Interim Report: Within-Group Scoring of the General Aptitude Test Battery (1988)

Pages
76

Size
5 x 8

ISBN
0309311357

Alexandra K. Wigdor and John A. Hartigan, Editors; Committee on the General Aptitude Test Battery; Commission on Behavioral and Social Sciences and Education; National Research Council

Interim Report : Within-Group Scoring of the General Aptitude Test Battery

Alexandra K. Wigdor and John A. Hartigan, *editors*

Committee on the General Aptitude Test Battery
Commission on Behavioral and Social Sciences and Education
National Research Council

NATIONAL ACADEMY PRESS
Washington, D.C. 1988

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

This report has been reviewed by a group other than the authors according to procedures approved by a Report Review Committee consisting of members of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine.

The National Academy of Sciences is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Frank Press is president of the National Academy of Sciences.

The National Academy of Engineering was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Robert M. White is president of the National Academy of Engineering.

The Institute of Medicine was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Samuel O. Thier is president of the Institute of Medicine.

The National Research Council was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Frank Press and Dr. Robert M. White are chairman and vice chairman, respectively, of the National Research Council.

The work of the Committee on the General Aptitude Test Battery is sponsored by the Employment and Training Administration and funded under U.S. Department of Labor Contract 99-7-3239-98-101-01.

Available from:
Committee on the General Aptitude Test Battery
National Research Council
2101 Constitution Avenue N.W.
Washington, D.C. 20418

Printed in the United States of America

COMMITTEE ON THE GENERAL APTITUDE TEST BATTERY

Contents

Introduction

THE CONTROVERSY

Officials of the public Employment Service are experimenting with a new referral system that uses the General Aptitude Test Battery (GATB) to screen job seekers for virtually all of the jobs handled through state Employment Service offices. With the encouragement and support of the U.S. Employment Service (USES), a part of the U.S. Department of Labor, many states have undertaken pilot studies of the new plan for assessing registrants and identifying which ones to refer to a particular employer. As an integral part of the plan, scores on the General Aptitude Test Battery are being converted to percentile ranks within the population categories of "black," "Hispanic," and "other" (which includes all of those not in the first two categories).

USES adopted the within-group percentile scoring device in 1981 when it began pilot testing the GATB-based referral system. It was a considered decision. In order to get the maximum economic benefit of a purportedly ability-based referral system, USES designed the system to rank order candidates by test score and to refer them from the top down. However, strict top-down referral would adversely affect the employment chances of black and Hispanic applicants, who as groups tend to score consistently lower than majority group members on standardized ability tests. To counteract this effect, the new referral system stipulates the conversion of raw scores into within-group percentile ranks, so

that the reported score reflects an applicant's standing with reference to his or her own racial or ethnic group, thus effectively erasing average group differences in test scores. A black applicant with a percentile score of 50 has the same ranking for referral as a white candidate with that percentile score, although their raw test scores would be very different. For example, in the category of semiskilled jobs, blacks at the 50th percentile have raw scores of 276; Hispanics, 295; and others, 308. Within the black group, a raw score of 308 is at the 84th percentile.

By combining this method of computing percentile scores with top-down selection of the applicants to be referred to prospective employers, USES sought to reach a workable compromise between an individual employer's interest in hiring the most able workers available and the need to meet federal equal employment opportunity and affirmative action goals. Without some sort of compensatory scoring system, in the agency's view, referral of candidates on the basis of GATB test scores from the top down would reduce the employment opportunities of minority group candidates and cause adverse impact problems for both the Employment Service and employers. But if top-down selection were completely abandoned, in the agency's view, workforce efficiency would suffer.

On November 10, 1986, Wm. Bradford Reynolds, Assistant Attorney General for Civil Rights, wrote the then-director of the U.S. Employment Service challenging the experimental GATB-based referral plan because of the way test scores are derived. Within-group scoring, in Mr. Reynolds' opinion, "not only classifies job applicants on the basis of their race or national origin, but . . . requires job service offices to prefer some and disadvantage other individuals based on their membership in racial or ethnic groups. Such a procedure constitutes intentional racial discrimination."

The Department of Labor adopted within-group scoring as a way to promote equal employment opportunity goals; the Department of Justice views the score conversions as illegally advancing the interests of one group at the expense of another. Their contradictory assessments of the use of race-conscious scoring to promote equal employment opportunity reflect the uncertain vision of the larger society.

In the years since the passage of the Civil Rights Act of 1964, the American people have not resolved a fundamental philosophical conflict between equality conceived of as color-blind law and equality defined as more equitable distribution of society's goods,

of making up for unequal opportunities. Many Americans who reject completely the system of laws and social conventions that relegated blacks to second-class citizenship for a hundred years after the abolition of slavery believe that the essential remedy is to ensure that the law is the same for all. Others feel that fairness requires more: that contemporary blacks carry the burden of disadvantage created by slavery and maintained by formal and informal segregation until the 1960s and deserve preferential treatment to enable them to join the mainstream of American economic and social life. The decline in the country's economic fortunes in the last decade has added a new dimension to the debate, as concern about productivity and American competitiveness affects perceptions of fair employment practices.

THE COMMITTEE AND ITS WORK

Charge to the Committee

As a consequence of the Justice Department challenge, the Labor Department sought guidance from the National Academy of Sciences, which, through the National Research Council, has convened a committee of experts to conduct a thorough, scientific evaluation of the proposed GATB-based referral system, including within-group scoring. In formulating its charge to the Committee on the General Aptitude Test Battery, the Department of Labor acknowledged that the important technical issues are rooted in a complex history of governmental policies and legal requirements and have serious economic and social consequences; these contextual realities must provide the framework for any specification of the policy options available to the U.S. Employment Service.

The evaluation of the GATB-based referral plan is to be carried out in two stages. The first stage is a preliminary study of the within-group percentile scoring system; the second stage is a full-scale evaluation of the whole referral plan and the likely consequences of its widespread adoption.

Stage One: Scoring Methods and Referral Rules

A principal component of the charge to the committee is the conduct of a careful review of alternative referral methods within the context of the issues raised by the within-group percentile

scoring system. This interim report represents the committee's preliminary findings, offered now in response to the Labor Department's request for rapid advice on the subject. We have analyzed issues theoretically in order to clarify the available policy options. But our mature conclusions about scoring and referral policy for the proposed GATB-based referral system must of necessity await the completion of the full study.

Stage Two: Full-Scale Evaluation of the Proposed Referral System

The second and major component of the committee's charge is "to conduct an impartial and comprehensive review" of the GATB-based referral program. We will study the psychometric quality of the GATB as an aptitude test, including the quality of its nine subtests. It is impossible to make recommendations about the proposed referral system without first establishing that the GATB can support useful inferences about performance in the many kinds of jobs handled by the Employment Service. Our evaluation of the GATB will include examination of the evidence supporting the claims that the GATB subtests measure unique aptitudes that are important contributors to the prediction of success in many jobs. We will examine the degree to which the predictive validities of the GATB are limited by the reliabilities of its subtests. We will consider the appropriateness of test item difficulties, the norms used to scale GATB scores, and other aspects of the psychometric quality of the GATB. In addition, we will evaluate the adequacy of the performance criteria, which are primarily supervisors' ratings, used to validate the GATB.

The principal scientific basis of the employment referral system proposed by the U.S. Employment Service is the work of Frank L. Schmidt and John E. Hunter, who in the 1970s brought statistical meta-analysis to the field of tests and measurement (see, in particular, Schmidt and Hunter, 1977, 1981; Hunter, 1982). Their theory, called validity generalization, has led some industrial psychologists and psychometricians to reconsider the traditional assumption that a new predictive validity study is necessary whenever a test is introduced in an employee selection system or when the nature of the job or character of the applicant pool changes substantially from that in existing validation evidence.

Since validity generalization is at the heart of the GATB-based referral system, we will consider both the theory of validity

generalization and its application in the USES referral program. This requires looking at the accumulated body of GATB validity research to see if predictive validities for virtually all jobs in the current *Dictionary of Occupational Titles* can be generalized from USES validity studies of particular jobs.

The second stage of the committee's task also includes analysis of the potential consequences of nationwide use of the GATB by the public Employment Service. If the USES-proposed system of employment referral is widely adopted, it could affect the economy of the United States, the composition of the workforce, the economic well-being of various groups of job seekers (including racial and ethnic groups, veterans, people with handicaps, and lower scoring majority group members), and the operation of the Employment Service. Furthermore, the USES system would set an example for public and private employers throughout the United States. The research agenda of the committee includes examination of each of these issues to the extent possible.

The Committee on the General Aptitude Test Battery

The National Research Council, the principal operating agency of the National Academy of Sciences, provides independent advice to the government on science and technology policy. Its work is accomplished by committees of volunteers, supported by a small professional staff. Members of National Research Council committees are chosen for their expertise, not to represent interested constituencies.

The members of the Committee on the General Aptitude Test Battery include experts in statistics and meta-analysis, psychometrics, industrial psychology, experimental psychology, economics, sociology, policy analysis, law, and the Employment Service—the expertise needed to address the broad range of technical and policy questions presented by the use of the GATB for referring job seekers to employers. Brief biographies of committee members and staff are in Appendix B.

The Liaison Group

The policy decisions ultimately made by the Department of Labor with regard to the GATB and its use in the public Employment Service will have an impact, perhaps a very great impact,

on the interests of various individuals, groups, and institutions. The Department of Labor and the National Academy of Sciences agreed that it is of vital importance that the committee be sensitive to the real-world implications of its findings. To ensure that the committee can call upon all relevant policy perspectives, experience, research data, and operating information in the course of its study, a liaison group has been appointed with the advice and cooperation of the Department of Labor. Its 27 members include representatives of the Departments of Justice and Labor, the Equal Employment Opportunity Commission, and state Employment Security agencies; spokesmen for business and labor, for minority group interests, for veterans, and for people with handicaps; and a number of scientists, including the developers of the theory of validity generalization. A list of the members of the liaison group is in Appendix C.

Each liaison group member has agreed to provide the committee with information, position papers, or data and, from time to time, to respond to requests for special briefings or policy statements on specific issues before the committee.

THE EMPLOYMENT SERVICE AND THE GENERAL APTITUDE TEST BATTERY

Congress established the public Employment Service during the Depression to help employers find workers and job seekers find work (Wagner-Peyser Act of 1933). It is a cooperative federal-state program that has grown over the years to include an extensive network of 2,000 local employment offices administered by the states. In 1985, the last year for which national figures are available, more than 20 million people were registered with a public employment office and, of these, 9.5 million received some "reportable" (by Employment Service staff) help in seeking a job.

Thus, at the state level, the Employment Service serves as an employment agency, or what an earlier generation called a labor exchange. Although there is great variety in procedures from state to state, and indeed from local office to local office, the basic functions of the state Employment Service agencies are to take job orders from employers, to take applications from job seekers, and to make referrals of applicants to job orders.

The Employment Service views itself in this role as an honest

broker, providing employers with access to a larger pool of potential employees than might otherwise be available to them and giving job seekers access to information about many job openings at a single location (unpublished USES communication to the committee, June 11, 1987). Employers send or phone in job orders to a local office of their state Employment Service, specifying the type of job(s) they need to fill; any special requirements for the job, such as educational credentials, work experience, or test results; and the number of applicants they would like the Employment Service to refer for each position. Each job order is assigned a code drawn from the Department of Labor's *Dictionary of Occupational Titles* (DOT), which classifies jobs according to a scheme of broadly defined performance requirements.

In day-to-day operations, it appears that the pool available to fill a particular job order is usually determined by the people who come into the office while the job order is current. But when an employer needs a large number of workers—for example, to put on an additional shift or staff a new facility—one or perhaps several local offices in a region will compile a large referral pool through advertising and file searches.

People in search of work who register at a local state Employment Service office are generally interviewed by a counselor who records information about their education, job experience, and preferences. In addition to the interview, a small number of applicants are given aptitude tests. On the basis of this information, one or more DOT codes is given to each registrant to reflect his or her job experience. These codes are the major means of matching people to jobs, although the Employment Service interviewer will also decide who to refer on the basis of an employer's special requirements, and the interviewer may make additional judgments about the suitability of the individual registrant for the job.

The federal part of the public Employment Service, the U.S. Employment Service, is a division of the Department of Labor; it provides research and technical support to state Employment Service agencies as well as some program monitoring and fiscal oversight. USES has carried out a variety of research programs over the years in support of the labor exchange functions of public employment offices. In 1939, USES produced the first of four editions of the *Dictionary of Occupational Titles*, the basic tool for matching workers and jobs. (The DOT was the subject of an earlier National Research Council report; Miller et al., 1980.) In

1947, it published the General Aptitude Test Battery (GATB) and made the test available to the states for use in vocational counseling and employment screening.

As its name indicates, the GATB is a general aptitude test, designed to be used for employment screening in many occupations. It consists of 12 separately timed subtests that were selected (on the basis of factor analysis results) from 59 occupation-specific tests that had been used prior to 1945. Eight of the subtests are pencil-and-paper tests, and four require manipulation of objects. The 12 subtests are intended to measure nine aptitudes:

Aptitude	*Subtests*
G-General Learning Ability	Three Dimensional Space Arithmetic Reasoning Vocabulary
V-Verbal Ability	Vocabulary
N-Numerical Ability	Arithmetic Computation Arithmetic Reasoning
S-Spatial Ability	Three Dimensional Space
P-Form Perception	Tool Matching Form Matching
Q-Clerical Perception	Name Comparison
K-Motor Coordination	Mark Making
F-Finger Dexterity	Rivet and Washer Assembly Rivet and Washer Disassembly
M-Manual Dexterity	Pegboard-place Pegboard-turn

The four subtests that require manipulation of objects and two of the eight paper-and-pencil subtests (name comparison and mark making) are designed to assess speed of work as a major component of the abilities they measure.

Between 1947 and 1980, some 700 validity studies were conducted to explore the relationships between performance on the GATB and performance on particular jobs. More than 400 Specific Aptitude Test Batteries (SATBs, combinations of GATB subtests related to 2, 3, or 4 aptitudes) were developed to predict performance in more than 500 occupations. Each validity study, in addition to specifying the most relevant aptitude composite for the job in question, suggested minimum passing scores derived from

the correlations between the aptitude composite and performance on the job (as measured by supervisors' ratings and other outcome variables).

After the passage of the Civil Rights Act of 1964 and subsequent implementing guidelines, increased emphasis was placed on studies of test fairness and on investigations of the validity of SATBs for members of minority groups.

In spite of this large research base, the GATB has not been extensively used in the Employment Service. Of the 9.5 million USES registrants who received some reportable service during fiscal 1985, less than 8 percent were given some or all of the GATB: about 387,000 registrants were given the GATB and perhaps as many as 316,000 others took one of the SATBs (Kelly, 1987). By comparison, approximately 1.8 million college applicants take one of two popular entrance examinations annually, and 1.3 million high school students take the Armed Services Vocational Aptitude Battery each year.

This picture could change dramatically, however. Pilot studies of the new referral system in 39 states resulted in 665,313 administrations of the GATB (in addition to regular administrations) in the program year from July 1, 1986, to June 30, 1987. If the GATB-based referral system is adopted throughout the Employment Service to screen job seekers for virtually all 12,000 jobs listed in the *Dictionary of Occupational Titles*, there would be a dramatic expansion of testing—so dramatic in fact as to deserve very careful consideration from both a scientific and a policy standpoint.

Part I
Balancing Efficiency and Equity: Issues in Science and Social Policy

SOCIAL GOALS, LEGAL CONSIDERATIONS, AND SCIENTIFIC ISSUES

The Committee on the General Aptitude Test Battery has been asked to conduct an impartial, comprehensive evaluation of a referral plan that the U.S. Employment Service hopes to promote throughout the Employment Service system. Central to the proposed plan is the use of the General Aptitude Test Battery (GATB), a test of cognitive, perceptual, and psychomotor abilities, to predict an applicant's expected performance in virtually all jobs.

This interim report focuses on one particular aspect of the proposed GATB-based referral plan: the conversion of raw scores on the GATB to percentile scores computed by groups based on applicants' racial or ethnic identities. The Department of Labor has asked for immediate consideration of the scoring issue because of the charge by the Department of Justice that within-group scoring has the effect of discriminating against job seekers in the majority group.

It is important to emphasize at the outset that this is a *preliminary* report. The bulk of the analysis deals theoretically with scoring methods and referral rules. We have found this analysis a useful and illuminating exercise, one that we believe can advance policy discussion. This report also makes some limited and highly qualified statements about how actual operation of a variety of referral rules in the state Employment Service agencies might be

expected to differ from the effects derived from the theoretical models. Our final conclusions and recommendations on the scoring issue will be presented when the full evaluation of the GATB and the referral plan is complete.

Neither the full-scale study nor this preliminary discussion of alternative referral rules will be of any great assistance to policy makers unless invested with an appreciation of the interests at stake and the legal and policy context in which employment selection now occurs. Questions about scoring methods and referral rules are not simply technical issues; indeed, at some level they are not primarily technical issues, but expressions of fundamental social policy.

Public debate over employment testing did not arise with the questioning of within-group scoring. Since the passage of the Civil Rights Act of 1964, testing has been the focus of controversy because of its explicit gatekeeping role. Some people consider employment tests artificial barriers, a device for maintaining racial and ethnic discrimination under the cloak of science. To other people, tests appear to promote both fairness and productivity by basing selection on ability. For everyone, the question of testing policy involves real economic interests and deep-seated political and social values. The following pages sketch the important and frequently conflicting social interests, frames of reference, and professional vocabularies that influence testing and employee selection. Our discussion centers on the economic, legal, and psychometric perspectives on testing that have been so influential in the evolution of policy and public opinion. Against this background, a representative array of possible scoring methods and referral rules and their policy implications are described in Part II.

THE ECONOMIC PERSPECTIVE: EFFICIENCY

One of the important insights of early social science was that the character and texture of life in a society is determined to an important extent by its economic arrangements. This insight has become part of the frame of reference of modern societies. Countries view themselves as capitalist or socialist or communist and in those concepts find substantial definition of their national character. As an example, the nineteenth century formulation of capitalism has had—and retains—an important hold on the American mind.

The fundamental premise of classical economic theory was that the operation of free and competitive markets would make a productive economy or, to put it another way, the pursuit of private gain would promote the public good. The assumptions of the free-market system were well suited to the new American context. In its respect for individual freedom and initiative, capitalism complemented democracy. The U.S. legal system was congenial to a capitalistic value system, being in large part devoted to protecting the rights of property. Above all, the emphasis of capitalist thought on traits like enterprise, mobility, and competitiveness was attractive in a nation undergoing great geographical and economic expansion.

In the contemporary capitalistic economy of the United States, economic policy continues to be guided by a presumption that, as a general rule, the free operation of the market will produce the most efficient, and therefore, the most prosperous, economy. The concomitant social virtues of competitiveness and individual enterprise continue to be highly valued. But nothing in the free-market system guarantees that the distribution of economic well-being will be socially desirable. Extreme concentrations of wealth and widespread poverty can easily coexist in capitalism, as the example of the United States makes clear.

In the course of this century, and particularly since 1960, many Americans have become convinced of the need for positive government intervention to produce a better distribution of economic well-being, both to ameliorate existing poverty and to ensure that economic opportunity is distributed equitably. Particularly as the civil rights movement developed, many Americans came to the conclusion that the abolition of slavery a century earlier had not brought economic, legal, or social justice to black Americans and that action was needed on all three fronts. Heightened by association with the civil rights movement, this broader concept of social justice created support for economic initiatives like the Great Society programs of the Johnson Administration and, with regard to employment opportunity, made the federal government an important force in the personnel offices of both public and private employers.

When a society moves from altruistic intentions to implementation, even modest redistribution of society's goods can be accompanied by resentment and controversy. There have been tensions between traditional free-market assumptions and the policies

aimed at a more equitable distribution of wealth and opportunity. Interventions to promote distributional goals are perceived by many to occur at the expense of economic efficiency. And these tensions have increased as America's competitive position in the world economy has faltered, despite recognition that the redistributive efforts are intended to mitigate market distortions caused by social policies of eras past.

The technique of within-group scoring was developed within the framework of an economic trade-off model that balances economic efficiency (in terms of quality of workers' job performance) against equity (in terms of increasing the participation of minority group members in the workforce). This conceptualization of the issue as a trade-off between economic efficiency and equity is consistent with the way employment testing has been treated as a public issue in the last 20 years or so. The Civil Rights Act of 1964 itself recognizes an employer's interest in hiring the most qualified staff possible at the same time as it requires the elimination of discriminatory employment practices (see *Griggs* v. *Duke Power*, 401 U.S. 424, 436).

The committee agrees that the trade-off model provides a useful way of thinking about many of the policy questions before us, and we have adopted the model in order to explore the effects of various referral rules in Part II of this report. However, the model, like every model, has limitations. As our discussion of the legal context surrounding the issue of within-group scoring makes clear (below), there are deep differences of opinion on the meaning of equity. There are also complexities in assessing economic efficiency that are not addressed by the model. In the conceptualization described here and used in this report, economic efficiency or productivity is considered at the level of the individual employer. This seems an appropriate first line of approach to the question, particularly since the subject of analysis involves employment testing. Nevertheless, time has shown the limitations of the classical economic approach of equating private gain and public good.

There could be social costs attached to the very use of ability tests that are masked by the approach taken here. For example, selection based on test scores implies that those persons with the lowest scores will rarely be referred for jobs. Although the number of people who are unemployed might not change, widespread

test-based selection would tend to place the same individuals perpetually at the bottom of the totem pole, creating a class of citizens with little chance of employment. An economic rationale for supporting policies that increase the employment of minorities and others outside of the economic mainstream could be made on the grounds of better utilization of human resources or encouraging investment in human capital even in the face of some decrement of efficiency at the level of individual employers.

One of the important contributions of economics as a discipline has been its approach to policy problems through trade-off models. Any important social issue requires policy makers to balance competing interests, and the trade-off model helps to locate the most productive point of compromise. But the notion that there might be a trade-off between efficiency and equity is not in harmony with the law regarding discrimination in employment: the aim of Congress and the courts has been to remedy a perceived wrong. Rather than encouraging a search for efficient trade-offs, as might be done if the economic perspective held sway, the courts have adopted an analytical framework that pits an employer's right to make a profit against the rights of minority group members to more equitable treatment. The apparent advantage may lie with employers, who have the right to adopt exclusionary practices if they can show that these practices serve their business interests. This is consonant with the fundamental premise of capitalism that the individual pursuit of economic self-interest will promote the larger social good. But, as documented in an earlier National Research Council report (Wigdor and Garner, 1982), courts and regulatory agencies can set standards that make it virtually impossible for employers to demonstrate the business necessity of their procedures. One is left with a situation where right conflicts with right, and compliance with one rule can lead to violation of another.

THE LEGAL PERSPECTIVE: EQUITY

Within-group scoring illustrates this conflict. The U.S. Employment Service developed a race-conscious method of scoring for its referral plan in order to meet what it understood to be the demands for group equity under the civil rights laws. Within-group scoring was developed in the spirit of a referral policy negotiated in 1972, between USES and the Department of Labor legal staff,

the Equal Employment Opportunity Commission (EEOC), and the Department of Justice, based on the newly emerging doctrine of adverse impact. The 1972 "Referral Ratio Policy" stipulated that for serving employer orders with tested applicants when the test battery had not been validated on minority groups: the ratio of minority applicants referred to employers would not be less than the ratio of minorities in the applicant pool; and test results would not be reported to employers (Quinlan, 1973). Now, however, Assistant Attorney General for Civil Rights Wm. Bradford Reynolds has challenged the scoring method because he believes it illegally infringes the rights of individuals whose opportunities may be diminished as a consequence of the benefits accorded minority group members.

Civil Rights and Fair Employment Practices Law

The legal context in which the Employment Service operates is provided by Title VII of the Civil Rights Act of 1964 and the 5th and 14th Amendments to the U.S. Constitution, supplemented in the regulatory arena by the 1978 *Uniform Guidelines on Employee Selection Procedures.* Title VII of the Civil Rights Act, entitled Equal Employment Opportunity, is a remedial statute intended to achieve equity in the workplace by eliminating discrimination in hiring, promotion, remuneration, or other conditions of labor. Employers and public and private employment agencies are prohibited from using race, sex, color, national origin, or religion as the basis of employment decisions. The act empowered the federal courts to remedy challenged discriminatory employment practices by ordering "such affirmative action as may be appropriate," including hiring or reinstating employees, with or without compensation.

While requiring equality of treatment, the language of the Civil Rights Act of 1964 did not go beyond the principle of color-blind practices. During congressional debate, supporters of the act such as Hubert Humphrey, who was floor manager for the bill in the Senate, repeatedly denied that the term discrimination would be read as mandating racial quotas. And indeed, Section 703(j) of Title VII states specifically that nothing in the statute shall be interpreted to require any employer to grant preferential treatment to any individual or group because of race, color, sex, etc. on account of any imbalance in the employer's workforce.

However, many people believed that the law against racial discrimination, like the earlier labor laws enacted to protect women and children, must recognize and protect members of groups at risk. As the previously cited National Research Council report (Wigdor and Garner, 1982:96-115) pointed out, there was an almost immediate and ultimately dramatic shift in government policy—as expressed both in administrative and judicial construction of Title VII and in the legislative history of the 1972 amendment of the act (Robertson, 1976:17-30)—from the requirement of equal treatment to a focus on the outcome of employment practices. When Congress extended the act to include nondiscrimination in federal government employment in 1972, it required each federal department and agency to develop "an affirmative program of equal employment opportunity," and reaffirmed their responsibilities under Executive Order 11478, which requires them "to promote the full realization of equal employment opportunity through a continuing affirmative action program."

The EEOC was instrumental in focusing the new law on employment practices having the "effect" of discriminating. Created by the Civil Rights Act of 1964, the EEOC was authorized to provide leadership and guidance on the meaning of Title VII, to promote compliance, and to prevent any person from engaging in unlawful employment practices. In a series of guidelines on the use of tests and other selection procedures issued in 1966, revised in 1970, and revised again in 1978 when the four federal agencies with authority in the matter (EEOC, the Departments of Justice and Labor, and the Office of Personnel Management) adopted the *Uniform Guidelines on Employee Selection Procedures*, the EEOC elaborated a definition of discrimination that turns on whether an employment practice "adversely affects," i.e., tends to exclude, members of a protected class. As stated in the 1978 *Guidelines* (41 CFR Ch. 60-3.3):

> The use of any selection procedure which has an adverse impact on the hiring, promotion, or other employment . . . opportunities of members of any race, sex, or ethnic group will be considered to be discriminatory . . . unless the procedure has been validated in accordance with these guidelines.

Accordingly, compliance authorities focused on the outcome of selection procedures as manifested in the composition of the workforce for evidence of such adverse affects.

The Department of Justice was also important in shifting the meaning of Title VII away from intentional discrimination to focus on the effects of employment practices. The Attorney General is empowered by the act to bring civil suit when there is reasonable cause to believe that an employer is engaged in "a pattern or practice of resistance" to the full enjoyment of the rights secured by the title. These "pattern or practice" suits quickly made workforce statistics rather than intent or motivation the primary medium of courtroom argument.

In *Griggs* v. *Duke Power Co.*, decided in 1971, the Supreme Court confirmed this interpretation of Title VII. The opinion focuses judicial attention on the consequences of a selection procedure rather than on intent or motive, saying that the Civil Rights Act proscribes "not only overt discrimination but also practices that are fair in form, but discriminatory in operation." Congress did not, in the opinion of the Court, intend to guarantee everyone a job, regardless of qualifications. What is required is the removal of "arbitrary, artificial, and unnecessary barriers to employment when they operate invidiously to discriminate on the basis of racial or other impermissible classifications." The touchstone, in the words of the Court, is "business necessity": a practice that operates to exclude blacks is prohibited *unless* it can be shown to be related to job performance (401 U.S. 424, 431). Since 1971, imbalance in the workforce has been accepted as prima facie evidence that discrimination had occurred. In order to rebut the inference of discrimination, an employer would bear the burden of showing that any given requirement has "a manifest relationship to the employment in question" (401 U.S. 424, 432).

Thus, early implementation of Title VII of the Civil Rights Act of 1964 established two of the more potent concepts in EEO policy: adverse impact and job relatedness. The result of this emphasis on the *consequences* of employment procedures rather than motive or intent has been to strengthen the redistributive effects of Title VII. However, it is important to distinguish between what Title VII requires and what it permits. The reasoning of the Court in *Griggs* is that Title VII does not require that preferential treatment be accorded to minorities; rather, qualifications are to be the controlling factor. Race, color, sex, ethnic origin, and religion are irrelevant. Later judicial interpretation of the act, on the other hand, has recognized that equal treatment might not result in fair treatment for blacks given the degree of disadvantage produced by

long-term, severe, and officially sanctioned inequality. Title VII is now understood to permit, in circumscribed instances, affirmative action programs intended to offer preferential treatment so that those subject to disadvantage would be better able to compete in the marketplace. And when proscribed discrimination is traceable to a particular employer, the courts have often required the employer to adopt race-conscious remedial measures.

The overall tendency of judicial and administrative implementation of the Civil Rights Act since *Griggs* has been to encourage employers in the direction of preferential hiring, thus blurring the distinction between nondiscrimination and affirmative action. At the same time, it has been more difficult to reach public consensus on the proposition that equity requires preferential treatment for protected groups, at least in the short term, than it was to generate enthusiasm for the original conception of the Civil Rights Act that the law should be color-blind.

Like the society at large, the courts and other legal authorities have expressed concern for the burdens imposed on majority group members by the extension of scarce benefits to protected classes. In the emerging case law, the Supreme Court has recognized the use of race- and gender-conscious employment practices in rather closely circumscribed situations for the purpose of remedying past or present unlawful discrimination or to foster appropriate affirmative action; on occasion, particularly where layoffs are concerned, race-conscious measures have been struck down in the lower courts on grounds of reverse discrimination. The Supreme Court seems to have rejected redress of a broad claim of "societal discrimination" as an acceptable basis for adopting race-conscious selection procedures that impose burdens on others (*Wygant* v. *Jackson Board of Education*, 106 S.Ct. 1842 [1986]), but commentators disagree about the boundaries of acceptable and unacceptable race-conscious procedures. In this context we consider the divided legal opinions about within-group scoring of the GATB.

Perspectives on Within-Group Scoring

Although the Supreme Court has not had occasion to address the subject of within-group scoring, a variety of score adjustment mechanisms intended to reduce adverse impact have been upheld at the appellate level, particularly in the Second Circuit. Nevertheless, there are significant differences of opinion about the

procedure as it is being used in pilot studies of the GATB-based referral system.

As detailed above, the Assistant Attorney General for Civil Rights has taken issue with the scoring system promoted by the USES because it classifies job applicants on the basis of their race or national origin and because it requires Employment Service offices to prefer some individuals and disadvantage other individuals based on their membership in racial or ethnic groups (Reynolds, 1986). The Justice Department finds the practice unconstitutional under *Wygant* v. *Jackson Board of Education* (106 S.Ct. 1842 [1986]) and *Local 28, Sheet Metal Workers' International Association* v. *EEOC* (No. 84-1656 [July 3, 1986]). The Justice Department position is that these cases make clear that racial preferences are permissible "only as a last resort to remedy persistent and egregious discrimination by the specific employer" (Reynolds, 1986). The GATB referral program, however, requires government agencies to extend racial preferences regardless of whether an employer has engaged in any racial discrimination, and it does so outside of a specific remedial context. The Justice Department also finds the score conversions illegal under Title VII, citing Section 703(j), quoted above, which denies any requirement for preferential treatment.

Other commentators, including the Lawyers' Committee for Civil Rights Under Law, read the case law rather differently. They point out that both Title VII and the Constitution permit (and sometimes require) the use of race-conscious selection procedures in appropriate circumstances. For example, the Supreme Court has held that private (*United Steel Workers of America* v. *Weber*, 443 U.S. 193 [1979]) and public (*Johnson* v. *Transportation Agency, Santa Clara County, California* (No. 85-1129 [March 25, 1987]) employers may, independent of any judicial finding of past discrimination, adopt race-conscious hiring or promotion plans as part of a voluntary affirmative action program to address a "conspicuous . . . imbalance in traditionally segregated job categories." As Justice Stevens wrote in a concurring opinion in *Johnson* (p.3):

> . . . since 1978 the Court has unambiguously interpreted the statute to *permit* the voluntary adoption of special programs to benefit members of minority groups for whose protection the statute was enacted.

In the *Weber* case, in which a white employee challenged

race-conscious admission to a new program to train workers for high-paying skilled crafts jobs, the opinion of the Court emphasized the temporary duration of the plan, its remedial purpose, the "voluntary" nature of the plan, and the fact that it did not abrogate preexisting rights since the whole training program was new. In *Johnson*, a female was promoted in preference to a male who had received a slightly higher rating, although both were rated as well-qualified for the job of road dispatcher. At the time, none of the 238 incumbents in the job category was female. The Court, guided by its decision in *Weber*, affirmed that voluntary employer action can play a crucial role in furthering Title VII's goal of eliminating the effects of discrimination in the workplace. In rejecting the district court's finding that the affirmative action plan was illegal because it was not temporary, the Supreme Court reasoned that: (1) the plan was flexible and did not impose quotas; (2) it did not authorize blind hiring by the numbers, but expressly directed that numerous factors be taken into account; and (3) the employer's plan was to "attain" a balanced workforce, not to "maintain" a permanent racial and sexual balance.

The determination of whether or not a given race-conscious procedure is lawful turns on the facts surrounding its use. Advocates of within-group scoring argue that since the procedure was adopted in order to comply with the requirements of Title VII and that, without the scoring adjustment, the GATB would result in significant adverse impact against minorities, judicial precedent supports the legality of the procedure as a reasonable measure to eliminate that impact.

Representatives of the Lawyers' Committee for Civil Rights Under Law pointed out to the committee by that the courts have approved the use of a variety of score adjustment mechanisms intended to reduce adverse impact. For example, an appellate court approved adding 250 points to the score of each minority candidate on the basis of evidence that the scores of minority candidates on the written portion of a promotional examination underpredicted their job performance (*Kirkland* v. *New York State Department of Correctional Services*, 628 F.2d 796 [2d Cir. 1980] cert. denied 450 U.S. 980 [1981]). In another case, the same court upheld a consent decree that called for a variety of race-conscious scoring procedures simply on the basis of a showing that the existing scoring and rank-ordered selection procedure had an adverse racial impact (*Kirkland* v. *New York State Department of*

Correctional Services, 711 F.2d 117 [2d Cir. 1983], cert. denied 465 U.S. 1005 [1984]). The race-conscious scoring procedures that the court approved included: (1) separate frequency distributions for minority and nonminority candidates; (2) establishing score zones in which a group of final examination scores are deemed the same for purpose of certification and appointment; and (3) elimination of particular items that resulted in statistically significant adverse impact among candidates of substantially equivalent ability.

In these and a number of other cases, courts have upheld methods of score adjustment that, if followed, would reduce or eliminate the adverse racial impact of the selection practice and avoid continued violation of Title VII. These cases may or may not apply to a race-conscious scoring system voluntarily adopted by the Employment Service outside of a remedial context. In addition, the emerging case law does not seem entirely consistent. In *San Francisco Police Officers' Association* v. *San Francisco* (812 F.2d 1125 [9th Cir. 1987]), the court of appeals rejected reweighting of three selection tests to eliminate an adverse impact against women on the grounds that the reweighting "unnecessarily trammeled the interests of nonminority police officers." The adjustment of scores, in the court's opinion, became the sifting device, rather than the examinations themselves (footnote 5). And in *Hammon* v. *Barry* (813 F.2d 412, petition for rehearing denied, 826 F.2d 73 [D.C. Cir. 1987]), the court of appeals rejected selection of fire fighters from rank-ordered lists compiled separately by race, sex, and ethnic group in proportion to their representation among those who passed the test. In this case the court failed to find the necessary "predicate of discrimination."

In addition to legal precedents, proponents of the adjustment of GATB scores point out that Section 6(A) of the *Uniform Guidelines* encourages the use of alternative selection procedures, including race-conscious procedures, as a way of achieving compliance with Title VII or achieving affirmative action purposes. (There is a caveat that the procedures must be legal.) The signatories to the *Uniform Guidelines* joined in adopting a set of "Questions and Answers to Clarify and Provide a Common Interpretation of the Uniform Guidelines" (*Federal Register* 43:12001) in March of 1979. The explication provided in questions 30 and 31 strongly suggest that no validation is required of alternative procedures adopted to eliminate adverse impact, because federal

law does not require a demonstration of the job relatedness of selection procedures that do not have adverse impact. In fact, under the *Uniform Guidelines*, use of alternative selection procedures to eliminate adverse impact is an option that is available to employers in lieu of validation. It would seem, then, that a within-group scoring procedure that eliminates adverse impact could fall within the *Uniform Guidelines*.

This committee is obviously not in a position to make a definitive statement about these conflicting interpretations of the legality of within-group scoring of the GATB. The evolution of fair employment law since 1964 has produced two grounds for race-conscious employment practices: the mitigation of adverse impact and voluntary affirmative action. It would appear that the Employment Service may not be able to justify use of score adjustments as part of its lawful affirmative action efforts because it is not acting as an employer. And, since in Justice Stevens's words, Title VII permits but does not require an employer to grant preferential treatment on the basis of race or gender, score adjustments for affirmative action purposes by a governmental employment agency might be found to constitute undue governmental interference with managerial discretion. If the scoring system is not justifiable as part of an affirmative action plan, then its acceptability would seem to depend on whether the weight of legal opinion will recognize the adoption of a generalized score adjustment, designed to prevent adverse impact, as an appropriate compliance effort under Title VII.

THE PSYCHOMETRIC PERSPECTIVE: VALIDITY

One of the ironies of the evolution of legal and regulatory policy is that in all probability, for within-group scoring to be adjudged "remedial," it would be necessary that the procedure be used primarily with tests that have not been validated in a manner that complies with the *Uniform Guidelines*.

From the point of view of psychometrics, validity is the paramount issue in the evaluation of the use of a test. This is the central premise of the *Standards for Educational and Psychological Testing* (American Educational Research Association et al., 1985:9) and the *Principles for the Validation and Use of*

Personnel Selection Procedures (Society for Industrial and Organizational Psychology, 1987:4), the published standards of measurement professionals. Without evidence of validity, there is no scientific justification for using a test.

Validity is a very broad concept. It refers to "the appropriateness, meaningfulness, and usefulness of specific inferences made from test scores" (American Educational Research Association et al., 1985:9). Validity is always a matter of degree, rather than a simple "valid or invalid" dichotomy. For example, research may reveal that there is a relationship between scores on a test and subsequent performance on a job, but the relationship may be weak or strong or at any level in between. (There will never be a perfect correspondence and a total lack of correspondence is unlikely.) The strength of the relationship demonstrated supports the interpretation that the test has a low, moderate, or high degree of validity as a predictor of performance on the job.

In validating employment selection procedures, one looks for positive relationships between predictors, such as test scores, and performance criteria, such as supervisor ratings or work sample tests. The type of validity referred to here, in the context of the GATB and its use by the USES, expresses statistically the degree of relationship between some (or all) of the GATB subtest scores and scores on some measure of job performance. For most of the GATB validity research, the job performance measure has been a supervisor rating of overall job performance as good, fair, or poor. The statistic used is the correlation coefficient, which can vary from -1.0 through 0 to 1.0. At zero correlation, using the test is no better than selecting at random. The further the value of the correlation coefficient is from 0, the stronger the linear relationship between the two measures being correlated. For correlations in the positive range (above 0), there is a direct relationship between test score and whatever measure of job performance is used. For correlations in the negative range the relationship is an inverse one.

Correlation coefficients in industrial psychology are typically lower than in educational psychology. The very complexity of human performance makes measurement difficult. Moreover, the relationship between the cognitive abilities traditionally assessed (verbal and numerical skills) and actual performance is less clear in the world of work than in academic settings. A low degree of validity, say .20, will have some, but probably limited, usefulness for

predicting job performance. A moderate degree of validity, say .40, is considered to be quite satisfactory by most industrial psychologists. What is considered a high degree of validity in this field, say .60, may occasionally be observed with specially developed criterion measures collected under carefully arranged circumstances. It is important to note, however, that even a high degree of validity, as represented by a correlation coefficient of .60, will still not provide complete accuracy in predicting job performance from test scores. Conversely, even a test with low or moderate degrees of validity provides better predictions of performance than random selection or the use of invalid procedures.

Although validity is a matter of degree, the use of a test for a given purpose, such as creating a list of job candidates to refer for a particular set of job vacancies, is usually dichotomous. That is, a person is either referred or not referred. A user must decide if the evidence about validity and the degree of validity obtained are sufficient to support that decision. In light of the potential economic gains and legal vulnerabilities described above, a user needs to weigh the consequences of using or not using the test.

Bias

An important psychometric consideration in the use of tests is the degree to which a test, or more accurately, the use of a test, might be biased against members of a particular group. In popular parlance, bias is often equated with adverse impact: that is, the simple fact of differential group performance is taken as evidence of bias in the test. That is not what the committee means by bias. From a psychometric perspective, bias may be thought of as a form of invalidity; it is the result of systematically inaccurate measurement involving either the predictor, the criterion measure, or both. Bias results when a test includes sources of difficulty that are irrelevant to the characteristic that the test is designed to measure and that artificially reduce the scores of an identifiable group of examinees; this bias leads to false inferences about the knowledge, skill, or ability of members of the group. For example, if black test takers or test takers aged 40 or more are less familiar with the new math and if a test of arithmetic skills is designed in a new math format, then inferences about the level of arithmetic skills of the black or middle-aged test takers could well be biased. Or, to extend the example of older test takers, if a test of

arithmetic skills had a very short time limit (this is called a highly speeded test), then older examinees might be at a disadvantage as compared with younger ones; the test scores would underrate the skills of the older examinees by confusing speed and ability.

If a test is used for selection and its use is supported by a demonstration that test scores are related to performance on the job, it is relevant to ask a further research question: are the predictions of job performance made from the test scores biased against members of a particular group. The *Uniform Guidelines*, addressing this issue under the rubric of fairness, caution:

> When members of one race, sex, or ethnic group characteristically obtain lower scores on a selection procedure than members of another group, and the differences in scores are not reflected in differences in a measure of job performance, use of the selection procedure may unfairly deny opportunities to members of the group that obtains the lower scores" (Sec. 1607.14 A.(6)(a)) [emphasis added].

The technical approach to addressing this issue involves the comparison of prediction equations that are estimated separately for the two groups in question. If test scores have the same predictive meaning regardless of the group to which a person belongs, then the predictions are considered unbiased and the use of the test equally valid for both groups.

The question of test bias is directly relevant to the issue of scoring methods and within-group score adjustments. If the GATB is found to predict job performance differently for different groups, an adjustment that corrected for these differences might be justified from a scientific standpoint. Presumably, such an adjustment would also be legally warranted. The committee's research agenda includes examination of the question of group differences in performance on the GATB.

Specificity Versus Generalizability

Until recently, personnel psychologists generally believed that the validity of a test for predicting job performance is highly dependent upon the specific situation. For example, demonstrating that a test was a good predictor of, say, clerical performance at a bank in Chicago would not imply that the same test would be a good predictor of clerical performance at a bank in New York,

much less that it would be a good predictor for clerks at a department store, for salespeople, or for medical technicians.

This assumption of specificity has been increasingly challenged during the past 10 years as a result of quantitative analyses of the results of validity studies. These studies, known as validity generalization studies, have been purported to show that, contrary to past belief, validity is highly generalizable from one situation to another. The influence of validity generalization research results is visible in the Society for Industrial and Organizational Psychology *Principles* cited above (1987:26), which conclude that "validities generalize far more than once supposed." The degree to which GATB validities can be generalized, the conditions that affect the degree of generalization, and the conditions under which it may be reasonable to infer that the GATB is valid for a particular job based only on the results of validity generalization research, will be a major focus of the work of the committee's continuing work.

Accuracy in Reporting

A final aspect of validity that is important in psychometrics, as it is to scientific enterprise in general, is that the relationships observed be reported accurately. It is a matter of what might be called numerical ethics that test scores should mean what they appear to mean. In the case of employment tests, two individuals with the same score should have the same predicted job performance. And the accuracy of the prediction should be reported to test users. According to professional standards, both test results and test limitations should be accurately reported; test makers even have the responsibility for dispelling common misinterpretations (American Educational Research Association et al., 1985:36). From this perspective, within-group percentile scores are problematical unless such conversions more accurately reflect the predicted job performance of members of each group than would unconverted scores.

Part II
Policy Alternatives: A Theoretical Exploration

BACKGROUND

In considering tests used in employment decisions, the conflicts between equity and efficiency goals and between individual and group conceptions of fairness take on a different status depending on one's vantage point. From a psychometric perspective, the validity of a test and the accuracy of test score reporting are paramount. From an economic perspective, employment procedures should contribute to efficiency or productivity (although this can be defined as more than the aggregate of individual employers' productivity gains). From the legal perspective, fair employment laws have no intrinsic interest in either test validity or optimal productivity: they focus on eliminating adverse impact in ways that do not unduly trammel the interests of majority-group job seekers. The challenge for the U.S. Employment Service is to come up with a referral plan that incorporates all three vantage points.

This challenge, of course, is not unique to USES. Before looking at specific referral rules, we briefly note what has been done in analogous situations. Private employers have faced these tradeoffs in their own personnel screening efforts, and the same issues have also been debated in education, especially with regard to selection into post-secondary and graduate courses of study.

Examples from Other Settings

The USES within-group scoring device is not the first such

attempt to deal with the exclusionary effects of ability tests by adjusting the scores of affected groups. For example, scoring of the Preliminary Scholastic Aptitude Test (PSAT) for the National Merit Scholarship competition is now based on a double weighting of verbal scores over quantitative scores, a weighting that was selected in part to correct for the underrepresentation of females when a simple sum of the two scores was used. The National Merit Scholarship program also uses score conversions to ensure proportional representation from the 50 states: the score necessary to be selected as a finalist is not the same in Mississippi as in Connecticut. Granting National Merit Scholarship awards to certain candidates whose raw scores are lower than those of candidates rejected in another state has a surface similarity to within-group scoring of the GATB. It would seem to achieve group parity at the expense of individual fairness, although one could also argue that achievement of the student must be considered in the context of the state and local educational systems serving the student.

In higher education admissions decisions, the most widespread practice followed to mitigate the adverse impact of selection tests on minority candidates is the use of multiple selection criteria. This is appropriate because high school grade-point average, for example, is a strong predictor of college success, and it has less adverse impact than test scores. Personal characteristics such as race, being the child of alumni, state residence, and extracurricular activities also influence selection decisions, but in relatively subtle ways that do not negate the dominant "merit" ranking determined by high school grades and test scores.

Recent empirical work by Willingham and Breland (1982) for the College Board documents how institutional goals for a balanced student body enter the decision-making process. For example, in the nine selective institutions studied, minority status increased the probability of a candidate's being admitted about 30 percent over the chances based on grades and test scores alone. Thus, a minority candidate with middling scores does not have the same chance of being selected as a majority candidate with a high score, but within a range of grades and test scores, when both candidates may be thought of as approximately equal on academic criteria, preference is shown to the minority candidate.

Although the principle of using multiple selection criteria is worth exploring in the USES context, the situation is not strictly analogous because high school status and grades apparently do

not predict job performance as well as they predict success in academic settings. The military, for example, has for years used high school graduation status as an important screening device, but as a predictor of likelihood of retention, not job performance. A large body of research has shown weak correlations between high school graduation and job proficiency. However, there are other criteria, particularly job experience, that have promise as supplements to test scores in predicting job performance.

A final development worth noting comes from testing professionals. Public pressure on test developers in the last 15 years has led to the development of more sensitive statistical techniques for discovering bias in test items. Various internal procedures can be used to identify items that pose *irrelevant* difficulties for minority examinees. In other words, if there is something in the item format or content that impedes the performance of knowledgeable minority test takers, then the substitution of other items that accurately elicit the knowledge of the examinees would reduce the test's adverse impact. Since the GATB was developed long before such techniques were invented, a very careful analysis of its items for possible bias would be appropriate, although this sort of analysis would not fully answer the larger question of fairness.

Distinguishing Recruitment, Referral, and Selection

The public Employment Service acts as an employment agency. Although from the point of view of fair employment law it is part of the selection process, the Employment Service only refers candidates for employment, it does not make hiring decisions. This distinction could be important in finding a legally defensible referral policy. If one looks upon the entire hiring process as a continuum from recruitment through selection, then the middleman function of employment agencies can be viewed as falling on the recruitment side of the scale. It is a mechanism by which employers seek and receive candidates to consider for job openings. The purpose of active recruitment is to attract more applicants than can be hired, because selection from a large applicant pool is likely to produce a more able workforce.

Courts have recognized race-conscious recruitment activities that are intended to increase the number of qualified minority candidates as a legitimate means of avoiding adverse impact and of achieving affirmative action hiring goals. While it is inarguably

the case that the Employment Service referral system works as a screening device as well as a catchment pool, its positive role in creating a large pool of qualified applicants from which an employer will select may make race-conscious policies such as the over-referral of minority applicants defensible.

Distinguishing Scoring Methods and Referral Rules

In thinking about referral systems based on testing, a distinction needs to be made between the method chosen to score test performance and the strategy used to select applicants for referral, i.e., the referral rule. In this report, we have been asked to address the issue of within-group score conversions, which seems to imply that the debate about the test score method is paramount. But the method used to compute scores is only one dimension of a referral system, and it is separable from the decision rule used to select which applicants to send to employers. If within-group percentile scores are found unacceptable, on legal or scientific grounds, there may be other combinations of scoring method and referral rules that could achieve the basic purpose of producing the most able or most productive workforce while reducing adverse impact.

REFERRAL RULES

In the following pages we describe six common referral rules that are based only on test score and, if pertinent, group membership. Each rule is then evaluated theoretically for its effect on expected job performance—as estimated by the correlation between performance on the test and subsequent performance on the job—and on the proportion of minority group members referred, under a range of assumptions about the relationship between test score and job performance, about the proportion of applicants referred to the job, about the distributions of minority and nonminority scores on the test, and about the proportion of the minority group in the applicant pool. Finally, we discuss the application of these rules to Employment Service referral policies.

Six Basic Rules

Raw-Score, Top-Down Referral

Applicants are referred in order of their scores on the test, from high to low.

Within-Group Percentile Score, Top-Down Referral

A percentile score is computed for each applicant by comparing the raw score for that applicant with the scores obtained by a norm group of the same race or ethnic group. This rule is race conscious. A score of 60 means that the applicant scored at least as high as 60 percent of the applicant's norm group.

Referral is in order of percentile scores for all groups in the applicant population. Thus, applicants with a percentile score of 60 are referred before applicants with a percentile score of 50, even though some of those at the 50th percentile will have higher raw scores than those at the 60th percentile from a different race or ethnic group.

The effect of within-group percentile referral is to increase the proportion of minorities referred in comparison with a raw-score, top-down referral system.

Minimum Competency Referral

Applicants with raw test scores exceeding some minimum value, the cut-off score, are referred at random.

Zone Score, Random Within-Zone Referral

The test score range is divided into zones, from 2 to 10 (or more). Typically, each zone contains the same number of applicants. All raw scores in a given zone are converted to the same zone score.

Applicants are referred in order of their zone scores from high to low, but at random from the lowest zone that is referred. For example, if there are five zones with 20 applicants in each zone and an employer wants 70 referrals, all applicants in the first three zones will be referred and one-half of the applicants in the fourth zone, chosen at random. When there are two zones, this rule is

similar to the minimum competency rule. When there are many zones, the rule is nearly identical to raw-score, top-down referral.

Zone Score, Preferential Within-Zone Referral

This rule is the same as within-zone referral, except that applicants from some racial or ethnic groups are referred before those from other groups within the lowest zone that is referred. This rule is race conscious.

Expected Performance Ratio Referral

This complex referral rule incorporates the degree of relationship between test score and job performance as well as the difference in average scores between the majority and minority groups.

Referral in order of raw test score refers minority group members in lower numbers than they would be referred in order of job performance, if job performance for all applicants were known, because test scores are usually imperfectly related to performance. As a result, there will be greater differences in test scores than there are in estimated job performance. This referral rule enables one to refer the proportion of minority group members who would be referred if referral were in order of job performance rather than test performance.

As with the within-group percentile rule, referral according to this plan is top down according to a modified test score. The proportion referred from each group is equal to the predicted proportion if referral were in order of performance. For example, assume that the correlation expressing validity between test score and productivity is 0.4, and that a minority group average test score is 0.5 standard deviations less than the majority group average test score. This referral rule is approximately the same as adding the quantity $0.5(1 - 0.4) = 0.3$ standard deviations to each minority score and referring in order of the adjusted score. In contrast, the within-group percentile rule would adjust the score by adding 0.5 standard deviations to each minority score, a value unadjusted for the degree of relationship between the test and job performance. This rule is race conscious.

This rule is a compromise between raw score and within-group

percentile referral. Like within-group referral, it takes group membership into account. It corrects the disadvantage to a minority group caused by less than perfect prediction of job performance from test score. If the test score predicts performance poorly, then the rule behaves like within-group percentile referral. If the test score predicts performance well, then the rule behaves like raw-score referral. This rule has the definite drawback that it is necessary to be able to predict performance explicitly from the test scores; that is, the validity coefficient for each job must be estimated.

Theoretical Assumptions Used

The committee computed the effects of each referral rule on the expected performance of referred applicants, and on the proportion of minority applicants in the group referred, under the following four assumptions:

(1) The employment decision is based only on the GATB score for a particular job family, and, if pertinent, on minority or majority group membership. (For convenience, this analysis posits only two populations.)

(2) The test score and job performance have jointly normal distributions with unit standard deviations and with the same correlation for the majority group and the minority group.

(3) The same regression line formula predicts mean performance for a given test score in both the majority and the minority group.

(4) The mean test score for the minority group is one standard deviation below the mean test score for the majority group.

The norm group for within-group percentile referral was taken to be the corresponding group of applicants. For the zone rules, we chose five zones, and the zones for zone scoring were computed so that the same number of applicants were in each zone. To illustrate the minimum competency rule, we chose the 30th percentile as the minimum passing score.

The effects of the rules depend on the correlation between test score and performance, on the referral ratio, and on the proportion of minority group members in the applicant pool. Therefore, we examined the rules for several values of correlation, referral ratios, and minority proportions.

A correlation of .2 indicates a modest relation between test score and performance, a correlation of .4 a moderate relation, and a correlation of .6 a strong relationship for this type of data. Even at a correlation of .6, performance is not perfectly predicted from test score. USES Test Research Report No. 43 (U.S. Department of Labor, 1983b) suggests a correlation between test score and job performance of 0.5, but it is prudent to allow for a wide variation in correlations for different jobs and different measures of job performance.

The referral ratio is the proportion of applicants referred to the job. Attractive jobs will have many applicants and low referral ratios; perhaps only 1 applicant in 10 will be referred. Less attractive jobs, or ones requiring special skills, will have fewer applicants; perhaps 1 in 2 of the applicants will be referred.

Tables 1, 2, and 3 report the results of the computations for selected values of correlation, referral ratio and minority proportion in the applicant pool (see Appendix A). They are examined below in the section "Discussion of Referral Rules." Before that examination, we consider the data and operations that are likely to qualify the theoretical findings.

Referral Rules in an Operational Context

Data Used in Evaluating Referral Rules

In considering the applicability of these theoretical findings to Employment Service referrals, we used data from Employment Service registrants in 1985 and other USES research data. In 1985, one-half the 18 million registrants were referred to some job, one-half of the referrals were hired, and two-thirds of these jobs lasted more than six months. Some fraction of the registered workers, perhaps as many as 25 percent, were required to register by law in order to receive unemployment insurance payments; a portion of these were temporarily laid off and not interested in referrals. About 40 percent of the minority registrants were referred, and about 60 percent of these were hired. So the overall employment rate for minorities was about the same as for the majority group. However, one-half of the females and minority group members hired were in the lowest wage group, compared with 25 percent of majority males hired, so there is some difference in job quality by sex and race.

TABLE 1 Expected performance (EP) of referred applicants, measured in majority group standard deviations from the overall mean, and percent minority (PM) in the group referred

For 20 percent minority applicants

	Correlation											
	.2				.4				.6			
	Referral Ratio											
	1:10		1:2		1:10		1:2		1:10		1:2	
Referral Rule	EP	PM	EP	PM	EP	PM	EP	PM	EP	PM	EP	PM
Raw-score, top-down	.37	3	.17	8	.74	3	.34	8	1.12	3	.51	8
Within-group, top-down	.35	20	.16	20	.70	20	.32	20	1.05	20	.48	20
Minimum competency (30% cutoff)	.11	11	.11	11	.22	11	.22	11	.33	11	.33	11
Random within-zone (five zones)	.30	4	.17	9	.60	4	.34	9	.90	4	.51	9
Preferential within-zone (five zones)	.30	8	.17	11	.60	8	.34	11	.90	8	.51	11
Expected performance ratio	.36	15	.16	17	.73	11	.34	15	1.11	7	.51	13

TABLE 2 Expected performance (EP) of referred applicants, measured in majority group standard deviations from the overall mean, and percent minority (PM) in the group referred

For 40 percent minority applicants

	Correlation											
	.2				.4				.6			
	Referral Ratio											
	1:10		1:2		1:10		1:2		1:10		1:2	
Referral Rule	EP	PM	EP	PM	EP	PM	EP	PM	EP	PM	EP	PM
Raw-score, top-down	.39	9	.18	22	.77	9	.36	22	1.16	9	.54	22
Within-group, top-down	.35	40	.16	40	.70	40	.32	40	1.05	40	.48	40
Minimum competency (30% cutoff)	.11	29	.11	29	.22	29	.22	29	.33	29	.33	29
Random within-zone (rive zones)	.32	12	.17	23	.64	12	.33	23	.95	12	.50	23
Preferential within-zone (five zones)	.31	22	.16	28	.63	22	.32	28	.94	22	.48	28
Expected performance ratio	.37	32	.17	36	.75	24	.34	32	1.15	18	.53	28

TABLE 3 Expected performance (EP) of referred applicants, measured in majority group standard deviations from the overall mean, and percent minority (PM) in the group referred

For 60 percent minority applicants

	Correlation											
	.2				.4				.6			
	Referral Ratio											
	1:10		1:2		1:10		1:2		1:10		1:2	
Referral Rule	EP	PM	EP	PM	EP	PM	EP	PM	EP	PM	EP	PM
Raw-score, top-down	.39	20	.18	42	.79	20	.35	42	1.18	20	.53	42
Within-group, top-down	.35	60	.16	60	.70	60	.32	60	1.05	60	.48	60
Minimum competency (30% cutoff)	.11	50	.11	50	.22	50	.22	50	.32	50	.32	50
Random within-zone (five zones)	.31	27	.18	42	.63	27	.35	42	.94	27	.53	42
Preferential within-zone (five zones)	.30	53	.15	50	.61	53	.31	50	.91	53	.46	50
Expected performance ratio	.37	51	.17	56	.76	42	.34	52	1.16	34	.53	48

TABLE 4 Mean GATB scores expressed in standard deviation from the majority mean, by job family, for blacks and Hispanics

Job Family		Percent of Job Orders	Deviation from Majority Mean	
			Hispanic	Black
I	Set up machinery	1	-0.6	-1.0
II	Feeding, offbearing	2	0.0	-0.4
III	Professional, supervisor	1	-0.8	-1.2
IV	Skilled trades, clerical	23	-0.6	-1.0
V	Semiskilled	73	-0.3	-0.8

Notes: Data from North Carolina Employment Security Commission, January-June 1987. The standard deviation is derived from the entire population of people and of jobs.

The differences between racial and ethnic groups for the GATB scores in five job families are estimated by the USES from norm groups of employed workers studied by the USES test development program since 1972. There are 8,000 blacks, 2,000 Hispanics, and 20,000 "others" in the norm groups.

The distribution of job orders over the five job families illustrated in Table 4, is based on data from the North Carolina Employment Security Commission for the first half of 1987. The distribution may be quite different in other states. We can expect, however, that most USES jobs are in families IV and V, and that the shift in mean scores is about 0.5 standard deviations for Hispanics and 1 standard deviation for blacks.

Employment Service Departures from the Theoretical Assumptions

We evaluated the referral rules under a set of four theoretical assumptions that may or may not match the population of Employment Service registrants. At this time, the committee lacks

empirical data to evaluate those assumptions, but this section discusses how the assumptions might be violated in the course of Employment Service operations.

(1) The employment decision is based only on the GATB score for a particular job family, and, if pertinent, on minority or majority group membership. In the Employment Service, many factors other than the GATB score and racial or ethnic group influence which applicants are referred and finally employed. Some employers specify required skills and experience in their job orders, and some require further tests and interviews of referrals. The people hired may be discharged for unsatisfactory work performance after a short time on the job. Registrants may decline the referral or the job, or may apply directly to the job even if not referred. Registrants are referred to several jobs. Employers and applicants have constraints on wages, location, working hours, and working conditions.

Thus, the effect of GATB-based referral rules must be evaluated in the context of the other factors that determine who applies and who is hired.

(2) The test score and job performance have jointly normal distributions with unit standard deviations and the same correlation for the majority group and the minority group. One might expect scores on a many-item test such as the GATB, for a homogeneous unselected population, to be nearly normally distributed. There is no similar reasoning supporting normality for the measures of job performance in USES studies, which are frequently supervisor ratings with three or five values.

If indeed the GATB predicts performance, one would expect, all things being equal, that workers of high GATB score and high performance are more likely to be currently employed, so that the distribution in USES registrants would be skewed towards the lower scores. The performance gains expected from selecting in GATB order would in that case be reduced, since the highly productive workers would, at least in times of low unemployment, most likely be already employed. The gains in performance in the theoretical models are made by selecting workers starting from the upper tail of the GATB distribution, but there are only so many workers in the upper tail.

We have not yet seen data on actual distributions of GATB scores in Employment Service registrants, but it appears likely

that the expected performance gains for the various rules would be lower than indicated in the theoretical analysis, which is based on the normal assumptions.

(3) The same regression line formula predicts mean performance for a given test score in both the majority and the minority groups. The validity of GATB scores for predicting performance, and the possibility that different predictive formulas apply to different minority groups are two of the major research questions to be investigated by the committee. In 51 validity studies supported by USES (U.S. Department of Labor, 1983a), there was no overall significant difference in the predictive regressions of job performance between blacks and whites. But the research is not definitive. Questions remain about the data connecting test scores and performance. For example, the measure of performance in many of the validity studies is a supervisor rating of workers as good, fair, or poor; such a rating system may be biased against minority workers. In addition, if the GATB becomes the principal basis for referrals, its validity might be affected by coaching or practice effects that could act unfavorably on minority groups. Tests administered in English may be biased against Hispanic applicants, depending on the requirements of the jobs to which they are being referred.

(4) The mean test score for the minority group is one standard deviation below the mean test score for the majority group. Most of the research on group differences in test scores has tended to report black/white differences of one standard deviation. The data on Hispanic subgroups is often spotty and less clear. The value of one standard deviation between black and white performance is also suggested by the USES norm groups that were developed as part of the conversion tables used to derive within-group percentile scores. Hispanic performance in these tables is about .50 below the majority mean.

There may be significant differences between employed workers and USES registrants. The unemployment rate is higher for minorities, and there are higher fractions of minorities among USES registrants than in the USES norm group. Therefore, one might expect the differences between minority and majority test scores to be lower in the employed norm group than in the USES registrant group, if indeed the test scores are correlated with performance. If this is the case, using *employed* workers as a norm

group will cause minority *registrants* to be referred at lower rates than others despite the intentions of the designers of the new GATB-based referral system.

The opposite effect could occur if the GATB test score distribution is not the same as the norm group for each specific job or for every local pool of registered applicants. The norm groups are for job families, but the applicant groups are for particular jobs. There are many very different jobs in a job family; the following list shows the 17 jobs most frequently filled by public employment offices during the first half of 1987 in Michigan, North Carolina, and Virginia:

Family	*Job Title*
V	Material handler
V	Construction worker II
V	Construction worker I
V	Hard packer
V	Cleaner, commercial
V	Clerk, general
V	Sewing machine operator
V	Security guard
V	Kitchen helper
IV	Cashier
IV	Clerk typist
IV	Waiter
IV	Cleaner, housekeeping
IV	Administrative clerk
IV	Cook, hotel
IV	Retail sales clerk
IV	Telephone solicitor

The average GATB score for workers in a particular job may vary widely over jobs in a job family; and the difference in average GATB score between minorities and others might be substantially less within some jobs than for the job family as a whole. Unless a score correction to eliminate adverse impact in a particular job is based on the *within-job* average difference in score, rather than the *within-family* difference, minority workers may be selected in higher proportions than they are present in the applicant group for all jobs in the job family.

Suppose, for example, that for the Family IV job, hotel cook, the average GATB score for blacks is .5 standard deviations below the average score for others, while the average GATB score for blacks for all Family IV jobs is 1 standard deviation below the average score for others. (We assume that the cook applicants have the same average score differences as employed cooks.) For the within-group percentile referral rule, black applicants would have 1 standard deviation added to their scores to adjust for the familywide score difference. Then 25 percent of those referred for this job would be black, although 20 percent of the applicant group is black.

DISCUSSION OF THE REFERRAL RULES

Findings

Raw-Score, Top-Down Referral

This rule gives the highest expected performance in the referred group and the lowest minority group proportion referred. For example, if the applicant group is 20 percent minority, the correlation between test score and job performance is .4, and 1 in 2 are referred, then the expected performance is .34 standard deviations higher than would result from random selection (see Table 1).

For the Employment Service, the gain in performance will be reduced if the high tail of the assumed normal distribution of GATB scores is attenuated by prior hiring of the more productive workers with their higher expected GATB scores. There are other reasons to expect that the performance gains would be more modest in the Employment Service context. Raw score referral and the other rules consider applicants for a single job, but the Employment Service has many jobs available. Although raw score referral provides the highest expected performance for any single job, it does not address the situation in which several job referrals are possible for each registrant and an allocation decision must be made. Maximizing total expected job performance of all referrals to all jobs would require that the highest scoring applicants be allocated to the most difficult or most critical jobs rather than simple referral to a particular job in order of expected performance of registrants.

Raw-score referral results in few minority referrals; in the example above, with moderate validity and a high referral ratio, only 8 percent of those referred to jobs would be minority. The effect is even more extreme when the referral ratio is low and the correlation between test and job performance is low. If the referral ratio is 1 in 10 and the correlation is .2, only 3 percent of those referred will be minority. However, given the low validity, many of the minority workers excluded would have performed just as well as the majority workers included. Whatever the correlation between test and job performance, this referral method would have significant adverse impact, and employers and the Employment Service would be subject to discrimination suits.

Within-Group Percentile Score, Top-Down Referral

This rule achieves the highest proportions of minority referrals, with slight overall losses in expected performance.

The loss in performance compared to raw score referral is slight unless test score and performance are highly correlated. For example, when the applicant group is 20 percent minority, the correlation is .4, and 1 in 2 are referred, the expected performance is .32 standard deviations higher than would result from random selection, which is negligibly different from the .34 standard deviation gain obtained for raw score referral. Consider another case: when an applicant group is 60 percent minority, the correlation is .6, and the referral ratio 1 in 2, the expected performance is .48 standard deviations higher than random selection, compared with .53 standard deviations higher for raw score referral. This case is relevant because of the population mix in certain geographical locations. In addition, it is prudent to expect many groups to press for within-group referrals or other advantages in the scoring if the GATB referral program is adopted nationally.

When the norm group is the applicant group, as is assumed in the computation, the proportion of minorities referred is the same as the proportion of minorities in the applicant group. In Employment Service operations, there will be considerable variation in the applicant groups for various jobs, and so the USES norm group of employed workers may be quite dissimilar to the applicant group in any particular job. It remains to be discovered which proportions of minorities will be referred for various jobs and what the average proportion will be for all jobs. Since within-group referral

is, for blacks, equivalent to raw-score, top-down referral with 1 standard deviation added to each minority score, its effect will always be to substantially increase minority referrals.

Minimum Competency Referral

We have used a 30 percent cutoff, corresponding to the fraction of Employment Service registrants who were not referred to any job in 1985 (discounting Unemployment Compensation registrants).

Expected performance is notably worse than all other referral rules for every combination of correlation, referral ratio, and minority proportion in the applicant group. For example, when the applicant group is 20 percent minority, the correlation is .4, and 1 in 2 are referred, expected performance is .22 standard deviations higher than random selection would give, compared with the .34 standard deviation gain obtained for raw-score referral. The poor theoretical performance of minimum competency referral derives in part from the assumption that only the GATB score and racial or ethnic group membership affects the referral decision. In practice, both employer and applicant will constrain referrals: the employer may require prior work experience and will interview referrals, so that the actual referrals will not be randomly selected from applicants with scores above the cutoff, the people hired will not be randomly selected from referrals, and the performance of those hired will be higher than that for those referred. Nevertheless, there are real performance gains for each employer in being referred workers by any of the rules *other* than minimum competency.

The proportion of minorities referred increases noticeably compared with raw-score referral. For example, if half the applicants are referred and 20 percent of the applicants are minorities, then 11 percent of the referred applicants will be minorities, compared with 8 percent for raw-score referral.

This rule achieves the highest minority referral rates of non-race-conscious rules, at substantial cost in expected performance.

Zone Score, Random Within-Zone Referral

We used five zones. Expected performance is comparable to raw-score referral when the applicant pool is small so that the

referral ratio is high, say, 1 in 2, but substantially worse than raw-score referral when there are large numbers of people available and the referral ratio is low, say, 1 in 10. When the referral ratio is 1 in 10, random within-zone referral with five zones is selecting at random from the highest 20 percent; for raw score referral, the selection would be from the highest 10 percent. If 10 zones are used, then zone referral would do as well as raw-score referral for a referral ratio of 1 in 10.

The proportion of minorities referred increases negligibly in comparison with raw-score referrals.

Zone Score, Preferential Within-Zone Referral

This rule behaves similarly to random within-zone referral, with slight losses in performance and slight gains in the proportions of minority applicants referred.

Expected Performance Ratio Referral

This rule produces expected performance negligibly different from raw-score referrals for all combinations of test-performance correlations, referral ratios, and minority group proportions.

The minority proportions referred under this rule are the second highest, after within-group referral; they are significantly lower than within-group referral when the correlation between test and performance is high and the referral ratio is low. For example, for 20 percent minority applicants, a correlation of .4, and a referral ratio of 1 in 2, this rule results in 15 percent minority group referrals, compared with 20 percent for the within-group system; a correlation of .4 and a referral ratio of 1 in 10 results in 11 percent minority group referrals. At the same time, the gain in expected performance over the within-group referral rule is slight, although it increases with the correlation.

A feature of the procedure is that minority group referrals are nearly in proportion to their number in the applicant pool when test-performance correlations are low, but are greatly reduced when the correlations reach .6. It is technically more difficult to do referrals this way, since it is necessary to estimate correlations between test scores and performance for each job, as well as differences in ethnic group mean test scores for each job.

Comparison of the Referral Rules

The tables show that expected job performance improves with higher correlation and lower referral ratio for all rules but one. The exception is minimum competency referral, which eliminates the advantage of lower referral ratios.

The principal referral-rule competitors for high performance are raw-score, top-down; within-group percentile score, top-down; and expected performance ratio. The remaining referral rules have distinctly lower performance when the referral ratio is low; for a referral ratio of 1 in 2, however, all referral rules but minimum competency have about the same performance. Minimum competency referral loses much of the advantage of being able to predict performance from the test score.

The principal competitors for producing a high proportion of minority applicants in the referred group are within-group percentile and performance ratio referral. Raw-score referrals contain the smallest minority proportions of all rules.

The zone rules are dominated by the performance ratio rule in that higher performance and higher minority proportions occur in performance ratio referrals for all combinations of parameters.

Raw-score, random-zone, and minimum competency are the rules that are not race conscious. Of these, the minimum competency rule produces the highest minority proportions, but at a very great cost in performance under the theoretical assumptions posited here.

Within-group percentile referral, which is race conscious, guarantees high minority proportions in the referred group with little loss in performance.

A Compromise Referral Policy

The committee is considering a number of referral rules that would allow employers to strike an appropriate compromise between the interests of productivity and racial balance in the workforce. To illustrate, a referral policy might use both across-group and within-group scoring methods to create the referral pool. The pool would be assembled in two stages. In the first stage, applicants would be chosen strictly in order of test score, as in the raw-score referral rule. Enough applicants would be chosen to meet an employer's job order. In the second stage, all minority

applicants would be added to the pool who would have been chosen if the within-group referral rule had been used. Raw scores would be reported to the employer.

Under such a compromise policy, employers could ignore race and select solely on the basis of predicted performance, or they could also select from the enriched pool that includes the most skilled minority group members available. From the point of view of the job seekers, there would be some equity gains. Any majority applicant who would have been selected on the basis of test scores alone is referred. Any minority applicant who would have been selected under the within-group percentile system is referred. The system is still race conscious, but it does not deny majority applicants opportunities they would otherwise have had.

Because the compromise plan combines straight rank ordering and within-group policies, it will yield a referral group that can be approximately as successful on the job as a group chosen by either of the other two policies. Moreover, this compromise plan would allow an employer to choose the balance between equity and productivity that is appropriate to the particular job and for the affirmative action policy of the company. We recognize that such a referral policy may pose practical difficulties, but it is important to explore all possible means of addressing the psychometric, legal, and economic requirements facing the Employment Service.

SCORE REPORTING

Although we make no claim that the assumptions supporting the above computations are well satisfied by Employment Service operations, a policy of referring applicants in rank order of their score on a *valid* test will ineluctably offer employers better performing workers, all other things being equal. At the same time, the computations make clear that a top-down referral system that incorporates within-group score conversions combines the advantages of productivity gains and racial equity in the workforce. However, the system of within-group percentile score reporting chosen by USES raises a number of legal and scientific concerns.

One major problem with reporting within-group percentile scores is that they are potentially misleading. While the purpose of the scoring method is to indicate an individual's predicted job performance with reference to other applicants within his own ethnic or racial group, employers may mistakenly infer that two

applicants with the same percentile score did equally well on the test, no matter what their racial or ethnic identity. Employers are not given the conversion tables and so have no way of determining the correspondence between scores obtained within different groups. This lack of information could lead employers to underestimate the magnitude of group differences in raw scores. (For example, on certain GATB composites, a raw score that places an applicant at the 50th percentile among blacks would place an applicant at the 16th percentile among whites.) It could also lead employers to underestimate the amount of overlap in test scores that exists between the groups.

Reporting raw scores is also potentially misleading. First, raw scores do not provide an employer who wants to carry out affirmative action goals with information about applicants' standing within their own group. Second, neither raw scores nor percentile scores provide an employer with information about the levels of job performance that can be expected from any particular range of scores. It is tempting for the employer to infer that a person at the 80th percentile of whatever norm group on the test score will be at the 80th percentile of the norm group in job performance. But the correspondence between test score percentile and job performance percentile depends on the correlation between test score and job performance. For example, if that correlation is .3, a person at the 80th percentile on the test score is expected to be at the 60th percentile on job performance. Finally, providing a score referenced to the total group without qualifying its relevance to a particular job could have a harmful effect on minority applicants who on the average score lower on the GATB. They will appear to be less qualified for the job, but the score may have little relevance to performance on the job.

The general principle is that employers should be given sufficient information to make correct inferences about job performance from the reported scores.

The drawbacks of both types of score reporting suggest that it might make sense to report both the within-group percentile and the raw test scores to employers, along with information about how relevant (valid) the test is for predicting performance on the jobs in question. With both types of scores in hand, employers could make selection decisions on the basis of relative performance within groups (to meet affirmative action goals), on the basis of absolute performance across groups (to maximize the job performance of

the selected group), or on the basis of a combination of these score types according to their needs. This type of score reporting would be appropriate if the compromise referral policy described above were adopted. In the current legal environment, however, providing both types of score may be subject to the Title VII injunction against racial classifications because the information would indirectly tell employers the race of the applicant.

Part III
Conclusions

In this interim report, the committee neither approves nor disapproves of the General Aptitude Test Battery (GATB), the theory of validity generalization, or the U.S. Employment Service's GATB-based validity generalization pilot referral programs. These issues, along with final conclusions about the GATB scoring system, remain under study by the committee.

The USES practice of converting an applicant's raw GATB score to a "black," "Hispanic," or "other" norm group percentile score is race conscious. Liaison group members and others have provided the committee with various reviews and interpretations of professional, legislative, and judicial commentaries on the use of race-conscious employment practices. At present, however, the committee has concluded that this evidence neither endorses nor proscribes unequivocally the use of such practices. Furthermore, it is beyond the committee's charge to determine the legality of race-conscious employment practices, let alone identify particular employment situations for which such use might be justified. Therefore, the committee's conclusions about the consequences of using race-conscious or non-race-conscious employment practices are similarly not to be interpreted as either approving or disapproving the use of such practices.

REFERRAL RULES

Two premises underlie our conclusions about referral rules:

- If test scores are positively correlated with job performance, selecting applicants with the highest scores will, other things being equal, contribute to a more productive workforce.
- When minority group members generally score lower on a test than do majority group members, a system that selects people for jobs in order of unadjusted test scores will have an adverse impact on the employment opportunities of minority group members.

We conclude:

1. If the will of society is to pursue both high levels of productivity and a racially balanced workforce and if a valid test that produces an adverse impact is used in the referral process, then a race-conscious referral policy is necessary.
2. While we do not at this time recommend a particular referral procedure, after considering a range of alternatives, some race conscious, some not, we find that a top-down within-group referral rule is an effective way to balance the conflicting goals of productivity and racial equity. This referral rule substantially reduces adverse impact with very little loss in the average predicted job performance of those referred. In contrast, the minimum competency rule, which has been used by the Employment Service with the SATBs, also reduces adverse impact, but with a substantially greater loss in predicted job performance.

However, the committee has a number of concerns regarding the current USES within-group referral rule. Its legal status has been challenged by the Justice Department because it is race conscious. From a psychometric perspective, this referral rule is insensitive to variations in test validities across jobs, differences between norm groups and applicant groups, regional differences in applicant populations, and other factors.

Because of these concerns, the committee is studying alternative referral methods.

SCORE REPORTING

Two principles underlie our conclusions about score reporting:

- It is misleading to report the same score for two individuals when the two individuals have widely different predicted levels of job performance.
- It is misleading to report widely different scores for two individuals when the two have very similar predicted levels of job performance.

We conclude :

1. The current U.S. Employment Service experimental policy of reporting only within-group scores risks violating the first principle above because employers might confuse within-group scores and raw or total-group scores.

2. We are concerned about the consequences of total-group score reporting or any method score reporting when a test is a poor predictor of job performance. A weak relationship between test and job performance is likely to result in larger differences in the reported scores of minority and majority group members than subsequent job performance would warrant.

3. We are not prepared in this interim report to approve or disapprove the current reporting practice, nor are we prepared to recommend an alternative. We considered recommending that USES report both total-group and within-group scores, a procedure that appears to combine the advantages of both scoring methods. However, it might be that in practice this procedure would combine the disadvantages of both scoring methods. Furthermore, such a procedure might not survive legal challenge because it would effectively reveal applicants' race to potential employers.

Our recommendations about referral systems and score reporting are contingent upon and must await a full evaluation of the GATB, the experimental referral program, and the social and economic consequences of the referral program.

References

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education

1985 *Standards for Educational and Psychological Testing.* Prepared by the Committee to Develop Standards for Educational and Psychological Testing. Washington, D.C.: American Psychological Association.

Hunter, John E.

1982 Test Validation for 12,000 Jobs: An Application of Job Classification and Validity Generalisation Analysis to the General Aptitude Test Battery (GATB). Report to the U.S. Employment Service.

Kelly, Brenda L.

1987 Letter from Test Research Specialist, Employment Division, Oregon Department of Human Resources, to Bill Simmons, Supervisor, Western Test Development Field Center, Utah Department of Employment Security (September 10).

Miller, Ann R., Donald J. Treiman, Pamela S. Cain, and Patricia A. Roos, editors

1980 *Work, Jobs, and Occupations: A Critical Review of the "Dictionary of Occupational Titles."* Committee on Occupational Classification and Analysis, Assembly of Behavioral and Social Sciences, National Research Council. Washington, D.C.: National Academy Press.

Quinlan, Pierce A.

1973 Implementation of Interim Policy on Use of SATB's in Selection. Field Memorandum No. 325-73, from Acting Associate Manpower Administrator for Field Director and Management to all Assistant Regional Directors for Manpower, Manpower Administration, U.S. Department of Labor (October 9).

Reynolds, Wm. Bradford

1986 Letter from Assistant Attorney General, Civil Rights Division, U.S. Department of Justice, to Richard Gilliland (Director, U.S. Employment Service, U.S. Department of Labor) (November 10).

Robertson, Peter C.
1976 Legislative History of the 1972 Amendment to the Civil Rights Act. Background paper prepared by Director, Office of Federal Liaison, Equal Employment Opportunity Commission, for H.L. Krieger, Director, Federal Personnel Compensation Division, General Accounting Office (August 29).

Schmidt, Frank L., and John E. Hunter
1977 Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology* 62:529-540.
1981 Employment testing: old theories and new research findings. *American Psychologist* 36:1128-1137.

Society for Industrial and Organizational Psychology, Inc.
1987 *Principles for the Validation and Use of Personnel Selection Procedures, 3d ed.* College Park, Md.: Society for Industrial and Organizational Psychology, Inc.

U.S. Department of Labor
1983a Fairness of the General Aptitude Test Battery: Ability Differences and Their Impact on Minority Hiring Rates. USES Test Research Report No. 46. Division of Counseling and Test Development, Employment and Training Administration. Washington, D.C.: U.S. Department of Labor.
1983b Overview of Validity Generalization for the U.S. Employment Service. USES Test Research Report No. 43. Division of Counseling and Test Development, Employment and Training Administration. Washington, D.C.: U.S. Department of Labor.

Wigdor, Alexandra K., and Wendell R. Garner, editors
1982 *Ability Testing: Uses, Consequences and Controversies. Part I: Report of the Committee. Part II: Documentation Section.* Committee on Ability Testing, Assembly of Behavioral and Social Sciences, National Research Council. Washington, D.C.: National Academy Press.

Willingham, W.W., and H.M. Breland
1982 *Personal Qualities and College Admissions.* New York: College Entrance Examination Board.

Appendix A
Analytic Methods Used to Compute Effects of Referral Rules

This appendix describes the calculation of the values in Tables 1-3. All of the calculations were based on a model in which test scores are normally distributed within the majority and minority groups, respectively. The (population) standard deviation of the test scores was set to 1 within both the majority and the minority group. The (population) mean test score within the majority group was set to 0 and the (population) mean within the minority group was set to -1. Because all of the selection procedures considered are invariant with respect to location and scale of test scores, this selection of means and standard deviations reflects the performance of the selection procedure for any linear rescaling of test scores in which the majority group outperforms the minority group by one standard deviation.

The calculations assume a common linear regression of job performance on test score, and they assume that the test has equal validity for minority and majority groups. Furthermore, it is assumed that test score and job performance are bivariate normal within the majority group and within the minority group. In all situations, the expected job performance is simply the product of the test validity and the mean performance of the applicants selected; this follows from the assumption that the employer selects at random from the referred group. The value given in the tables is the difference between expected job performance for a particular rule and expected job performance if referral were at random from the whole population. (Under random referral, expected job

performance is 0 (for the majority group) minus the product of the minority group proportion and test validity, since the minority is assumed to have average test scores of -1.)

The values for mean performance and for the proportion of minority group members selected were calculated analytically on the basis of the above assumptions about score distributions and the relationship between score distributions and performance. We do not defend these assumptions as true of all or even *any* actual test data. We present these assumptions and the calculations based on them as an idealized, but not unreasonable, model for examining the gross properties of various referral rules.

In the discussion that follows we use the random variable X to represent the majority group test scores and the random variable Y to represent the minority group test scores. In conventional notation, our distributional assumptions for X and Y are therefore that

$$X \sim N(0, 1)$$

and

$$Y \sim N(-1, 1).$$

Let π be the proportion of the total applicant pool that is minority. Then a randomly selected applicant has probability π of being a member of the minority group and probability $(1 - \pi)$ of being a member of the majority group. Letting random variable Z represent the test scores of the entire applicant pool (minority and majority group members combined), we see that Z is a mixture of X and Y with mixing coefficients π for the minority group distribution and $(1 - \pi)$ for the majority group distribution.

RAW-SCORE, TOP-DOWN REFERRAL

Let $F_X(x)$, $F_Y(y)$, and $F_Z(z)$ represent the cumulative distribution functions of X, Y, and Z respectively. For a given selection ratio p the mean test score of selected applicants was obtained by first calculating the cutoff score, $c = F_Z^{-1}(1 - p)$. The mean test performance of selected applicants was then computed as $E[Z|Z > c]$. The proportion of minority applicants selected was calculated as the conditional probability of being a minority given selection:

$$P\{\text{Minority} \mid \text{Selection}\} = \frac{[1 - F_Y(c)]\pi}{[1 - F_Y(c)]\pi + [1 - F_X(c)](1 - \pi)}.$$

WITHIN-GROUP PERCENTILE SCORE, TOP-DOWN REFERRAL

Within-group ranking for a selection ratio p explicitly selects the top 100p percent minority applicants and the top 100p percent majority applicants. Hence the proportion of those selected that is minority will always be π (except for sampling variation). The cutoff scores for majority applicants will therefore be

$$c_X = F_X^{-1}(1 - p)$$

and the cutoff score for minority applicants will therefore be

$$c_Y = F_Y^{-1}(1 - p).$$

The mean test score of all applicants selected will be a weighted combination of the means of minority and majority applicants selected:

$$(1 - \pi)E(X \mid X > c_X) + \pi E(Y \mid Y > c_Y).$$

The proportion of the selected group who are minorities will always be π.

MINIMUM COMPETENCY REFERRAL

In this model the test is scored dichotomously. Applicants are selected randomly from the group of all applicants with test scores above a certain percentile 100p. Consequently the mean score and proportion of minority applicants selected is identical to those obtained with selection based on raw scores with a selection ratio of (1 - p).

ZONE SCORE, RANDOM WITHIN-ZONE REFERRAL

Zone scoring selects from the highest test score downward by zones. The critical element in the computation is the lowest

zone from which applicants are selected, which depends on the number of zones, k, and the selection ratio, p. Denote this zone the (j + 1) zone. To select a proportion p of the total applicant population, we select a proportion j/k from the first j zones that are completely filled. That leaves (p - j/k) to be selected from the (j + 1) zone. The proportion of the total scores in zone (j + 1) is 1/k by definition. Since the proportion of the total scores to be selected from zone (j + 1) is only (p - j/k < 1/k), we select [(p - j/k)/(1/k)] = (kp - j) of the scores in zone (j + 1) at random.

Given k zones, the test scores representing the boundaries of the zones are

$$c_{k-1} = F_Z^{-1}(1/k), c_{k-2} = F_Z^{-1}(2/k), \ldots, c_1 = F_Z^{-1}[(k-1)/k].$$

The average test score for all applicants selected was obtained by first calculating the expected test score separately for the selected minority applicants and then for the selected majority applicants. These were then combined to yield the expected test score for all applicants selected. The expected value of the test score for the selected majority applicants is

$$\frac{E(X \mid X > c_j)P(X > c_j) + E(X \mid c_{j+1} < X < c_j)P(c_{j+1} < X < c_j)(kp - j)}{P(X > c_j) + P(c_{j+1} < X < c_j)(kp - j)}$$

and the expected value of the test score for minority applicants selected is

$$\frac{E(Y \mid Y > c_j)P(Y > c_j) + E(Y \mid c_{j+1} < Y < c_j)P(c_{j+1} < Y < c_j)(kp - j)}{P(Y > c_j) + P(c_{j+1} < Y < c_j)(kp - j)}$$

The probability π_M of being selected given that the applicant is a minority is

$$\pi_M = [P(\text{Selection}|\text{minority})\ P(\text{minority})] / \{[P(\text{Selection}|\text{minority})\ P(\text{minority})] + [P(\text{Selection}|\text{majority})\ P(\text{majority})]\},$$

that is,

$$\pi_M = \frac{\alpha\pi}{\alpha\pi + \beta(1-\pi)},$$

where

$$\alpha = P(Y > c_j) + P(c_{j+1} < Y < c_j)(kp - j)$$

and

$$\beta = P(X > c_j) + P(c_{j+1} < X < c_j)(kp - j).$$

ZONE SCORE, PREFERENTIAL WITHIN-ZONE REFERRAL

This model is the same as the previous model except that in the final zone, minorities are selected at random first and majority group members are selected only after the pool of minority applicants has been exhausted. The computations are similar to those for zone scoring with random selection within zones except that the probability of selection within the final zone differs for majority and minority group members. It is useful to compute this probability separately in two cases.

Case 1

If the proportion (p - j/k) of the total group to be selected in the final zone exhausts the minorities, that is, if

$$\pi P(c_{j+1} < Y < c_j) < p - j/k,$$

then the conditional probability of selection given minority status is just $P(Y > c_{j+1})$. In this case the overall proportion of scores left to be selected from the majority group in zone (j + 1) is

$$(p - j/k) - \pi P(c_{j+1} < Y < c_j).$$

Consequently, the proportion of the majority group selected in zone (j + 1) is

$$\pi^*_M = \frac{p - j/k - \pi P(c_{j+1} < Y < c_j)}{(1 - \pi)P(c_{j+1} < X < c_j)}.$$

Therefore, the expected test score for minority applicants selected is just $E(Y|Y > c_{j+1})$ and the expected test score for majority applicants selected is

$$\frac{E(X \mid X > c_j)P(X > c_j) + E(X \mid c_{j+1} < X < c_j)P(c_{j+1} < X < c_j)\pi^*_M}{P(X > c_j) + P(c_{j+1} < X < c_j)\pi^*_M}.$$

The proportion of minorities selected is just the conditional probability of being selected given that the applicant is a minority, or

$$\pi_M = \frac{\alpha\pi}{\alpha\pi + \beta(1-\pi)}$$

where

$$\alpha = P(Y > c_{j+1})$$

and

$$\beta = P(X > c_j) + P(c_{j+1} < X < c_j)\pi^*_M.$$

Case 2

If the proportion (p-j/k) of the total group to be selected in the final zone does not exhaust the minorities, that is, if

$$\pi P(c_{j+1} < Y < c_j) > p - j/k,$$

the situation can be handled by interchanging the labels of the groups and using the method given for Case 1.

EXPECTED PERFORMANCE RATIO REFERRAL

This method uses overall ranking on a modification of the observed test score as the basis for selection. It is equivalent to subtracting $\mu(1 - \rho)$ from each score where μ is the mean of the group from which the score was obtained and ρ is the validity of the test. Let the random variable X^* represent the modified test scores of the majority group and the random variable Y^* represent the modified test scores of the minority group. Since the majority group mean is set to 0 and the minority group mean is set to -1 in our model,

$$X^* = X \sim N(0, 1)$$

and

$$Y^* \sim N(-\rho, 1).$$

Using the random variable Z^* to represent the modified test scores of the entire applicant pool (minority and majority group members combined), we see that Z^* is a mixture of X^* and Y^* with mixing coefficients $(1 - \pi)$ for the majority group and π for the minority group. Let $F_{Y^*}(y)$ and $F_{Z^*}(z)$ denote the cumulative distribution functions of Y^* and Z^*, respectively.

For a given selection ratio p, we first computed the cutoff score in the modified score distribution, $c^* = F_{Z^*}^{-1}(1 - p)$. The proportion of the selected group who are minorities is just the conditional probability of selection given minority status or

$$\pi_M = \frac{[1 - F_Y(c^*)]\pi}{[1 - F_Y(c^*)]\pi + [1 - F_X(c^*)](1 - \pi)}.$$

Computations of expected job performance are based on the unmodified test scores. The expected unmodified test score of the applicants selected is

$$E(X \mid X > c^*)(1 - \pi_M) + E[(Y \mid Y > (c^* + \rho - 1)]\pi_M.$$

Appendix B
Biographical Sketches, Committee Members and Staff

JOHN A. HARTIGAN is the Eugene Higgins professor of statistics and director of the Statistical Computing Laboratory at Yale University. His teaching and research interests center on the foundations of probability and statistics, Bayes theory, classification, statistical computing, and graphical methods. He is an elected fellow of the American Statistical Association and the Institute of Mathematical Statistics and an elected member of the International Statistical Institute and the Royal Statistical Society. He received BSc and MSc degrees in mathematics from the University of Sydney and a PhD degree in mathematical statistics from Princeton University.

LORRIE A. SHEPARD is professor and chair of Research and Evaluation Methodology in the School of Education at the University of Colorado. She is past president of the National Council on Measurement in Education and past editor of both the *Journal of Educational Measurement* and the *American Educational Research Journal*. Her research includes applied psychometric studies, aimed at topics such as standard setting and bias detection, and policy studies addressing issues of test use. She received a BA degree from Pomona College in history, and an MA degree in counseling and a PhD degree in educational research from the University of Colorado.

MARCUS ALEXIS is dean of the College of Business Administration at the University of Illinois, Chicago. An expert on

decision making, marketing, and the economic role of minorities, he has taught at Macalester College, De Paul University, University of Rochester, and Northwestern University and been a Ford Foundation Faculty Study Fellow at Harvard University and the Massachusetts Institute of Technology. Dr. Alexis was a commissioner of the Interstate Commerce Commission from 1979 to 1981. He is currently deputy chairman of the Federal Reserve Bank of Chicago, a trustee of the Teachers Insurance & Annuity Association (TIAA), and a member of the board of directors of the Metropolitan Planning Council for the City of Chicago. He is a member of the Board of Governors of the Beta Gamma Sigma society, a member of the American Economic Association, the National Economic Association, and the Econometric Society, and a past member of the board of the Caucus of Black Economists. He is on the editorial board of the *Review of Black Political Economy* and has served on the board of economists of *Black Enterprise*. He received AB and honorary Doctor of Humane Letters degrees from Brooklyn College, an MA degree from Michigan State University, and the PhD degree in economics from the University of Minnesota.

MANFRED EMMRICH is director of the North Carolina State Employment Service Division of the Employment Security Commission of North Carolina, responsible for the operation of employment services provided by 84 local Job Service Centers and branch offices across the state. From 1978 to 1985, Emmrich served as a senior associate with MDC, Inc., a nonprofit research and development group with special interests in productivity and employment and training issues. Emmrich was chair of the Employment Security Commission of North Carolina from 1973 to 1978, responsible for the state's employment service, unemployment insurance, and labor market information programs. From 1962 to 1973, Emmrich held various management and executive positions in the Macke Company. Prior to entering the private sector, Emmrich served in the U.S. Army as an officer in Army Intelligence and Security. A former president of the Interstate Conference of Employment Security Agencies, Inc., Emmrich currently serves on that organization's Employment and Training Committee. Emmrich has a BA degree in economics from Davidson College.

LARRY V. HEDGES is associate professor and chair of the Measurement Evaluation and Statistical Analysis Program in the Department of Education at the University of Chicago. His research is concerned with statistical methods for combining evidence from replicated research studies, statistical models in cognitive psychology, and the social psychology of scientific research. He is the coauthor (with Ingram Olkin) of *Statistical Methods for Meta-Analysis* and is the associate editor of the *Journal of Educational Statistics* and *Psychological Bulletin*. He received a BA degree from the University of California, San Diego, and MS and PhD degrees from Stanford University, all in statistics.

IRA J. HIRSH is Edward Mallinckrodt Distinguished University Professor of psychology and audiology at Washington University, where he has also been chair of the Department of Psychology and dean of the Faculty of Arts and Sciences. His research on speech, hearing, and deafness has been carried out at the Central Institute for the Deaf, where he was formerly director of research and is presently senior research scientist. He is a member of the National Academy of Sciences and a fellow of the Acoustical Society of America (past president), American Psychological Association, and American Speech and Hearing Association. From 1982 to 1987 he was chair of the Commission on Behavioral and Social Sciences and Education of the National Research Council. He has an AB degree from the New York State College for Teachers at Albany, an MA degree from Northwestern University, and a PhD degree in experimental psychology from Harvard University.

RICHARD M. JAEGER is professor of education and director of the Center for Educational Research and Evaluation at the University of North Carolina, Greensboro. His research is concerned with educational measurement and applied statistics. He is coeditor of *Minimum Competency Achievement Testing* (1980), author of *Statistics: A Spectator Sport* (1983) and *Sampling in Education and the Social Sciences* (1984), and editor of *Complementary Methods for Research in Education* (1987). He is past editor of the *Journal of Educational Measurement* and on the editorial boards of several journals. He is past president of the National Council on Measurement in Education and a member of the American Statistical Association, the American Educational Research Association, and the American Evaluation Association. He received a BA degree in mathematics from Pepperdine College and MS and

PhD degrees in mathematical statistics and educational research methodology, respectively, from Stanford University.

STEPHEN P. KLEIN is a senior research scientist with the RAND Corporation where he directs policy research studies in the fields of education, health, and criminal justice. He also serves as a consultant to several professional licensing boards on matters relating to testing. He is a member of the American Psychological Association, the American Educational Research Association, and the National Council on Measurement in Education. He received a BS degree from Tufts University and MS and PhD degrees in industrial psychology from Purdue University.

ROBERT L. LINN is professor of education at the University of Colorado. His research is directed at applied and theoretical problems in educational and psychological measurement. He is a former president of the Division of Evaluation and Measurement of the American Psychological Association, former president of the National Council on Measurement in Education, and former vice president of the American Educational Research Association for the Division of Measurement and Research Methodology. He has served as editor of the *Journal of Educational Measurement* and was vice chair of the committee that developed the 1985 *Standards for Educational and Psychological Testing*. He received an AB degree in psychology from the University of California, Los Angeles, and an MA degree in psychology and a PhD degree in psychological measurement from the University of Illinois, Urbana-Champaign.

JOHN M. RAUSCHENBERGER is a personnel research consultant in the Workforce Research and Selection Systems Section, Employee Development Office, Ford Motor Company. He is a member of the American Psychological Association, the Society for Industrial and Organizational Psychology, the Academy of Management, and the Equal Employment Advisory Council's Subcommittee on Employee Selection Procedures. He serves as a special reviewer for the *Journal of Applied Psychology* and is on the editorial board of *Personnel Psychology* journal. He received a BS degree in psychology and MA and PhD degrees in industrial psychology, all from Michigan State University.

MICHAEL ROTHSCHILD is professor of economics and dean of social sciences, University of California, San Diego. His research

concerns the economics of information, financial economics, law and economics, and industrial organization. He is a member and fellow of the Econometric Society and a member of the American Economic Association. He received a BA degree from Reed College, an MA degree from Yale University, and a PhD degree in economics from the Massachusetts Institute of Technology.

PAUL R. SACKETT is associate professor of psychology at the University of Illinois, Chicago; he was previously on the faculty of the School of Business at the University of Kansas. He has published extensively in the areas of assessment of managerial potential, job analysis, honesty in the workplace, and psychometric issues in employee selection. He is coauthor (with George F. Dreher) of *Perspectives On Employee Staffing and Selection* (1983), editor of *Personnel Psychology*, and serves on the editorial board of the *Journal of Applied Psychology*. He received a PhD degree in psychology from the Ohio State University.

O. PETER SHERWOOD is solicitor general of New York State. A litigator, he has tried cases and argued appeals in many state and federal courts, including several cases involving challenges to employment testing practices under Title VII. Before joining the Office of the New York Attorney General, he was an assistant counsel at the NAACP Legal Defense and Educational Fund, Inc., where his practice was focused on fair employment practices litigation. Until 1987 he was an adjunct assistant professor of law at the New York University School of Law, where he taught constitutional law and fair employment practices law. He received a BA degree from Brooklyn College of the City University of New York and a JD degree from New York University School of Law.

HOWARD F. TAYLOR is professor of sociology at Princeton University. His research interests encompass the methodology of test score heritability estimation, social psychology, and race and ethnic relations. His books include *The IQ Game: A Methodological Inquiry Into the Heredity-Environment Controversy* (1980) and *Balance in Small Groups* (1978). He is a member of the American Sociological Association, a fellow of the Sociological Research Association, vice president of the Eastern Sociological Society, and he has been a member of the editorial boards of several journals. He received a BA degree from Hiram College and MA and PhD degrees in sociology from Yale University.

ALEXANDRA K. WIGDOR is study director of the Committee on the General Aptitude Test Battery and also serves as study director of the Committee on the Performance of Military Personnel. Previously, as study director of the Committee on Ability Testing, she coedited (with Wendell R. Garner) *Ability Testing: Uses, Consequences, and Controversies* (1982). Trained as an historian, her research interests now include human performance assessment, the legal and social dimensions of psychological testing, and the development of governmental policy on testing and selection. A member of Phi Beta Kappa, she received BA and MA degrees from the University of Missouri, and studied further at the Free University of Berlin, the University of Maryland, and the Institute for Historical Research, University of London.

HILDA WING is research associate for the Committee on the General Aptitude Test Battery and the Committee on the Performance of Military Personnel. Previously she was with the Psychological Corporation, served as chief of the predictor development team at the U.S. Army Research Institute for the Behavioral and Social Sciences, and was research psychologist at the U.S. Office of Personnel Management. Her primary area of expertise is personnel testing. She is a member of the American Psychological Association, has served as chair of its Committee on Employment and Human Resources, and is currently chair of its Committee on Psychological Tests and Assessment. She received an AB degree in mathematics from Middlebury College and MA and PhD degrees in experimental psychology from the Johns Hopkins University.

Appendix C
Members, Liaison Group

ROBERT BOLDA, Personnel Research Division, General Motors Corporation, Detroit, Mich. (Ret.)
CLINT BOLICK, Office of the Assistant Attorney General (Civil Rights Division), U.S. Department of Justice
RALPH G. CANTRELL, Virginia Employment Commission, Richmond
WAYNE F. CASCIO, Graduate School of Business Administration, University of Colorado, Denver
SUZAN CHASTAIN, Office of Civil Rights, Office of the Solicitor, U.S. Department of Labor
CONSTANCE L. DUPRE, U.S. Equal Employment Opportunity Commission (Ret.)
PATRICIA J. DYER, IBM Corporate Employment and Placement, Armonk, New York
KENNETH EDWARDS, Skill Improvement Department, International Brotherhood of Electrical Workers, Washington, D.C.
MANFRED EMMRICH, North Carolina State Employment Service, Raleigh
BOB FUNSTON, Oklahoma Employment Security Commission, Oklahoma City
JOHN E. HUNTER, Department of Psychology, Michigan State University

PAUL LEUNG, Medical School, University of North Carolina at Chapel Hill; Committee on Disabilities and Handicaps, American Psychological Association
HENRY LEVIN, Departments of Education and Economics, Stanford University
PHILIP B. LYONS, U.S. Equal Employment Opportunity Commission
CHARLES F. NIELSON, Texas Instruments, Inc., Dallas, Texas
EDWARD E. POTTER, McGuiness & Williams, Washington, D.C.; Equal Employment Advisory Council
NAMBURY S. RAJU, Department of Psychology, Illinois Institute of Technology, Chicago
DENNIS K. RHOADES, National Economic Commission, American Legion, Washington, D.C.
WILLIAM L. ROBINSON, Lawyers' Committee for Civil Rights Under Law, Washington, D.C.
WILLIAM W. RUCH, Psychological Services, Inc., Glendale, California
ROBERT A. SCHAERFL, U.S. Employment Service, U.S. Department of Labor
FRANK L. SCHMIDT, College of Business Administration, University of Iowa
NEAL W. SCHMITT, Department of Psychology, Michigan State University
DONALD J. SCHWARTZ, Office of Research and Analytic Services, U.S. Equal Employment Opportunity Commission
JAMES C. SHARF, Career Entry Group, U.S. Office of Personnel Management
WILLIAM TRACEY, Employment Services, New Jersey Department of Labor, Trenton
DENNIS L. WARMKE, Philip Morris U.S.A., Richmond, Virginia
RAUL YZAGUIRRE, National Council of La Raza, Washington, D.C.