

Book Review

Re-examining Language Testing: A philosophical and social inquiry.

Glenn Fulcher

London: Routledge 2015 pp i-ix-232

On the subject of *testing* Glenn Fulcher is has a world-wide reputation and consequently this book is important to help teachers reflect on what is involved in something that is very much part of their teaching life. If we simply take the chapter titles in this book, we see chapter headings of what would be expected in a book on testing. The book discusses the major features such as: *Inference, Measurement, Language, Numbers, Validity, Meritocracy* and *Values*. However, this book is much more than that as the full title indicates it is ‘a philosophical and social inquiry’.

The author outlines what is involved in language testing.

- The tasks should be relevant to the decision to be made.
- The responses elicited by those being tested should be useful in decision making.
- The scores should summarize the responses.
- Scores can be used as necessary but not sufficient evidence for any decision made.

Fulcher discusses what is testing as opposed to assessment and the power and politics of English high-stakes examinations in which English is embedded in local economies. He looks at the ways in which demand for English is part of the larger picture of ELT as it is tied up with forms of power and prejudices shedding light on the assumptions of privilege that are associated with native speaker authority. As a consequence of this an important issue is evidence from *inference* and how often this is little more than consensual professional judgment. In language testing this has typically degenerated in face-validity claims that the test ‘looks’ authentic in content or format.

As regards *validity* and *reliability*, statistically this often involves some form of measurement such as the Bell Curve. That is to say that when we have large numbers of observations (results/scores) the distribution is close to normal probability distribution. So if the mean =50 and the SD =5, we know that probably 68% (2/3) of the students are within the range of scores 55 to 45 indicating a fairly homogeneous group. However, there is still a fundamental assumption that the data are normally distributed and in the pursuit of psychometric responsibility, too often we ignore findings from the way language is used in society, particularly in second/foreign language situations. Research into English in a global context is inherently more fluid and dynamic and therefore predicts that variable performance will be the norm rather than the exception.

In a later chapter the author points out that the language of numbers is often considered to be symbols of objectivity and give the illusion of truth and legitimacy. Certain numbers have become icon such as 6.5/7 on the IELTS test. However this scaling has changed little since its forerunner the English Language Testing System (ELTS) developed in the 70s. Score users and the public have come to accept the truth that this is the point on a nine-band scale at which a test taker is capable of undertaking English medium studies in higher education.

Then there is the issue of *standards* and *validity*. In any set of standards each level is defined by a descriptor as for example the nine levels of IELTS. However, validity problems for language testing has arisen because standard documents have been treated not as policy statements but descriptors of proficiency. This is also the case of *criterion-reference* assessment in which the criterion is interpreted as a *standard*. If we take for example the standard of *Common European Frame of Reference* (CEFR) means that any test organization that wishes its test to be used for the purpose of assessing students' level of English will produce a 'cut-off' score that links the test to the CEFR required level. The implication is that once a test has been scaled and standards set to external standards, it has been 'criterion referenced'.

Applying this to the ASEAN region has resulted in educational policies of some countries requiring a C2 for University teachers (IELTS 8-9), a C1 for teachers in Higher Education (6.5/7) and a B2 for teaching in secondary education (5.5/5) and a B1 (4.0) for primary. The issue then is since language is by nature variable to some extent, do these descriptors and consequent cut-off points reflect *real-life* situations in this part of the world?

These are just a few issues raised in Fulcher's book in later chapters, he

looks at the *ethics* and *fairness* of testing. Importantly, given the reality of English as a global language, the challenge for English language testers in this 'uneven' world is to move away from simply native-speaker correctness. There is a need to look at the global sociolinguistic reality. Indeed the *wash-back* effect of the present forms of international testing, in particular, promote an outdated view of communication in English as being relatively fixed and based on native speaker norms. What Fulcher advocates is *pragmatic realism* which holds that progress is achieved through open questioning of theories and evidence and that the 'thinking' teacher should pose for themselves the question of whether validity in tests, especially high-stakes testing reflect reality of how the target language, in this case English is used in real-life.

J.A.Foley

Graduate School of English

Assumption University