they represent the permanent challenges to the profession that time and experience do not resolve.

Barnwell details the differing answers over the years to the question of how language testing conceives of language: over time the view has changed from language as literature, to literary knowledge, literary criticism, metalinguistic knowledge, vocabulary, grammar, communication, culture.

Barnwell himself is attracted by the new style, objective tests of the 1920s, but he seems more comfortable with the 'real life' type test. He argues in his positive account of the FSI tests:

In validity, the FSI interview on the face of it surpassed almost any other kind of testing technique. In order to find out how well a person spoke, he or she was asked to speak....there was no outside mediation between the ability and the expression of that ability. (: 138)

Barnwell, regrets that most language test activity – both academic and institutional – does not deal with what he regards as the major language testing problem, that of the classroom (: 47). He is correct that too little has been done on classroom assessment, but he is quite wrong that what has been the focus – on adults and on proficiency – is unimportant.

Surveying foreign language testing in the US from its beginnings to the present, Barnwell hands the accolade to the 1920s decade. Since then the trajectory has been downhill:

the 1920s in many ways stand as a high point in the history of foreign language testing. (: 54)

He bases his claim on the following:

1. the decades that followed made no progress and simply returned to the advances of the 1920s:

such issues as reliability and validity....these terms were thought of for the first time as being best established by experiment and statistics rather than through intuition and experience....the breakdown of the old test format which had been rooted in the teaching and testing of the classical languages and in which metalinguistic knowledge occupied a central role. This was replaced by tests that sought to measure 'functional' knowledge of the language. (: 54)

2. the involvement of academics who 'enthusiastically committed themselves to language teaching and testing' (: 54/5).

And this seems to be what attracted Barnwell to the 1920s – the teaching–testing fusion. What primarily interested him, and what his book is substantially about is the history of foreign language **teaching** in the U.S. Hence his approval of Ben Wood's experiment (Wood, 1927) in New York City using the 1925 Regents examinations which found that:

thousands of students were misplaced (which) Wood called....the sacrificing of bright students on the altar of mediocrity. (: 72)

And later Henmon's more ambitious project (Henmon, 1929), the findings of which replicated those of Wood (: 41).

Barnwell's concern is primarily with achievement within an educational system. His chapter on 'The Proficiency Movement' makes clear his distaste for the expenditure of effort by language testers on concerns with adult learners which, in his view, are somewhat peripheral:

Foreign language testers thus need to go beyond measuring proficiency and should ask why measured proficiency is so low in the language-learning population. (: 190)

Barnwell is more persuasive when he is dealing with language teaching and achievement than when he tackles what I would regard as more narrowly

language testing issues. He complains about the use of correlations (: 38), of statistics more generally (: 54) and of the use of the native speaker as a criterion (Chapter 8 *passim*). On this last issue, the native speaker, he seems confused (Davies, 2003). He writes:

our knowledge of native speakers is at this stage quite inadequate to allow us to predict how they will assess non-native speech. (: 156)

And then, just a few pages later, in discussing the training of apprentice raters of the ACFTL/ETS scale, he writes:

the more formal training given to apprentice raters, the more their experience diverged from that of the native speaker. (: 164)

But wait a minute! Hasn't he just told us that our knowledge of the native speaker is quite inadequate? How then can we know how these apprentice raters' experience diverged from that of the native speaker'?

Barnwell quotes Paikeday (1985) with approval (: 155) and proposes an alternative model to emulate 'the good foreign language speaker' (: 170).

Such an ambition is not all that far from recent proposals for an English as a Lingua (ELF) variety (Jenkins, 2000; Seidlhofer, 2001). But like these proponents of ELF, Barnwell has not reckoned with the problems attached to a vague criterion which will be used as a norm. Communicative success is not in itself generalizable, not with the analyses currently available.

Barnwell's praise for the 1920s is, as I have suggested, a plea for the fusion of language teaching and testing. Although he discusses the contribution over the decades of academia, government agencies, and testing organisations such as ETS, he finds them all wanting because they are so narrowly concerned with testing alone. This is a principled position, no doubt, and his many animadversions, notably against the ACTFL/ETS scale are well made, but the book does not give a serious account of language testing as it has been practiced and professionalised in the last half century, in the U.S. just as much as elsewhere.

Barnwell ends his book pessimistically:

what has a century of testing achieved?...the real goal of testing is not to find out anything profound but merely to give grades, rewards and punishments - to classify because that is what the educational system needs. (: 190)

And then he reveals what he thinks language testing should do:

The teacher who assigns As, Bs, and Cs at the end of a French 101 college course is not really making any prediction about how his students will handle themselves on the streets of Paris...Typically, there is no externally predictive element to the foreign language grade that is assigned...Contrast this with the need for tests of, say, medical or engineering ability; these must reliably predict how well people will subsequently use that knowledge in building bridges or saving lives. (: 190)

Such an indictment indicates just how unaware Barnwell really is of the problems, the repeating problems, of language testing. The connection between a French 101 course and the streets of Paris is not as straightforward as his comparison examples, medical graduates and saving lives and engineering graduates and building bridges. Here Barnwell shows how little he understands of language and of testing.

Barnwell's book was published in 1996: does it still have resonance for us today? I suggest in three ways: the first is that he has assembled a wealth of often arcane information which future historians will be grateful for. The second is that he is correct in his complaint that language testing has just not directed enough of its attention to the largest group of language learners, students in schools and universities enrolled in foreign language programs. The third is that for a critique to be valid it must examine a profession in terms of its stated purpose and not for not being something else.

Review by Alan Davies

University of Edinburgh

References

- Davies, A. (2003). *The Native Speaker: myth and reality*. Clevedon: Multilingual Matters.
- Henmon, V. A. C. (1929). *Achievement tests in the modern foreign languages*. New York: Macmillan.
- Jenkins, J. (2000). *The Phonology of English as an International Language: new models, new norms, new goals*. Oxford: Oxford University Press.
- Paikeday, T. M. (1985). *The Native Speaker is Dead!* Toronto and New York: Paikeday.
- Seidlhofer, B. (2001). 'Closing a conceptual gap: the case for a description of English as a lingua franca.' *IJAL*, 11, 133-158.
- Spolsky, B. (1995). *Measured Words* Oxford: Oxford University Press.
- Wood, B. D. (1927). *New York experiments with new-type modern language tests*. New York: Macmillan.