

ADVANCES IN EDUCATIONAL AND PSYCHOLOGICAL TESTING

edited by

Ronald K. Hambleton

Jac N. Zaal

Kluwer Academic Publishers

Boston/Dordrecht/London

Advances in Educational and Psychological Testing: Theory and Applications

edited by
Ronald K. Hambleton
Jac N. Zaal



Kluwer Academic Publishers
Boston/London/Dordrecht

Distributors for North America:
Kluwer Academic Publishers
101 Philip Drive
Assinippi Park
Norwell, Massachusetts 02061 USA

Distributors for all other countries:
Kluwer Academic Publishers Group
Distribution Centre
Post Office Box 322
3300 AH Dordrecht, THE NETHERLANDS

Library of Congress Cataloging-in-Publication Data

Advances in educational and psychological testing : theory
and applications / edited by Ronald K. Hambleton, Jac N.
Zaal. p. cm.—(Evaluation in education and
human services series)

Includes bibliographical references.

ISBN 0-7923-9070-9

1. Psychological tests. 2. Educational tests and
measurements.

I. Hambleton, Ronald K. II. Zaal, Jac N. III. Series:
Evaluation in education and human services.

BF176.A42 1990

150'.28'7—dc20

90-31493
Rev. CIP

Copyright © 1991 by Kluwer Academic Publishers

All rights reserved. No part of this publication may be reproduced,
stored in a retrieval system or transmitted in any form or by any
means, mechanical, photocopying, recording, or otherwise, without the
prior written permission of the publisher, Kluwer Academic Publishers,
101 Philip Drive, Assinippi Park, Norwell, Massachusetts 02061.

Printed on acid free paper.
Printed in the United States of America.

ADVANCES IN EDUCATIONAL AND PSYCHOLOGICAL TESTING: THEORY AND APPLICATIONS

EVALUATION IN EDUCATION AND HUMAN SERVICES SERIES

Editors:

George F. Madaus
Boston College
Chestnut Hill, Massachusetts, USA

Daniel L. Stufflebeam
Western Michigan University
Kalamazoo, Michigan, USA

Previously published books in the series:

Kelleghan, T., Madaus, G., and Airasian, P.: THE EFFECTS OF STANDARDIZED TESTING
Madaus, G. (editor): THE COURTS, VALIDITY, AND MINIMUM COMPETENCY TESTING
Brinkerhoff, R., Brethower, D., Hluchyj, T., and Nowakowski, J.: PROGRAM EVALUATION, SOURCEBOOK/CASEBOOK
Brinkerhoff, R., Brethower, D., Hluchyj, T., and Nowakowski, J.: PROGRAM EVALUATION, SOURCEBOOK
Brinkerhoff, R., Brethower, D., Hluchyj, T., and Nowakowski, J.: PROGRAM EVALUATION, DESIGN MANUAL
Madaus, G., Scriven, M., and Stufflebeam, D.: EVALUATION MODELS: VIEWPOINTS ON EDUCATIONAL AND HUMAN SERVICES EVALUATION
Hambleton, R., and Swaminathan, H.: ITEM RESPONSE THEORY: PRINCIPLES AND APPLICATIONS
Stufflebeam, D., and Shinkfield, A.: SYSTEMATIC EVALUATION
Nowakowski, J.: HANDBOOK OF EDUCATIONAL VARIABLES: A GUIDE TO EVALUATION
Stufflebeam, D.: CONDUCTING EDUCATIONAL NEEDS ASSESSMENTS
Cooley, W., and Bickel, W.: DECISION-ORIENTED EDUCATIONAL RESEARCH
Gable, R.: INSTRUMENT DEVELOPMENT IN THE AFFECTIVE DOMAIN
Sirotnik, K., and Oakes, J.: CRITICAL PERSPECTIVES ON THE ORGANIZATION AND IMPROVEMENT OF SCHOOLING
Wick, J.: SCHOOL-BASED EVALUATION: A GUIDE FOR BOARD MEMBERS, SUPERINTENDENTS, PRINCIPALS, DEPARTMENT HEADS, AND TEACHERS
Worthen, B., and White, K.: EVALUATING EDUCATIONAL AND SOCIAL PROGRAMS
McArthur, D.: ALTERNATIVE APPROACHES TO THE ASSESSMENT OF ACHIEVEMENT
May, L., Moore, C., and Zammit, S.: EVALUATING BUSINESS AND INDUSTRY TRAINING
Abrahamson, S.: EVALUATION OF CONTINUING EDUCATION IN THE HEALTH PROFESSIONS
Glasman, N., and Nevo, D.: EVALUATION IN DECISION MAKING: THE CASE OF SCHOOL ADMINISTRATION
Gephart, W., and Ayers, J.: TEACHER EDUCATION EVALUATION
Madaus, G., and Stufflebeam, D.: EDUCATIONAL EVALUATION: CLASSIC WORKS OF RALPH W. TYLER
Gifford, B.: TEST POLICY AND THE POLITICS OF OPPORTUNITY ALLOCATION: THE WORKPLACE AND THE LAW
Gifford, B.: TEST POLICY AND TEST PERFORMANCE
Mertens, D.: CREATIVE IDEAS FOR TEACHING EVALUATION
Osterlind, S.: CONSTRUCTING TEST ITEMS
Smith, M.: EVALUABILITY ASSESSMENT
Ayers, J., and Berney, M.: A PRACTICAL GUIDE TO TEACHER EDUCATION EVALUATION

Preface

Over the last 20 years there have been a large number of technical advances and changes in the field of educational and psychological testing. According to Anne Anastasi,

The decade of the 1980's has been a period of unusual advances in psychological testing. Technological progress, theoretical sophistication, and increasing professional responsibility are all evident in the fast-moving events in this field (A. Anastasi, *Psychological Testing*, Sixth Edition. New York: Macmillan, 1988).

On the psychometric front, advances in topics such as item response theory, criterion-referenced measurement, generalizability theory, analysis of covariance structures, and validity generalization are reshaping the ways that ability and achievement tests are constructed and evaluated, and that test scores are interpreted. But psychometric advances, as substantial and important as they have been, are only a fraction of the major changes in the field of testing. Today, for example, the computer is radically changing the ways in which tests are constructed, administered, and scored. Computers are being used to administer tests "adaptively." That is, the sequence of questions an examinee is administered depends upon his or her performance on earlier administered items in the test. Tests are "adapted" to the ability levels of the examinees who are being assessed. One result is shorter tests with little or no loss in measurement precision. Computers are also being used to store or bank test items. Later, items of interest can be selected, and the computer is used to print copies of the test.

The field of cognitive psychology is also having a major influence on testing practices. Researchers such as Robert Sternberg, Robert Glaser,

Richard Snow, and Susan Embretson in the United States have highlighted the importance of new types of cognitive variables in instruction and assessment as well as new ways for identifying important cognitive variables (i.e., through cognitive task analyses). Development of tests to measure these variables proceeds from a theory-based orientation unlike the more common pragmatic test development approaches that require extensive validation work after tests are constructed.

Also, test results are increasingly being used to address policy questions. National and state assessments, for example, which exist in many countries, are now being used to monitor the quality of educational progress and direct financial resources to areas of concern.

In view of the recent rapid, extensive, and important developments in educational and psychological testing, the International Test Commission (ITC) felt that it could further its mission by organizing a collection of papers from researchers in several countries which focused attention on the technical advances, advances in applied settings, as well as emerging topics in the testing field. The result was this book which contains 14 chapters from 22 contributors. A brief description of the ITC itself follows this preface.

The book is divided into three main parts. Part I—*Methodological Advances*—contains six chapters. Ronald Hambleton and H. Jane Rogers, from the University of Massachusetts at Amherst in the United States, describe the concept of criterion-referenced measurement and how it is reshaping assessments in schools, industry, and the military. The testing focus is on assessing what it is persons know and can do rather than on how persons compare with one another. The latter testing focus is central in norm-referenced measurement. David Weiss and Michael Yoes, from the University of Minnesota in the United States, describe a new theory of measurement, item response theory (IRT). IRT has become immensely important as a new measurement framework for test construction and test evaluation. It provides descriptors of test items that do not depend upon the particular samples of persons used in item calibrations, and provides ability estimates that do not depend upon the particular choices of test items from the domain of items measuring the ability of interest. Both measurement features are very useful in the hands of measurement specialists. Dato de Gruijter and Leo van der Kamp, from Leiden University in The Netherlands, and H. Swaminathan, from the University of Massachusetts at Amherst in the United States, prepared chapters on generalizability theory and analysis of covariance structures, respectively. Both chapters address research concerning the analysis of data for test validation purposes. Generalizability theory represents an

extension of classical test theory, while covariance structure analysis is an extension of factor analysis methods. Wim van der Linden, from the University of Twente in The Netherlands, describes some of his research involving testing and decision making in the presence of uncertainty and different consequences associated with errors in classification decisions. Finally, John Hunter, from Michigan State University, and Frank Schmidt, from the University of Iowa in the United States, have addressed issues and methods of meta-analysis or validity generalization, as it is often called. This line of research has become important in industry and the military as attempts are made to make strong statements about test validity in the presence of many small sample studies with different predictor tests, criterion measures, test designs, and groups.

Part II—*Developments in Applied Settings*—contains five chapters. Robert Roe and M. Greuter, from the University of Delft in The Netherlands, have addressed many of the current issues and methods associated with personnel selection. With the need to improve job performance in industry as well as to prepare for lawsuits brought by rejected applicants clearly in mind, current research on personnel selection is substantial and important. From inside the Dutch government, Jac Zaal describes some of the current research that is focused on the uses of assessment centers in performance evaluations and personnel selection. Stan Scarpati, from the University of Massachusetts at Amherst in the United States, reviews some of the current practices and assessment issues confronting professionals who work with handicapped children and adults in public settings. Testing of the handicapped is common in the United States, but a large number of problems and technical issues need to be overcome to insure that it is done well. Fons J. R. van de Vijver and Ype H. Poortinga, from Tilburg University in The Netherlands, provide an overview of the problems of test administration in a cross-cultural context and focus considerable attention on the detection of biased test items. The issue of item bias or fairness has become an important topic for test development. Currently, there appears to be a considerable amount of misunderstanding about item bias definitions and approaches for identifying it. Katherine MacRury, Philip Nagy, and Ross Traub, from the Ontario Institute for Studies in Education in Canada, provide a review of the concept of large-scale assessment programs and their expected effects. In view of the newness of large-scale programs as well as the substantial interest that exists worldwide, they also offer an agenda for further research.

Part III—*Emerging Topics*—contains three chapters. The first chapter, by Ronald Hambleton, Jac Zaal, and J. P. M. Pieters—the latter, from the University of Nijmegen in The Netherlands—introduces readers to

computerized-adaptive testing (CAT). It appears that CAT can reduce testing time and examinee frustration without any loss in measurement precision. These authors also address both hardware and software issues as well as guidelines for evaluating CAT. Robert Sternberg, from Yale University in the United States, provides an introduction to cognitive theories and their role in building new psychometric instruments. Finally, Walt Haney and George Madaus, from Boston College in the United States, review the important developments over the years that have influenced the current ethical and technical standards for test construction and test usage.

The ITC is grateful to the contributors for their willingness to give of their time, experiences, and scholarship to this book. We hope that educational researchers and psychologists around the world will find the book useful in their work.

Ronald K. Hambleton
Jac N. Zaal

INTERNATIONAL TEST COMMISSION COMMISSION INTERNATIONALE DES TESTS

Goal

The International Test Commission (ITC) is a collection of national psychological societies, test commissions, test publishers, and other organizations who use tests. ITC's goal is to assist in the exchange of testing information among member and affiliate organizations on matters pertaining to the construction, distribution, and uses of educational and psychological tests. To achieve its goal, ITC carries out the following activities on a regular basis:

- organizes symposia at international psychology meetings
- publishes a journal addressing testing issues and developments
- encourages and/or directs research studies and reviews on important testing issues and developments.

Current Projects

Presently, ITC has several projects under way which should provide important information on tests and testing practices to countries all over the world:

- The extent to which tests are adapted and used in foreign countries.
- Copyright abuse of tests.
- Survey of test legislation.
- Organization of a symposium on adapting tests for uses in different cultures at the 1990 meeting of the International Association of Applied Psychology in Kyoto, Japan.
- Collection of national society ethical and technical standards for test use.

ITC reports the findings from its projects at international meetings of psychologists, in its journal, and in other publications.

Benefits from Membership in the ITC

Several benefits accrue to members from participation in the ITC:

- communication with other organizations involved with tests to discuss problems, issues, needs, and possible solutions
- opportunity to influence the specific directions and research program of the ITC
- opportunity to contribute to a Commission whose goal is to improve the quality of testing practices around the world

History of the International Test Commission

The idea of setting up an international body as a central stimulus for better use of tests worldwide was first raised by Professor Cardinet in Switzerland in the late 1960s. Through correspondence with national societies and meetings of interested individuals during international conferences in 1969 and 1971, the initial idea for an organization began to take shape. The International Test Commission was formally established and, by 1974, 15 countries were represented at a general meeting held during the Montreal conference. At that time, the Netherlands Committee on Testing Affairs accepted responsibility for administration of the ITC, an international Council was appointed, and Dr. Ype Poortinga was elected President and Editor of the Newsletter. Although initiated with encouragement from the International Association of Applied Psychology, the ITC was established as an autonomous body—primarily because delegates to the Commission were to participate as representatives of their national societies, or test commissions, unlike the IAAP, whose membership is on an individual basis. At the 1975 IUPsyS Conference in Montreal, the Council prepared a draft constitution and by-laws which were amended and provisionally accepted in Paris in 1976, then formally ratified at the quadrennial business meeting held during the IAAP conference in Munich in 1978. The constitution of the ITC was based on the document “Recommendations Concerning the Construction, Distribution, and Use of Psychological Tests” which was approved by the IAAP general assembly in 1971.

Memberships

There are two types of membership possible in the ITC, *full* and *affiliate*. A *full* member of the ITC will be a national commission, a national psychological society, or other national body which has been recognized by the association(s) of psychologists in a country and is working towards the following objectives:

- to advance professional test development and to raise the standards of psychological tests
- to protect the public against the personal and societal consequences of the use of inadequate or inappropriate psychodiagnostic procedures and of the use of tests by unqualified persons or in a manner objectionable on scientific or ethical grounds.

Currently, 13 national societies are members of the ITC, and three other national societies have filed for membership.

Affiliate membership is open to other organizations with an interest in tests, such as test publishers, units of government, hospitals, or universities (i.e., departments) involved in psychometric work, or groups of psychologists who work with tests. Affiliate members do *not* have voting powers. Currently, there are 20 affiliate members of the ITC, with four other testing organizations having filed for membership and awaiting admission decisions.

The ITC Bulletin (Bulletin de la Commission Internationale) which is bilingual (English and French) is published biannually and sent to all full and affiliate members, and subscribers.

Application for Membership

National psychological societies, test commissions, test publishers, and major test users who are interested in joining the ITC may obtain an application for full or affiliate membership from the ITC by writing:

Ronald K. Hambleton
ITC President (1990–1994)
University of Massachusetts
Hills South, Room 152
Amherst, MA 01003
U.S.A.

The fees for full membership are related to the number of members in the national organization. For affiliate members, there is a fixed fee.

Full Members

1 to 100	members	£ 15.00	(\$ 26 U.S.)
101 to 1,000	members	£ 35.00	(\$ 60 U.S.)
1,001 to 10,000	members	£105.00	(\$180 U.S.)
10,000+	members	£225.00	(\$380 U.S.)

Affiliate Members

National or international organizations	£105.00 (\$180 U.S.)
Other organizations	£ 35.00 (\$ 60 U.S.)

Membership Applications

Applications for membership are usually confirmed at a general meeting of the ITC. The Council calls such a meeting at least once a year. Postal votes of the ITC Council Members are also used to process applications for ITC membership.

Contents

Preface	vii
International Test Commission	xi
I	
Methodological Advances	1
1	
Advances in Criterion-Referenced Measurement	3
Ronald K. Hambleton and H. Jane Rogers	
2	
Generalizability Theory	45
Dato N. M. de Gruijter and Leo J. Th. van der Kamp	
3	
Item Response Theory	69
David J. Weiss and Michael E. Yoes	
4	
Analysis of Covariance Structures	97
H. Swaminathan	
5	
Applications of Decision Theory to Test-Based Decision Making	129
Wim J. van der Linden	
6	
Meta-Analysis	157
John E. Hunter and Frank L. Schmidt	

II		
Developments in Applied Settings		185
7		
Developments in Personnel Selection Methodology		187
Robert A. Roe and Martin A. M. Greuter		
8		
Assessment Centers at the Dutch Central Government		227
Jac N. Zaal		
9		
Current Perspectives in the Assessment of the Handicapped		251
Stan Scarpati		
10		
Testing Across Cultures		277
Fons J. R. van de Vijver and Ype H. Poortinga		
11		
Reflections on Large-Scale Assessments of Student Achievement		309
Katherine MacRury, Philip Nagy, and Ross E. Traub		
III		
Emerging Topics		339
12		
Computerized Adaptive Testing: Theory, Applications, and Standards		341
Ronald K. Hambleton, Jac N. Zaal, and Jo P. M. Pieters		
13		
Cognitive Theory and Psychometrics		367
Robert J. Sternberg		
14		
The Evolution of Ethical and Technical Standards for Testing		395
Walt Haney and George Madaus		
Chapter Abstracts		427
Contributing Author Biographies		435
Index		443

I METHODOLOGICAL ADVANCES

