PERSPECTIVES ON INDIVIDUAL DIFFERENCES

Perspectives on Bias in Mental Testing

Edited by CECIL R. REYNOLDS
and ROBERT T. BROWN

Perspectives on Bias in Mental Testing

Edited by

CECIL R. REYNOLDS

Texas A&M University
College Station, Texas

and

ROBERT T. BROWN

University of North Carolina
Wilmington, North Carolina

PLENUM PRESS • NEW YORK AND LONDON

Library of Congress Cataloging in Publication Data

Main entry under title:

Perspectives on bias in mental testing.

(Perspectives on individual differences)
Includes bibliographical references and indexes.
1. Psychological tests. 2. Educational tests and measurements. 3. Test bias. I. Reynolds, Cecil R., 1952– . II. Brown, Robert T. III. Series.
BF176.P47 1984 153.9'3 84-9800
ISBN 0-306-41529-1

A Division of Plenum Publishing Corporation
233 Spring Street, New York, N.Y. 10013

Printed in the United States of America

Contributors

ERNEST M. BERNAL, School of Education and Human Development, California State University, Fresno, California

ROBERT T. BROWN, Department of Psychology, University of North Carolina, Wilmington, North Carolina

H. J. EYSENCK, University of London, London, England

ROBERT A. GORDON, Department of Sociology, Johns Hopkins University, Baltimore, Maryland

GORDON M. HARRINGTON, Department of Psychology, University of Northern Iowa, Cedar Falls, Iowa

ASA G. HILLIARD III, Georgia State University, Atlanta, Georgia

LLOYD G. HUMPHREYS, Department of Psychology, University of Illinois, Champaign, Illinois

JOHN E. HUNTER, Department of Psychology, Michigan State University, East Lansing, Michigan

REX JACKSON, Applied Educational Research, Princeton, New Jersey

ARTHUR R. JENSEN, Institute of Human Learning, University of California, Berkeley, California

WINTON H. MANNING, Educational Testing Service, Princeton, New Jersey

JANE R. MERCER, Department of Sociology, University of California, Riverside, California

JOHN RAUSCHENBERGER, ARMCO Corporation, Middletown, Ohio

CECIL R. REYNOLDS, Department of Educational Psychology, Texas A & M University, College Station, Texas

FRANK L. SCHMIDT, U.S. Office of Personnel Management and George Washington University, Washington, D.C.

Preface

The cultural-test-bias hypothesis is one of the most important scientific questions facing psychology today. Briefly, the cultural-test-bias hypothesis contends that all observed group differences in mental test scores are due to a built-in cultural bias of the tests themselves; that is, group score differences are an artifact of current psychometric methodology. If the cultural-test-bias hypothesis is ultimately shown to be correct, then the 100 years or so of psychological research on human differences (or differential psychology, the scientific discipline underlying all applied areas of human psychology including clinical, counseling, school, and industrial psychology) must be reexamined and perhaps dismissed as confounded, contaminated, or otherwise artifactual. In order to continue its existence as a scientific discipline, psychology must confront the cultural-test-bias hypothesis from the solid foundations of data and theory and must not allow the resolution of this issue to occur solely within (and to be determined by) the political *Zeitgeist* of the times or any singular work, no matter how comprehensive.

In his recent volume *Bias in Mental Testing* (New York: Free Press, 1980), Arthur Jensen provided a thorough review of most of the empirical research relevant to the evaluation of cultural bias in psychological and educational tests that was available at the time that his book was prepared. Nevertheless, Jensen presented only one perspective on those issues in a volume intended not only for the scientific community but for intelligent laypeople as well. What is needed is a more thorough analysis of the issues of bias in mental testing that is written from a variety of perspectives, including interdisciplinary ones. Further, the field has been anything but static in recent years: Much research has been published since Jensen completed his book; new theoretical and conceptual issues have been raised; and new areas involving test bias have arisen, such as those concerning tests published by the Educational Testing Service. Finally, the issue of test bias has hardly been resolved in either its

scholarly or its applied forms, as indicated by the opposing judicial verdicts handed down in the cases of *Larry P.* v. *Riles* and *PASE* v. *Hannon.*

This volume presents the views of several prominent authors in the fields of cultural test bias specifically and of individual differences generally. We have attempted to bring together proponents of the major views currently available, although space limitations have obviously prevented the presentation of all positions or all possible contributors. We have, however, tried to justify the use of the word *perspectives:* Our authors are from different academic disciplines, cultural backgrounds, and employment settings. They represent scholars, test publishers, private practitioners, and psychologists in private industry. Further, they have employed a variety of research approaches in addressing the issue of bias, from complex correlational approaches to those that are complex experimental. In dealing with this area, there are few simple approaches. Each author deals with a topic of specific relevance to a full evaluation of the cultural-test-bias hypothesis. Competing views have been purposely presented so as to make clear where the areas of disagreement lie. The request to the authors was simple and straightforward: They were asked to focus on the empirical, scientific evaluation of the cultural-test-bias hypothesis and to avoid emotional issues. The responses to this request, the authors' chapters, are presented with minimal editorial alteration, so that readers can clearly see the positions of experts in the field at this time and can find the basis for an evaluation of the viability of their views.

Some notes on the organization and content of the book follow. The first two chapters provide background information. The first chapter, an introduction to the issues by the editors, is an overview of the scientific and emotional issues surrounding the test bias hypothesis and provides a historical perspective that acts as a context for the rest of the book. The chapter by Hunter, Schmidt, and Rauschenberger is a sophisticated presentation of the methodological and statistical considerations that may lead to problems in interpreting research in the area, as well as of the ethical considerations that must guide that research.

Chapters 3 through 10 present specialized areas of research or topics of interest relevant to test bias. As can be seen, these chapters cover a diversity of topics from a diversity of perspectives, and by design, there has been no attempt to bring unity or any particular organization to them. They well represent the variety of methodological and theoretical approaches that characterizes the field at this

time. Harrington, in Chapter 3, offers an unusual perspective on test bias: a true experimental model that has attempted, through the use of animal subjects, to test the assumptions that underlie standardized psychological tests. As might be expected, his approach is not universally accepted, but it is certainly thought provoking. Bernal, in Chapter 5, describes a new research project using different types of instructions to evoke different levels of performance on intelligence tests by different cultural groups. Manning and Jackson discuss the charges of test bias that have been leveled against tests produced by the Educational Testing Service and describe research that they feel successfully answers those charges. Eysenck, in Chapter 8, faces head-on one of the issues that underlies the current interest in the cultural-test-bias hypothesis: racial differences in intelligence test performance.

The other chapters are more theoretically oriented. Hilliard (Chapter 4) and Mercer (Chapter 9) criticize present intelligence tests from, respectively, a black and a sociological perspective. Both believe that the cultural-test-bias hypothesis is basically correct. Gordon (Chapter 10) provides a sociological perspective that is not only opposed to but highly critical of Mercer's. Humphreys (Chapter 7) discusses the issue of test bias in the larger context of the theoretical construct of general intelligence. In the last chapter, Jensen updates his position as presented in *Bias in Mental Testing* and critiques the other chapters in this volume. Among his criticisms is a statistical reevaluation of Bernal's results that leads to questions concerning Bernal's interpretation. In a postscript at the end of the book, Bernal replies to Jensen's criticism.

We suffer under no illusion that the test bias issue will be resolved to the satisfaction of all for some time to come. Our hope is that this book will clarify the positions and the supporting data of various camps. We hope that the volume will expose areas where little evidence has been gathered and where additional research is needed. We hope additionally that the volume will demonstrate that even areas of considerable emotional impact and political and social implications will benefit from scientific evaluation.

There are many to whom we owe thanks for assistance in the preparation of this book. We appreciate the cooperation of the authors in the timely submission of their contributions. The cooperation and the encouragement of Leonard R. Pace, formerly of Plenum Press, has helped to smooth the development of this volume and the series on individual differences of which it is the initial volume. The patience and the support of our wives, Brenda and Sue,

and our families has contributed in a major way to the successful completion of this book. Finally, the invaluable assistance of the office staff at University of North Carolina at Wilmington (Martha Jo Clemmons and Eleanor Martin), University of Nebraska at Lincoln (Linda Weber), and Texas A & M University (Lori Powell) is gratefully acknowledged.

CECIL R. REYNOLDS
ROBERT T. BROWN

Contents

CHAPTER ONE

Bias in Mental Testing

An Introduction to the Issues

CECIL R. REYNOLDS
and
ROBERT T. BROWN

> *Public controversy deals with stereotypes, never in subtleties. The Luddites can smash up a device, but to improve a system requires calm study. . . . Sound policy is not for tests or against tests, but how tests are used. But all the public hears is endless angry clamor from extremists who see no good in any test, or no evil.*
>
> Cronbach, 1975, p. 1

Cultural bias in mental testing has been a recurring issue since the beginning of the testing movement itself. From Binet to Jensen, many professionals have addressed the problem, with varying and inconsistent outcomes. Unlike the pervasive and polemical nature–nurture argument, the bias issue has been until recently largely restricted to the professional literature, except for a few early discussions in the popular press (e.g., Freeman, 1923; Lippmann, 1923a,b). Of some interest is the fact that one of the psychologists who initially raised the question was the then-young Cyril Burt (1921), who even in the 1920s was concerned about the extent to which environmental and motivational factors affect performance on intelligence tests. Within the last few years, however, the question of cultural test bias

CECIL R. REYNOLDS • Department of Educational Psychology, Texas A & M University, College Station, Texas 77843. ROBERT T. BROWN • Department of Psychology, University of North Carolina, Wilmington, North Carolina 28401.

has burst forth as a major contemporary problem far beyond the bounds of scholarly academic debate in psychology. For approximately the last decade, the debate over bias has raged in both the professional and the popular press (e.g., Fine, 1975). Entangled in the larger issues of individual liberties, civil rights, and social justice, the bias issue has become a focal point for psychologists, sociologists, politicians, and the public. Increasingly, the issue has become a political and legal one, as reflected in numerous court cases and passage in the state of New York and consideration elsewhere of "truth-in-testing" legislation. The magnitude—and the uncertainty—of the controversy and its outcome is shown in two highly publicized recent federal district court cases. The answer to the question "Are the tests used for pupil assignment to classes for the educably mentally retarded biased against cultural and ethnic minorities?" was "Yes" in California (*Larry P.*, 1979) and "No" in Chicago (*PASE*, 1980).

Unfortunately, we are all prisoners of our language. The word *bias* has several meanings, not all of which are kept distinct. In relation to the present issue, bias as "partiality towards a point of view or prejudice" and bias as "a statistical term referring to a constant error of a measure in one specific direction as opposed to random error" frequently become coalesced. If the latter meaning did not drag along the excess baggage of the former, the issue of bias in mental testing would be far less controversial and emotional than it is. However, as indicated in the *Oxford English Dictionary*, bias as partiality or prejudice can be traced back at least to the sixteenth century and clearly antedates the statistical meaning. We are likely ourselves to be biased toward the meaning of partiality whenever we see the word *bias*. The discussion of bias in mental testing as a *scientific* issue should concern only the statistical meaning: whether or not there is systematic error in the measurement of mental ability as a function of membership in one or another cultural or racial subgroup (Reynolds, 1982b).

As would be expected, Jensen's *Bias in Mental Testing* (1980a) has hardly stilled the controversy. Since his now-classic "How Much Can We Boost IQ and Scholastic Achievement?" (1969), virtually anything with Jensen's name on it has had earthquake impact, some publications evoking higher Richter-scale readings than others. Whatever its actual effect, *Bias in Mental Testing* was intended to be a scholarly presentation of the statistical concepts and the empirical research concerning test bias. After an exhaustive evaluation of the literature, Jensen concluded that the charge of bias in mental tests could not be substantiated. Although virtually all scholars respect

Jensen's scholarship and integrity, by no means do all accept his conclusions. Other recent independent, extensive reviews of the empirical literature have also generally come to the same conclusion as has Jensen (e.g., Reynolds, 1982a; Vernon, 1979). Jensen's book has been subjected to peer evaluation not only in multiple individual book reviews, but in a recent issue of *Behavioral and Brain Sciences* (1980b), in which his position was strongly supported by some professionals and just as strongly attacked by others. The present book is designed to provide a more extensive presentation of a variety of views on bias than is available elsewhere. It is not intended as a critique of Jensen, but as an original scholarly contribution to the literature on bias. For that reason, authors have been chosen who have a variety of opinions on the issue, but who, for the most part, share an interest in an empirical as opposed to an emotional resolution of the issue. Unfortunately, it was not possible to include all those qualified to write on the issue. As the acknowledged authority in the field, Jensen was invited to comment on the other papers.

At the outset, we would like to make our position as editors clear: In order to maintain its scientific integrity, psychology must treat the cultural-test-bias hypothesis as a serious scientific question, to be argued and resolved on the basis of empirical research and theory rather than on one or another of our many politicosocial *Zeitgeister*. The history of genetics under T. D. Lysenko in Stalinist Russia should be a warning to us all of the danger of political resolutions to scientific questions. However impossible full objectivity is in science (e.g., Kuhn, 1962) and however much we are trapped by our metatheoretical views, as well as by our language, we must view all socially, politically, and emotionally charged issues from the perspective of rational scientific inquiry. Moreover, we must be prepared to accept scientifically valid findings as real—even if we do not like them. Otherwise, we risk psychology's becoming an impotent field whose issues will be resolved not by scholars in courts of scientific inquiry but by judges in courts of law, and whose practitioners will be viewed as charlatans with opinions of no more validity than the claims of patent-medicine-pushing quacks. Further, it behooves us as social scientists to be aware of and sensitive to the historical perspective from which some concerned groups may view an issue.

The need for an objective answer to the question "Is there bias in mental testing?" is as manifest as the need to eliminate any such bias found.

THE CONTROVERSY: WHAT IT IS AND WHAT IT IS NOT

Systematic group differences on standardized intelligence and aptitude tests occur as a function of socioeconomic level, race or ethnic background, and other demographic variables. Black-white differences on IQ measures have received extensive investigation over the past 50 or 60 years. The preponderance of these studies has been reviewed by Shuey (1966) and Tyler (1965). Jensen (1980a) and Willerman (1979) have reviewed several more recent studies. Although the results occasionally differ slightly, depending on the age groups under consideration, random samples of blacks and whites show a mean difference of about one standard deviation, with the mean score of the whites consistently exceeding that of the black groups. The differences have persisted at relatively constant levels for quite some time and under a variety of methods of investigation. When a number of demographic variables are taken into account (most notably socioeconomic status), the size of the mean black-white difference reduces to .5 to .7 standard deviations (e.g., Kaufman, 1973; Kaufman & Kaufman, 1973; Reynolds & Gutkin, 1981) but remains robust in its appearance. All studies of racial and ethnic group differences on ability tests do not show higher levels of performance by whites. Although not nearly as thoroughly researched as black-white groups, Oriental groups have been shown to perform consistently as well as, or better than, white groups (Pintner, 1931; Tyler, 1965; Willerman, 1979). Depending on the specific aspect of intelligence under investigation, other racial and ethnic groups show performance at or above the performance level of white groups. There has been argument over whether any racial differences in intelligence are real or even researchable (e.g., Schoenfeld, 1974), but the reliability across studies is very high, and the existence of the differences is now generally accepted. It should always be kept in mind, however, that the overlap among the distributions of intelligence test scores for the different races is much greater than the degree of differences between the various groups. There is always more within-group variability than between-group variability in performance on mental tests. The differences are nevertheless real ones and are unquestionably complex (e.g., Reynolds & Jensen, 1983).

The issue at hand is the explanation of those group differences. It should be emphasized that both the lower scores of some groups and the higher scores of others need to be explained, although not necessarily, of course, in the same way. The problem was clearly

stated by Eells in his classic study of cultural differences (Eells, Davis, Havighurst, Herrick, & Tyler, 1951):

> Do the higher test scores of the children from high socioeconomic backgrounds reflect genuine superiority in inherited, or genetic, equipment? Or do the high scores result from a superior environment which has brought about real superiority of the child's "intelligence"? Or do they reflect a bias in the test materials and not any important difference in the children at all? (p.4)

Eells *et al.* also concisely summarized the cultural-test-bias hypothesis as it applied to differences in socioeconomic status (SES):

> If (a) the children from different social-status levels have different kinds of experiences and have experiences with different types of material, and if (b) the intelligence tests contain a disproportionate amount of material drawn from the cultural experiences with which pupils from the higher social-status levels are more familiar, one would expect (c) that children from the higher social-status levels would show higher I.Q.'s than those from the lower levels. This argument tends to conclude that the observed differences in pupil I.Q.'s are artifacts dependent upon the specific content of the test items and do not reflect accurately any important underlying ability in the pupils. (p. 4)

Eells was aware that his descriptions were oversimplifications and that it was unlikely that all of the observed group differences could be explained by any one of the three factors alone. More recently, Loehlin, Lindzey, and Spuhler (1975) concluded that all three factors were probably involved in racial differences in intelligence. In its present, more complex form, the cultural-test-bias hypothesis itself considers other factors than culture-loaded items, as will be seen below. But the basics of Eell's summary of the cultural-test-bias hypothesis still hold: Group differences stem from characteristics of the tests or from aspects of test administration. Because mental tests are based largely on middle-class white values and knowledge, they are more valid for those groups and are biased against other groups to the extent that these groups deviate from those values and knowledge bases. Thus, ethnic and other group differences result from flawed psychometric methodology and not from actual differences in aptitude (Harrington, 1975, 1976). As is described in some detail below, this hypothesis reduces to one of differential validity, the hypothesis of differential validity being that tests measure intelligence more accurately and make more valid predictions about the level of intellectual functioning for individuals from the groups on which the tests are mainly based than for those from other groups. Artifactually low scores on an aptitude test could

lead to pupil misassignment to educational programs and unfair denial of admission to college, graduate school, or other programs or occupations in which such test scores are an important decision-making component. This is the issue over which most of the recent court cases have been fought. Further, there would be dramatic implications for whole areas of psychological research and practice if the cultural-test-bias hypothesis is correct: The principal research of the last century in the psychology of human differences would have to be dismissed as confounded and largely artifactual because much of the work is based on standard psychometric theory and testing technology. The result would be major upheavals in the practice of applied psychology, as the foundations of clinical, school, counseling, and industrial psychology are strongly tied to the basic academic field of individual differences. The issue, then, is one of the most crucial facing psychology today (Reynolds, 1980c).

On the other hand, if the cultural-test-bias hypothesis is incorrect, then group differences are not attributable to the tests and must be due to one of the other factors mentioned by Eells *et al.* (1951) or to some combination of them. That group differences in test scores reflect real group differences in ability should be admitted as a possibility, and one that calls for scientific study.

The controversy over test bias should not be confused with that over the etiology of any obtained group differences in test scores. Unfortunately, it has often seen inferred that measured differences themselves indicate genetic differences, and therefore the genetically based intellectual inferiority of some groups. Jensen has himself consistently argued since 1969 that mental tests measure, to a greater or lesser extent, the intellectual factor *g*, which has a large genetic component, and that group differences in mental test scores may then reflect group differences in *g*. Unless one reads Jensen's statements carefully, it is easy to overlook the many qualifications that he makes regarding these differences and conclusions.

But, in fact, Jensen or anyone else's position on the basis of actual group differences should be seen as irrelevant to the issue of test bias. However controversial, etiology is a separate issue. It would be tragic to accept the cultural-test-bias hypothesis as true if it is, in fact, false. In that case, measured differences would be seen as not real, and children might be denied access to the educational environment best suited to them. Further, research on the basis of any group differences would be stifled, as would implementation of programs designed to remediate any deficiences. Further, from our perspective, the most advantageous position for the true white racist and