

## BOOK REVIEWS

### ■ The ITC International Handbook of Testing and Assessment

Leong, F.T.L., Bartram, D., Cheung, F.M., Geisinger, K.F. & Iliescu, C. (Eds.) (2016). New York: Oxford University Press

As the editors say in their introduction to this book; in a world which is heading towards globalization, psychology, if it wants to remain viable and successful, has no alternative but to integrate models, theories and perspectives with a transnational focus. In testing and psychological and educational assessment the International Test Commission (ITC) has been a key organization in implementing this international perspective, and this concern has inspired this volume. With this focus, the book sends us on a journey, not via psychometric methods (it is not just another book about psychometry), but rather via psychology, education and health, where evaluation technologies (testing amongst them) have a central role. This cross-sectional tour is in two strands: geographical, and the fields of application of evaluation tools in social and health sciences (especially psychology). In line with their proposed model, the book is divided into six sections.

#### *I. Introduction and overview*

This first section reviews the history of tests and the ITC itself. In terms of the latter, it emphasizes the main successes embodied in the five guideline documents from the work of its various committees, to wit; Test use; Computer-based testing; Internet security of tests and other assessments; and Quality control in scoring, test analysis, and the reporting of tests scores. This section also summarises the ITC's collaboration with other associations such as the EFPA (European Federation of Psychologists' Associations).

#### *II. Domains of testing and assessment*

This section covers some of the most important topics evaluation has been concerned with: intelligence and other abilities, motivation and emotion, attitudes, values, interests, personality, psychopathology and neuropsychology. Each chapter in this section gives a reminder of the current status of the definition of the construct being dealt with and the usual evaluation procedures in order to pose alternatives, from a perspective of constructive criticism, which consider the differences that may arise in different geographical areas of application.

#### *III. Settings*

Through considering various settings (educational, labor, clinical, forensic, health, and rehabilitation) this section reflects

on some of the burning issues in, for example, procedures for accountability in the education system, large-scale evaluations for international comparisons (PISA, TIMSS, etc.), the predictive capacity of personnel selection and promotion procedures in ever more internationalized settings, the "exportability" of clinical instruments such as the MMPI or the Rorschach test to non-western cultures, the need to consider the particular characteristics of the population to evaluate and the context of forensic practice, and the advantages of using a model which considers the interaction between factors of physical health, activities and social participation, personal and environmental factors in an integrated vision of health and rehabilitation.

#### *IV. Special populations*

Over six chapters this section addresses situations which must be faced when the individuals being evaluated present certain differential traits. It considers the accommodation of evaluation tests when applied to people with disabilities, and advocates the construction of more accessible tests. When evaluating children and adolescents, the developmental sequence is a factor that cannot be separated from the evaluation of those groups and something which must be considered in the design of test batteries. This same developmental perspective must be maintained in the evaluation of populations of older adults which raises the issue of the equivalency of measures that are not developed specifically for this age group. The two following chapters focus on language, the first is concerned with evaluation using a language other than the evaluation subjects first language, the second looks at evaluation of immigrants lacking fluency in the language of the country they are in, and consequently, being second-language learners. This section ends with the consideration of those positions defending the need for a "local focus" in evaluations as opposed to the more classical western position defending the universality of psychological constructs created in our environment.

#### *V. Methodological advances*

This, the most strictly psychometric section, examines some of the technical challenges that the best evaluation instruments must meet in the field of psychology and education, the tests. It covers topics such as the classics of reliability and validity, the identification, avoidance and correction of response biases, and more recent tests such as measurement and structural equivalence via various groups using Confirmatory Factor Analysis. It also looks at the influence of advancing technology on evaluation.

#### *VI. Problems and challenges*

The final three chapters address future lines for this international perspective on evaluation. The first looks at the need to deal with ethical aspects of professional practice, paying special

attention to test construction and application. The second focuses attention on the creation and operation of standards that respond appropriately to new, ever more supranational circumstances of test creation and application. The final chapter addresses the challenges to be faced in the ever-controversial topic of the use of tests in high-stakes decision making, especially in an educational setting, both in the evaluation of educational centers and in admission processes to higher education.

In sum, what we have before us is a book about challenges. Challenges presented by the universalization of psychology and transculturality now and in the near future. This future must be approached from the understanding that it is not enough to transfer dominant western models to other cultural contexts but rather that universality must be achieved from a global vision with local nuance, we must combine the *etic* (universal) with the *emic* (indigenous). The new playing field we are starting to see makes this book recommended reading, not only for those with an interest in the field of evaluation, but also those interested in behavior in general, as globalization is a process which affects us all, and the transculturality alluded to many times is no longer a synonym of transnationality, but rather something we also experience on a daily basis inside our own borders.

Reviewed by:

**Marcelino Cuesta Izquierdo**

Department of Psychology

University of Oviedo

Spain

## ■ Educational measurement: From Foundations to Future

Wells, C. S., & Faulkner-Bond, M. (Eds.) (2016). New York: The Guilford Press

This handbook is inspired by the contributions made by Dr Ronald K. Hambleton to the field of educational measurement. The structure of this book, edited by C. S. Wells & M. Faulkner-Bond, follows the major assessment methodologies investigated and advanced during a career filled with ground-breaking innovation. The esteem, affection and admiration shown by his colleagues in the prologue, acknowledgments and epilogue, clearly illustrate how far-reaching Dr Hambleton's influence and inspiration has been during his career; it is not surprising that so many advances in educational measurement parallel the career of this distinguished professor from the University of Massachusetts Amherst, School of Education.

Education and psychology experts – more than 50 from all over the world – have contributed to this work, which can be considered an extension of the “Teach your Children Well: A conference Honoring Ronald K. Hambleton” held in 2012. Amongst educational psychologists, this conference is affectionately known as “The Conference”.

This handbook is aimed at professionals and students of education, psychology and other related disciplines, with the objective of providing both historical and applied perspectives of the key areas of educational measurement in an accessible manner. In this sense, this work is intended to encourage practice by reaching a wider range of society who can benefit from the knowledge described.

The handbook begins with an exploration of the historical foundations of the concepts involved. This is followed by the core of the book, which is divided into five main parts. Each of these parts is structured similarly and is based on an exhaustive revision of previous literature; the parts each cover one of the main topics of educational measurement: Criterion-Referenced Testing (CRT); Item Response Theory (IRT); Computer-Based Testing (CBT); Cross-Lingual Assessment; and Accountability Testing and Score Reporting. In the final section, the handbook unites the past, present and future, culminating with discussion and suggestions from the authors of future directions in educational measurement, and of the hopes and challenges confronting those working in the field. The future possibilities stemming from this work are great.

The structure and development of this handbook have been guided by two principles: practicality and utility. Thus it is both an inspirational and a useable guide for people who are concerned about education and want to improve teaching and learning outcomes through improved and better-utilized assessment of students. Although this idea has often not been clearly understood in the past, this book considers on-going and meaningful assessment as an integral part of the learning paradigm, instead of something external or redundant to the process.

### *I. The roots*

The assessment practices of the 21st century are rooted in the historical perspectives covered in this first part of the book. This section establishes the 200-year path of the development of educational assessment, including developments in related fields – mathematics, psychology and sociology – as well as highlighting contributions from different cultures, events, and technology from throughout the world.

Firstly, the contributions of the most representative figures from the history of psychometry – including Galton, Fisher, Catell, Binet and Terman – are explored, and their relevant contributions presented. This is followed by a literature review of the basic concepts which underpin educational assessment, such as the trend for large scale assessment related to intelligence or scholastic admission.

Secondly, this part of the book explores the development of the different theories and psychometric measures that form the basis for current knowledge and understanding about assessment and classifies these into two main groups: Classical Test Theory and Item Response Theory. Following this, contemporary assessment trends, such as the Criterion-Referenced Testing movement and Computer-based and Adaptive Testing (CAT) approaches, are reviewed.

This section concludes with a discussion of the idea that, to most efficiently and effectively incorporate advances in technology and psychometrics as we strive towards improved professional practices, we should not simply dive into the future; rather, we should learn from and build upon our knowledge of the past.

### *II. Criterion-referenced testing*

In the second section, the historical study of Criterion-Referenced Testing (CRT) leads to an in-depth exploration of assessment consistency, as well as a discussion of the advances made, since the 1970s, of Standard Setting methods. In the implementation of Standard Setting methods, it is reliability and its assessment that are key to developing correctly standardized educational assessment measures. CRT is intimately linked to the development of new technologies and educational assessment reports, which are widely known and relevant nowadays. In this context, the necessity of the Standard Setting methods is highlighted. This part provides a step-by-step detail of the process, as well as explaining the key

concepts necessary, for correct measurement standardization. The most commonly used and most efficient and effective methods are covered before a reflection regarding future directions and challenges.

This section concludes with three chapters that have a practical focus. The first of these presents validity interpretation as a relevant component of the test design process and then explores the interpretation of this validity in light of results obtained from Performance Level Descriptors (PLDs). The second chapter explores how best to achieve accurate and consistent classification in decision-making using practical examples and different procedures based on the three models: Classic Test Theory (CTT), IRT, and Measurement Decision Theory (MDT). Finally, in the third of these chapters, exploring some of the challenges relating to educational assessment leads us to Generalizability Theory and its applicability to data interpretation when considering the context and taking error into account. Despite this chapter concluding with several basic and practical examples of CRT theory, a deeper understanding, such as that referenced in the first part of this section, would nonetheless be required for meaningful implementation.

### *III. Item Response Theory (IRT)*

This section introduces the concepts relevant to Item Response Theory (IRT), beginning with consideration of the different probability-based models associated with IRT, and their benefits and drawbacks. Notably, the chapters highlight how conceptual, mathematical (estimation models) and technical advances have contributed to the development of this theory. Due their relevance to IRT and to the ideas of Professor Hambleton, Diagnostic Cognitive Models (DCM) and Cognitive Diagnostic Assessment (CDA) are also described.

This then leads us to the difficult process of advancing from unidimensional IRT to the more complex models required by Multidimensional Item Response Theory (MIRT), a process which requires the incorporation of certain concepts from the different models of psychometric analysis. Within this context, analyzing the differential functioning of test items and item sets becomes important for interpreting the multidimensionality of variables and methods. This results in an increased use of Experience Sampling Method (ESM) as well as an increased requirement that the latent variables are effectively controlled. The different models are also represented graphically, helping to highlight the similarities and differences between item representations, test information, true score representations and conditional representations in the context of MIRT.

This section concludes by establishing the potentialities of the IRT models, explaining its necessity, the main advantages and disadvantages of these models, the assumptions, and the relationships with real data. Consistent with this objective, it shows the relevance of the role of dimensionality, local Independence and the Item Characteristic Curve (ICC). These concepts are very useful in psychometric assessment as they allow for the correct implementation of different analyses such as Confirmatory and Exploratory Factor Analysis and the Bifactor Model, as well as non-parametric approaches, in a reliable and robust way.

### *IV. Computer-based testing*

This section begins by investigating the applicability of technology in psychometric and educational assessment. The history of CBT is, understandably, not very extensive. Its various implementations however, are endless and it will obviously play a large and very positive role in the future of educational assessment. There is a clear parallelism as ICT (Information

and Communication Technology) continues its expansion into education systems globally and into society.

Historically, the use of technology has been related to its application in instruction and in the development of new testing methodologies and policies. These contexts are illustrated through several initiatives in the USA, such as CAT (Computerized-Adaptive Testing) programs and Item Banking. This section then explores the need to further develop the infrastructure around CBT as it continually adjusts to ongoing technology changes. The main issues currently faced by this form of evaluation are then analyzed and the main reasons to use this approach – accuracy, convenience, speed of results and tools – are highlighted. Models for developing CBT are also explained, with emphasis on the CAT model and its advantages, including the use of these models as a part of instruction in education. This form of testing, with its short but fast growing history, presents some challenges as it expands into newer developments including flexibility of testing location, artificial intelligence, calibration and open-source platforms. Finally, the last two chapters present new paradigms in CBT, highlighting gaming and advanced simulations and generation of automated scoring, discussing their main advantages and disadvantages. More specifically, the focus of the final chapter is Simulation and Game-Based Assessments (SGBAs); the structure, the design and the possibilities that these tools can offer in terms of data analysis (also considering the increase of psychometric validity in the last few years) are all explored.

### *V. Cross-lingual assessment*

In what can be considered the most internationally-focused section, some of the technical challenges faced by our dynamic global society in applying meaningful and equitable educational assessment across language boundaries are explored. Main method, design and concepts involved in the generation of cross-lingual tests are reviewed before moving on to suggestions and discussion of future trends in cross-lingual assessment and taking into account the distinction from cross-cultural assessment. The second chapter of this section is an analysis, from a professional perspective, of the transfer and adaptation of tests to different contexts – the idea being to replace test translations with test adaptations. Both current projects and future innovation possibilities are considered with the aim being to develop integral assessment, with comparable and standardized tests, at international level. Finally, following the guidelines of the International Test Commission (ICT), the process of standardizing adapted tests is described with differences between countries (legal context, language, culture, use ...) considered. In short, this section brings together the considerations and process for anyone in the international community to follow for successful adaptation of any test.

### *VI. Accountability in testing and score reporting*

The six chapters in this section focus on accountability and testing, considering in particular the requirements related to educational accountability during the assessment process. Concrete aspects such as assessment procedures, dealing equitably with learning disabilities and aspects of teacher evaluation are explained. The use of assessment measures to improve educational practice is undoubtedly one of the most important aims of educational measurement in the professional setting. An emphasis is placed on considering the so-called “Growth Models” and on the importance of measuring gains in learning. The relevance of assessment reports and the effective comparison of results across different places, cultures and education systems is then discussed. Finally, the last three chapters describe the history of testing special

populations and review the main milestones encountered so far during the development of meaningful assessment in learning disability and language learning environments. With respect to the validity of the instruments and adaptations particularly relevant aspects that should be considered when measuring specific groups are described. It is also worth noting the alternative assessment processes proposed in this section, with the USA perspective used as a clear example for understanding scoring criteria, the large-scale assessment system called AA-AAS, achievement level, and connections with instruction, among other factors.

#### *VII. Ongoing debates and future directions*

This handbook concludes with a synthesis explaining the current trends in educational assessment (and associating these with the corresponding chapter) and then offering presenting a globally-focused, summarized view of the entire handbook. Considering their possible future impacts in the fields of education

and psychometry, the many and varied topics addressed in this handbook are of great significance.

This book, developed as a tribute to Professor Ronald K. Hambleton and his career, brings together the focusses of his work in a practical and accessible manner; thus, his work can continue to guide and inspire investigators and practitioners well into the future as the evolution of educational measurement, with its increasing challenges, continues. New educational paradigms make this book timely reading for anyone interested in educational assessment, and an essential reference for professionals dealing with assessment in education, psychology and other related fields.

Reviewed by:

**Celestino Rodríguez**

Department of Psychology

University of Oviedo

Spain