

CONTENTS

Fairness in Employment Testing

Foreword

- Linda S. Gottfredson and James C. Sharf. *Fairness in Employment Testing* 225

Introduction

- Lillian M. Avery. *Origins of and Reactions to the PTC Conference on Fairness in Employment Testing* 231

Major Contributions

- James C. Sharf. *Litigating Personnel Measurement Policy* .. 235
Frank L. Schmidt. *The Problem of Group Differences in Ability Test Scores in Employment Selection* 272
Linda S. Gottfredson. *Reconsidering Fairness: A Matter of Social and Ethical Priorities* 293
Clint Bolick. *Legal and Policy Aspects of Testing* 320
Richard T. Seymour. *Why Plaintiffs' Counsel Challenge Tests, and How They Can Successfully Challenge the Theory of "Validity Generalization"* 331
W. B. Allen. *Rhodes Handicapping, or Slowing the Pace of Integration* 365

Commentaries

- I. A. Ryanen. *Commentary of a Minor Bureaucrat* 379
V. Jon Bentz. *Comments on Papers Concerning Fairness in Employment Testing* 388
Henry M. Levin. *Issues of Agreement and Contention in Employment Testing* 398
John H. Bishop. *Employment Testing and Incentives to Learn* 404
Robert A. Gordon, Mary A. Lewis, and Ann M. Quigley. *Can We Count on Muddling through the g Crisis in Employment?* 424
Barry L. Goldstein and Patrick O. Patterson. *Turning Back the Title VII Clock: The Resegregation of the American Work Force through Validity Generalization* 452
Robert J. Delahunty. *Perspectives on Within-Group Scoring* 463

- Author Index for Volume 33 478
Cumulative Subject Index for Volumes 32-33 480

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission from the copyright owner.

The appearance of the code at the bottom of the first page of an article in this journal indicates the copyright owner's consent that copies of the article may be made for personal or internal use, or for the personal or internal use of specific clients. This consent is given on the condition, however, that the copier pay the stated per copy fee through the Copyright Clearance Center, Inc. (27 Congress Street, Salem, Massachusetts 01970), for copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to other kinds of copying, such as copying for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale. Copy fees for pre-1988 articles are as shown on the article title pages; if no fee code appears on the title page, the copy fee is the same as for current articles.

0001-8791/88 \$3.00

MADE IN THE UNITED STATES OF AMERICA

JOURNAL OF VOCATIONAL BEHAVIOR
(ISSN 0001-8791)

Published bimonthly by Academic Press, Inc.,
1 East First Street, Duluth, MN 55802

1988: Volumes 32-33. Price: \$170.00 U.S.A. and Canada; \$206.00 all other countries
1989: Volumes 34-35. Price: \$180.00 U.S.A. and Canada; \$225.00 all other countries
All prices include postage and handling.

Information concerning privileged subscription rates to individuals may be obtained by writing to the Publishers. All correspondence and subscription orders should be addressed to the office of the Publishers at 1 East First Street, Duluth, MN 55802. Send notices of change of address to the office of the Publishers at least 6 to 8 weeks in advance. Please include both old and new address. POSTMASTER: Send changes of address to *Journal of Vocational Behavior*, 1 East First Street, Duluth, MN 55802.

Second class postage paid at Duluth, MN, and at additional mailing offices.

Copyright © 1988 by Academic Press, Inc.

FOREWORD

Fairness in Employment Testing

LINDA S. GOTTFREDSON¹

Department of Educational Studies, University of Delaware

AND

JAMES C. SHARF¹

Washington, D.C.

This special issue of the *Journal of Vocational Behavior* (JVB), *Fairness in Employment Testing*, focuses on the fairness of using objective tests in hiring workers. It follows up the December 1986 special issue of the JVB [L. S. Gottfredson (1986). The *g* Factor in Employment [Special issue]. *Journal of Vocational Behavior*, 29(3)]. The 1986 issue focused on the role of ability tests, and of one ability factor in particular (*g*, or the general mental ability factor), in predicting performance across jobs. Like that earlier issue, the current one is based on a catalytic conference held in the preceding year by the Personnel Testing Council (PTC) of Southern California. As described by Lillian Avery, the conference coordinator, the 1987 PTC conference was stimulated directly by local and national reactions to the earlier 1985 conference and the 1986 journal issue stemming from it. The agendas of both conferences reveal an acute sense on the part of the personnel measurement profession of where the leading edge of employment testing controversy was located at those particular moments in history.

THE CHANGING NATURE OF DEBATE ABOUT EMPLOYMENT TESTING

Debate about fairness in employment testing is evolving in potentially momentous ways as a result of recent developments in law and science and as a consequence of experience with current equal employment

¹ Guest Editor.

Please address reprint requests to Linda S. Gottfredson, Department of Educational Studies, University of Delaware, Newark, DE 19716.

opportunity enforcement policy. Because employment selection plays such a vital role in the distribution of society's rewards and responsibilities, the change that is occurring in the employment arena may be foreshadowing a major shift in the larger social-political debate about fairness in a democracy.

This special issue of the JVB describes some of those recent developments in employment testing—including research and litigation concerning “validity generalization,” the Supreme Court's *Watson* decision, and the imminent report from the National Research Council (NRC) Committee on the General Aptitude Test Battery. It also provides a retrospective view of personnel selection research and law since passage of the Civil Rights Act of 1964. But more importantly, the issue brings into sharp relief the changing nature and broad importance of current debates over fairness in employment testing. As noted in several contributions here, testing research and policy have been driven in the last two decades by group differences in scores on employment tests. Early debates focused on relatively limited questions such as how many tests are biased against minority groups, how tests can be improved if they are biased, and what evidence is necessary for showing that a test is job related and therefore permissible under *Griggs* when the test has adverse impact. Now the debate is shifting to address the question of what to do when minorities differ on tests that *meet* the earlier criteria for fairness. What is the meaning of fairness in this new context? Clearly, this question is a most troubling one because it suggests that certain of our fundamental values may be in conflict and so it compels us to consider what kind of society we in the United States wish ours to be and to become.

If there is one single overarching debate among the contributors to this special issue, it is directed toward this question: How deeply troubled should we be over the policy question of what to do when valid tests have adverse impact? Some view this situation as being extremely serious, even contending that group differences in ability constitute a social crisis, labeled in one commentary as the “*g* crisis.” Others take the question seriously but attach less significance to group differences in the long run because they believe them to be diminishing or disappearing. Some contributors would dismiss the question entirely either because they believe that most tests are biased or because they are not convinced that tests are useful for improving productivity even if not biased. Clearly, these are quite different perspectives on essentially the same question and they cannot all prove correct.

CONTRIBUTORS

All six speakers at the 1987 PTC conference have contributed articles to this volume. Six commentaries on these articles have been solicited

from people of diverse perspectives in order to extend the debate and broaden the inquiry. A seventh commentary by Robert Delahunty, an attorney in the Civil Rights Division of the U.S. Department of Justice, is a letter from the U.S. Justice Department to the NRC Committee on the General Aptitude Test Battery. That letter, which discusses the legality of race-norming, is included here because of the importance of that issue in the present context and because of the potential influence of both the Justice Department and the NRC Committee on future uses of employment tests.

Contributors were chosen for their divergent perspectives as well as for their expertise, and they include major protagonists in current personnel testing research and litigation. Frank Schmidt, for example, and his co-workers have revolutionized personnel testing theory through their work on validity generalization. Richard Seymour of the Lawyers' Committee for Civil Rights under Law is undoubtedly one of the nation's most influential civil rights attorneys in the employment discrimination arena. Lawyers from the NAACP Legal Defense and Educational Fund (Barry Goldstein and Patrick Patterson) and from the U.S. Justice Department (Clint Bolick and Robert Delahunty) are also represented, as is a member (and now chair) of the U.S. Commission on Civil Rights (William Allen). In addition to several industrial psychologists (Mary Lewis, Ann Quigley, Frank Schmidt, and James Sharf), the contributors include personnel professionals who have directed personnel research and policy in two large organizations, one for a 30,000-member workforce in the public sector (I. A. Ryanen) and one for a 400,000-member workforce in the private sector (V. Jon Bentz). Insights from several of the other relevant disciplines dealing with man and society are provided by labor economists John Bishop and Henry Levin, sociologists Robert Gordon and Linda Gottfredson, and political scientist William Allen.

CONTENTS OF THE INTRODUCTION AND MAJOR CONTRIBUTIONS

Avery describes how the 1987 PTC conference, “Fairness in Employment Testing,” originated in the concerns over group fairness and other unresolved issues that were raised by the 1985 PTC conference, “The *g* Factor in Employment Testing.” Her survey of reactions to the 1987 conference indicated that, regardless of their perspectives on the difficult social issues raised, personnel professionals are restive that these issues have not been addressed fully in an open and informed debate.

Sharf reviews the history and application of Title VII of the Civil Rights Act of 1964 in order to argue that the adverse impact definition of test discrimination embodied in *Griggs* and its progeny is the product of successful advocacy of unelected bureaucrats rather than a product of public deliberation. He argues further that “hair trigger” application

of the adverse impact definition ignores skill deficiencies among certain minorities and guarantees that even the use of valid tests will be litigated. He suggests that validity generalization may be a general rebuttal to *prima facie* presumptions of discrimination based on adverse impact.

Drawing heavily from validity generalization research, Schmidt summarizes evidence to dispute the frequently proposed theories that objective employment tests are either biased against minorities or lack utility. He argues that group differences in test scores do not result from deficiencies in the tests but in the skills they measure. He suggests that measurement-based research does not provide the solution to the problem of group differences, but that it does reveal its nature more clearly.

Gottfredson argues on the basis of earlier research on the importance of *g* and black-white differences in *g* that many valid, unbiased tests can be expected to produce high levels of adverse impact when used in a race-neutral manner, especially in high-level jobs. She argues further that unrealistic expectations regarding racial parity often lead employers to adopt practices that are expedient in the short-term but which may in the long run compromise the integrity of their personnel systems, lower worker productivity, and stigmatize minorities.

Like Sharf, Bolick argues that Title VII has been turned into its opposite, namely, a retreat from nondiscrimination, and he counsels "a more reasonable approach to determining job-relatedness." Bolick suggests that quota systems, race-norming of tests, and other forms of preferential treatment, even though they may sometimes constitute low-cost insurance against litigation, "bode devastating real-world consequences for the purported beneficiaries." In line with his claim that fresh approaches are needed, he suggests some specific policy initiatives for improving economic opportunities for minorities.

Seymour challenges the claims and assumptions of the foregoing four contributors regarding the legal history, validity, utility, and fairness of employment tests, and argues that tests "can be an engine for the exclusion of racial minorities more permanent and thorough than individual ill will." He argues that industrial psychologists have overlooked evidence that many tests are biased and that their more ambitious claims for validity generalization are based on faulty science. He outlines what plaintiff's counsel looks for in deciding to try a testing case, and he provides a primer on how to challenge validity generalization.

Drawing on examples from college athletics (specifically NCAA Proposition 48), education, and personal experience, Allen describes the racial unfairness both of using scholastic tests where they should *not* be used and of not using them where they *should* be. Both can deny opportunities to demonstrate relevant talent, and thus both also constitute a rejection of the merit principle. He suggests that greater consideration be given to the reasons for administering or withholding tests and whether such action is appropriate in particular situations.

THEMES AMONG THE COMMENTARIES

The commentators voice agreement or disagreement, sometimes in unpredictable patterns, with various themes in the major contributions. Many commentators note that Title VII and its subsequent interpretation and application in the Uniform Guidelines and Supreme Court cases have redirected both personnel practice and research, primarily in order to eliminate racial differences. They disagree about whether this has had salutary effects (for example, Goldstein and Patterson arguing that it has been essential for progress toward racial equality) or not (for example, Ryanen suggesting that it has stultified personnel selection research and led to new forms of racial discrimination and patronage in public institutions). The commentaries thus echo disagreements among the major contributors about whether the legal context of testing is sufficiently "rational" (Seymour and Goldstein and Patterson) or not (Sharf, Bolick, and Ryanen). Ryanen and Bentz also provide special insight into how the legal context has affected the work of personnel professionals.

As a group, the commentators devote considerable attention to test validity and utility—the *sine qua non* for "job-relatedness" and hence a critical issue in discussions of test fairness. In their commentary, Gordon, Lewis, and Quigley dispute Seymour's claim that he successfully demonstrates pervasive test bias against blacks in the General Aptitude Test Battery (GATB), and they reanalyze his data to argue that the phenomena Seymour describes have no relation to race and can be observed, for example, between whites of different average ability levels. Levin, like Seymour and Goldstein and Patterson, argues that not enough is known about the utility of tests to justify their widespread use. He also urges industrial psychologists to adopt a more sophisticated view of the nature and determinants of economic productivity when estimating utility. Bishop, another labor economist, does not question the validity or utility of tests, which he argues is considerable, but he argues that employment tests that assess material taught in high school can also promote productivity through the incentives they provide for prospective workers to develop their own skills.

The social context and consequences of testing practices also receive considerable attention in the commentaries. Like Gottfredson, Gordon et al. speak of a miseducation of the public with regard to testing and ability differences, which hinders the development of more rational public policy. Consistent with Seymour, Goldstein and Patterson argue that affirmative action has improved minority employment whereas validity generalization would "reestablish the old patterns of racial segregation and hierarchy," and that there is no compelling evidence either that affirmative action depresses productivity or that validity generalization enhances it. Race-norming, which is a way of adjusting test scores to obtain equal hiring rates when groups differ in test scores, is acceptable to Bishop because he believes that it would detract little from productivity.

In contrast, Delahunty presents the case that race-norming as practiced by the U.S. Employment Service is both statutorily suspect and unconstitutional. Ryanen, like Bolick and Gottfredson, argues that using test scores in such a race-conscious manner harms blacks more than it helps them. Bentz objects to the adversarial nature of the testing context, while Gordon et al. suggest that adversarial debate is essential when ethical decisions are at issue. Goldstein and Patterson paint very starkly the developing collision between validity generalization and adverse impact case law, which Sharf also discusses. The specter of the winner-take-all outcome that they imply, with its devastating losses for blacks, is bleaker than the scenario envisioned by Gottfredson and Schmidt, who discuss how the search for valid supplements to cognitive tests in the selection process can improve productivity while reducing adverse impact to some extent. And Bentz calls on psychologists to pursue new methodologies for achieving equity and efficiency, as well as for encouraging a "flowering of intelligence."

Another major, but implicit, theme running through both the commentaries and the major contributions bears attention. That theme can be characterized as one of crisis and frustration—of things going wrong, of institutions in jeopardy, of the betrayal of deeply held values, including that of scientific objectivity. Ryanen and Gordon et al., echoing especially Sharf and Gottfredson but also Allen and Bolick, speak of dishonest social systems and the irrational application of law, of declining productivity and worsening race relations, and of the repudiation of merit. Echoing Seymour, Goldstein and Patterson fear massive backsliding on the road to racial equality as a result of the quick acceptance of what they view as bad science. With his remark that we are "on the horns of a scientific and social dilemma," Bentz suggests that this sense of crisis may proceed from increasingly keenly felt conflicts among fundamental principles with regard to group equality, individual equity, and economic productivity. The poignancy of Bentz's piece also suggests that it is less clear now to many people what it means to be "fair."

LOOKING AHEAD

This special issue does not provide the answer to the question of what is fair in employment testing. It makes clear, however, that the issue of fairness is not a scientific question; rather, it is a matter of political philosophy, and hence a question that requires thoughtful reflection and public debate. By diagnosing problems and analyzing the implications of suggested solutions, science can, however, encourage more informed judgment among judges, lawyers, personnel professionals, politicians, and others who influence personnel selection policy. We hope that this volume encourages both the research and the debate that are essential for creating employment policy that is both effective and fair.

Received: November 2, 1988

INTRODUCTION

Origins of and Reactions to the PTC Conference on Fairness in Employment Testing

LILLIAN M. AVERY

California Test Development Field Center, Inglewood, California

Since its organization over 30 years ago, the Personnel Testing Council of Southern California (PTC) has served as a forum for professionals in the field of personnel testing, selection, and research, through published materials, workshops, and semiannual conferences designed to provide information on current practices, new ideas and techniques, legal developments, and solutions to testing and selection problems. This article describes the perceptions and reactions to the 1987 PTC conference on Fairness in Employment Testing, which had its origins in the 1985 PTC conference on The *g* Factor in Employment Testing. As with the *g* conference, the Fairness in Employment Testing conference stimulated further discussion and raised additional concerns and issues. Audience reactions to the conference are also described. © 1988 Academic Press, Inc.

In October, 1985, the Personnel Testing Council of Southern California (PTC) presented a conference on The *g* Factor in Employment Testing. That conference brought together leading researchers, including Arthur R. Jensen, Robert L. Thorndike, John E. Hunter, James Crouse, Linda S. Gottfredson, and John Hawk, for the purpose of summarizing recent research on the nature of *g* (general intelligence) and the relevance of mental tests in personnel selection. Papers presented at that conference, along with commentaries contributed by recognized experts in mental testing, Leona Tyler, Richard Arvey, Lloyd Humphreys, and Robert Linn, were published in a special issue of the *Journal of Vocational Behavior* in December, 1986.

Although the 1985 conference on *g* brought together a wealth of evidence and theory about the role of *g* and the implications surrounding the use of *g*-loaded tests in educational and employment settings, discussion was

The author served as coordinator for the 1987 Fall Conference of the Personnel Testing Council of Southern California, *Fairness in Employment Testing*, Newport Beach, California, October 1-2, 1987. Requests for reprints should be addressed to Lillian M. Avery, 1015 Hedgepath Avenue, Hacienda Heights, CA 91745.