TEST INTERPRETATION AND DIVERSITY: ACHIEVING EQUITY IN ASSESSMENT

Jonathan Sandoval, Craig L. Frisby, Kurt F. Geisinger, Janice Dowd Scheuneman and Julia Ramos Grenier (editors)

American Psychological Association, Washington, DC 1998
ISBN: 1-55798-509-X (hbk) £37.50

Reviewed by Rae Vermeulen
Disability Services,
1st Floor,
236 Grays Inn Road,
London,
WC1X 8HL

This book is the report of the Task Force of Test Interpretation and Diversity, formed by the American Psychological Association’s Committee on Psychological Tests and Assessment. The aim in writing the book was to provide guidelines to practising psychologists in the USA on interpreting test results of individuals with cultural, linguistic or economic backgrounds that differ substantially from the majority population, and of individuals with disabilities. A recurrent theme throughout the book is the development of critical thinking rather than the adoption of a cookbook approach to test interpretation. Although the book is about the practice of psychology in the United States, there are many useful ideas and suggestions for good practice for psychologists in any country.

The book is divided into four parts:

• Part 1 reviews psychometric issues for test interpretation, discusses the critical thinking needed in test interpretation, and includes political perspectives;

• Part 2 deals with three contexts in which high-stakes decisions are made: educational, employment and clinical, each having different kinds of tests, different legal constraints, and different challenges for test interpretation;

• Part 3 addresses various social and psychological dimensions of diversity: language, acculturation, poverty, vision, hearing, learning difficulties including dyslexia;

• Part 4 discusses training needs for psychologists and future changes in the technology
of measurement and testing.

This book provides a thorough review of the key assessment issues in using tests with people with different linguistic backgrounds and disabilities. It therefore includes detailed discussions of all aspects of test fairness including the validity of test scores. The book is an excellent source of information on the literature of test use in assessment with a full list of references being given at the end of each chapter.

As each chapter is written by a different expert, there is inevitable repetition of some points and variability in the quality of writing. This is anyway not a book to be read from cover to cover but, with a comprehensive author and subject index, relevant sections are easily found. With such a broad review, only a few points of particular interest can be covered.

Following on from the article by Lanchbury and Kearns (2000) on accommodations, there is an interesting suggestion on ranking accommodations depending on their impact on psychometrics:

- Level 1 accommodations would have little or no appreciable impact eg using a writing slope;
- Level 2 accommodations would possibly have an impact eg extending the time limits for speeded tests;
- Level 3 accommodations would have the most impact eg omitting tests from a test battery.

The psychologist would then need to weigh up the risks associated with the accommodations with the importance of the decisions being made. For example, accommodations for examinations for professionals would be incompatible with the high level risks to test validity posed with accommodations at levels 2 and 3. For people with dyslexia, the particular problem is that the target variables (eg tests of academic achievement) that are measured in competitive, professional situations are usually those that are most problematic for people with this disability.

“Flagging” test scores is another problem area in testing in the States. Scores are flagged to show that test accommodations have been used. Where information on the nature of the disability has been given to explain the flagging, this has been challenged as a violation of privacy. However, test scores that are flagged without the explanatory information needed to interpret the test scores may violate the ADA - the American equivalent of the DDA. No doubt we will hear more.

There is a full discussion on evaluating test fairness in employment testing and an explanation of the increasing use of “test bands” in selection. Test bands aim to balance the social goal
of reducing adverse impact with the measurement goal of maximising selection utility. They operate from the premise that all observed scores contain measurement error and that applicants with different observed scores may actually have similar true scores. Estimates of measurement error are used to set up test score bands within which applicants are considered to be equal. In order to reduce adverse impact, candidates may be selected in proportion to their representation in the total pool rather than being selected using a simple top-down approach.

There are interesting snippets of information which we may wish to pursue in the UK. The most recent edition of the adult Wechsler scales has a formal American Sign Language translation that was developed and distributed by the test publisher. Should we suggest a BSL translation of the WAIS-III to the Psychological Corporation?

In the chapter on learning disabilities (dyslexia in British terminology), the problem of definition is seen as arising from a failure to distinguish between classification which is driven by theory and which is measurement free, and identification which requires assignment to a category and which is bound by measurement. The difficult nature of dyslexia assessment is acknowledged: most American states adopt an aptitude-achievement discrepancy for diagnosing dyslexia but there is little agreement on how to quantify this. As test users, we can take encouragement from the theory development in this area that, the author assures us, lies ahead.

Test users should also feel encouraged by the increasing efforts of test publishers to build accommodations into tests. The WAIS-III, for example, does not require reading or writing and its one-to-one administration allows the administrator to adjust to problematic behaviours such as inattention or distractibility.

This publication includes much of interest to occupational psychologists who need to make judgements about people. There is sound advice on assessment and interpretation of test scores particularly in part 3 of the book where I found the chapters on psychological dimensions of diversity - language, vision, hearing, dyslexia and other disabilities - particularly useful.

I would recommend this book strongly to anyone with an interest in psychometrics and to those psychologists delivering or receiving Level A training. In my opinion, most of the chapters would be too detailed for Level A qualified test users without a background in psychology. For the rest of us, this book makes a sensible and considered contribution to the responsible assessment of all our clients.

REFERENCES
