TABLE OF CONTENTS

AUTHORITIES CITED

Cases

Statutes and Rules of Court

Congressional Materials

Administrative Materials

Other Authorities (Legal)

Other Authorities (Non-Legal)

N. S. Petersen and M. R. Novick, "An Evaluation of Some Models for Culturo-Fair Selection," J. Educ. Measurement (in press) 11

R. L. Pothoff, Statistical Aspects of the Problem of Biases in Psychological Tests, University of North Carolina Institute of Statistics Mimeo Service No. 479 (1966) ... 11

F. Schmidt, et al., "Racial Differences in Validity of Employment Tests: Reality or Illusion?" 58 J. Applied Psych. 5 (1973) 11

Technical Recommendations for Achievement Tests (1955), reprinted in O.K. Buros, Tests in Print, 367-391 (1961) 9

Technical Recommendations for Psychological Tests and Diagnostic Techniques, Psych. Bull, Supp. 1954, 1 ... 9

R. L. Thorndike, "Concepts of Culture-Fairness," 8 J. Educ. Measurement 63 (1971) 11

R. L. Thorndike, Personnel Selection (1949) 41

U.S. Office of Strategic Services, Assessment Staff, Assessment of Men (1948) 10

IN THE

SUPREME COURT OF THE UNITED STATES

OCTOBER TERM, 1975

No. 74-1492

WALTER E. WASHINGTON, et al.,
Petitioners,

vs.

ALFRED E. DAVIS, et al.,
Respondents.

On Writ of Certiorari to the United States Court of Appeals for the District of Columbia Circuit

MOTION OF THE EXECUTIVE COMMITTEE OF DIVISION 14 OF THE AMERICAN PSYCHOLOGICAL ASSOCIATION FOR LEAVE TO FILE BRIEF AND FOR ORAL ARGUMENT AS AMICUS CURIAE

The Executive Committee of Division 14 of the American Psychological Association ("Executive Committee," "Division 14" and "APA", respectively) moves the Court under Rule 42 for leave to file a brief herein as *amicus curiae* and under Rule 44(7) for special leave of Court for oral argument.

The consent of the attorney for the petitioners herein has been obtained, but the attorney for the respondent herein refused to consent to the filing of a brief by the Executive Committee of Division 14 of the APA as *amicus curiae*. Neither petitioner nor respondent consents to oral argument on its side by the Executive Committee of Division 14 of the APA.

The Executive Committee of the Division of Industrial-Organizational Psychology[1] is the elected governing body of a professional association of approximately 1,400 members. The membership is believed to include the large majority of the practicing industrial psychologists in the United States. There is no other recognized regional or national organization of industrial psychologists. Division 14 is one of the specialized professional associations within the American Psychological Association, which is the major national professional association of psychologists in the United States. A critical area of interest to Division 14's membership is the use of psychological procedures in employee selection and utilization. Members of Division 14 who work in this facet of the profession are concerned with operational and ethical aspects of the procedures that are used to make employee selection, placement, and promotion decisions.

The stated objectives of the APA (and Division 14 as a component part) as set forth in its By-Laws include the desire to

---

[1] "Division 14" is the common term of reference for the Division of Industrial-Organizational Psychology of the American Psychological Association and will be used as such hereinafter. Its Executive Committee is elected annually by popular vote of all members and is hereinafter referred to as "the Executive Committee." The Executive Committee of Division 14 represents the members of the Division only. It does not speak for the entire Association. The Executive Committee and its members have no financial interest in the outcome of this litigation. Dr. Mary L. Tenopyr, whose affidavit is of record in this case, is an elected member of the Executive Committee. She has not participated in voting to submit this Brief or in its preparation. This Brief does not represent the views of any individual member of Division 14, including Dr. Tenopyr. Dr. Tenopyr is no longer employed by any party to this litigation.

advance psychology as a science, as a profession and as a means of promoting human welfare through the advancement of high standards of scientific and ethical conduct and through the dissemination of psychological knowledge to promote the public welfare.

While Division 14 shares many of the concerns raised by the American Society for Personnel Administration ("ASPA") in its *amicus curiae* brief filed in the instant case,[2] its interest lies chiefly in ensuring that legal principles governing validation efforts be consistent with professional practice and the current state of the art.

The accompanying brief is adverse to many of the positions of the immediate parties, one of whom has consented to its filing at this time. However, it is hoped that consideration of this case from the scientific perspective of the profession most directly involved will be of substantial assistance to the Court.

Division 14 recognizes that different principles may apply to public and private employers, but no position is taken regarding this question. The issues addressed herein by Division 14 transcend any such distinction and require guidance from this Court.

After consideration of various relevant principles, Division 14 seeks to apply them to the record in this case so as to show the errors which, it is respectfully submitted, were committed by both the court of appeals and the district court. It is felt that such errors are typical of those found in recent decisions of lower courts. However, such exposition is utilized solely for

---

[2] The Executive Committee of Division 14 is in substantial agreement with the *amicus curiae* brief herein of ASPA. Accordingly, the analysis contained therein will be repeated only to the extent necessary for proper presentation of the matters addressed in this brief.

illustrative purposes, and Division 14 takes no position on the ultimate merits of the case at bar.

In view of its conviction that both the court of appeals and the district court erred in this case, the Executive Committee of Division 14 of the APA believes that the views it seeks to present to this Court will contribute to a proper resolution of questions more important than the interests of the respective litigants.

Wherefore, the Executive Committee of Division 14 of the American Psychological Association respectfully requests this Court to permit it to file the brief *amicus curiae* which is submitted herewith and for special leave for oral argument.

Respectfully submitted,

R. LAWRENCE ASHE, JR.

SUSAN A. CAHOON

WILLIAM A. WRIGHT
3100 Equitable Building
100 Peachtree Street
Atlanta, Georgia 30303
(404) 522-3100
Attorneys for the Executive Committee of Division 14 of the American Psychological Association

Of Counsel

KILPATRICK, CODY, ROGERS,
McCLATCHEY & REGENSTEIN

IN THE

SUPREME COURT OF THE UNITED STATES

OCTOBER TERM, 1975

No. 74-1492

WALTER E. WASHINGTON, et al.,
Petitioners,

vs.

ALFRED E. DAVIS, et al.,
Respondents.

On Writ of Certiorari to the United States Court of Appeals for the District of Columbia Circuit

BRIEF

INTEREST OF THE AMICUS CURIAE

The interest of the Executive Committee of the Division of Industrial-Organizational Psychology (Division 14), American Psychological Association, as *amicus curiae* is set forth in the foregoing Motion.

SUMMARY OF ARGUMENT

In fair employment cases involving the job-relatedness of personnel selection systems, the lower courts lately have tended to regard published standards, designed to provide flexible and ideal guidelines to industrial psychologists charged with assessing whether systems are job-related, as a checklist against which validation studies offered as evidence in fair employment litigation should be matched on an "all-or-nothing" basis. Too often the courts have failed to appreciate that industrial psychology as a professional discipline is a young, evolving science possessed of few settled principles susceptible of simplistic and universal application. Rather, the validation studies performed by industrial psychologists essentially are applied research efforts which must invariably be tailored to the particular characteristics of the selection system, the job, and the employee or applicant population being analyzed.

Both the district court and appellate decisions in this case reflect misunderstandings of industrial psychology in their analysis of the limited evidence adduced in support of petitioners' and respondents' respective motions for summary judgment. In particular, the definition of substantial adverse impact has been left confused and uncertain, the legality of using success in training as a criterion has been questioned despite its recognized utility in the ordinary research strategies of our profession, and the necessity for a more complete record in order to draw reasonable inferences about the quality of a validation study has been ignored.

ARGUMENT

I. In Order Properly to Assess the Work of an Industrial Psychologist, a Court Must Recognize That Industrial Psychology Is Not a Discipline Possessing Standardized Formulae to Distinguish Acceptable From Inadequate Research Studies.

Following this Court's landmark pronouncement in *Griggs* v. *Duke Power Co.*, 401 U.S. 424 (1971), that henceforth selection procedures "measure the person for the job and not the person in the abstract,"[1] employers turned to industrial psychologists to develop new procedures which would satisfy Title VII's command and to analyze whether existing procedures were demonstrably job-related as required by *Griggs*. Thus, industrial psychology acquired a major new dimension in its long-standing professional objective of developing selection procedures which effectively predict which candidates are more likely to succeed on the job. For the first time the industrial psychologist has been required to consider not only whether his work satisfies his own standards of professionalism, but also whether a court of law would understand and concur in that assessment.

In the wake of *Griggs* the industrial psychologist has become a familiar forensic fixture in fair employment litigation, especially in cases involving paper-and-pencil tests used by employers as screening devices for hiring or promotional selections.[2]

---

[1] 401 U.S. at 436.

[2] *E.g., U.S. v. Georgia Power Co.*, 474 F.2d 906 (5th Cir. 1973); *Chance v. Board of Examiners*, 458 F.2d 1167 (2d Cir. 1972); *Allen v. City of Mobile*, 331 F.Supp. 1134 (S.D.Ala. 1971), *aff'd per curiam*, 466 F.2d 122 (5th Cir. 1972), *cert. denied*, 412 U.S. 909 (1973); *James v. Stockham Valves & Fittings Co.*, 394 F.Supp. 434 (N.D. Ala. 1975); *Arnold v. Ballard*, 390 F.Supp. 723 (N.D. Ohio 1975).

Understandably, the courts have sought to find standards to assist them in evaluating the conflicting conclusions of the litigants' respective expert witnesses, for the typical case involves a "battle of experts" who draw opposite conclusions from the same data. In order to place into perspective the propriety of current trends in finding such standards, this Court must first appreciate the evolution of certain written guidelines developed within the profession or by administrative agencies charged with implementing fair employment policies.

A. **Psychologists have only recently formulated written standards and do not intend them to be comprehensive minimum statements applicable in every research study.**

Psychologists have had a major role in the development and application of tests and other selection procedures in a wide variety of situations, including educational, military, and employment contexts. As early as 1890, James McKeen Cattell described tests he was using in an effort to demonstrate a relationship between test scores and performance in college. By 1901, this effort was far enough along to have produced data on correlations between mental ability tests and academic performance at Columbia University.[3] By 1911, Dr. Hugo Munsterberg of Harvard University was investigating the relationships of ability tests to performance on the job and to training of streetcar motormen and telephone operators.[4] As a consequence of this and subsequent research, Mr. Freyd was able by 1923 to offer an extensive review of basic requirements for competent test validation.[5] Subsequently, much research on employee selec-

---

[3] Hull, C.L., *Aptitude Testing.* World Book Company, 1928.

[4] Munsterberg, H., *Psychology and Industrial Efficiency.* Houghton, Mifflin, 1913.

[5] Freyd, M., "Measurement in Vocational Selection: An Outline of Research Procedure." 2 *Journal of Personnel Research* 215-49, 268-84, 377-85 (1923).

tion has been undertaken and an extensive literature (articles and books) on the subject has appeared.

In 1954, committees of the APA, the American Educational Research Association and the National Council on Measurement in Education collaborated to prepare the first professional summary intended to offer recommendations of appropriate research techniques and evaluation standards for the use of test developers and publishers.[6] Thereafter, a joint committee of members of the three associations consolidated, modified, and revised these publications in an attempt to present working guideposts and ideals to those involved in preparing and distributing tests. The culmination of this effort was the APA's *Standards for Educational and Psychological Tests and Manuals,* which was published in 1966 and then revised and republished in 1974 as *Standards for Educational and Psychological Tests.*[7]

Despite extensive effort, however, the technology of psychological testing and other employee selection procedures remains relatively embryonic. The evolution of the technology has been slow and somewhat erratic. There have been spurts of interest and effort and other periods of limited progress. The two World Wars provided both the incentive and the resources to bring about major advances in selection methodology. The armed services found the use of tests and other selection procedures of enormous value in screening recruits for the many kinds of "jobs" entailed in managing a modern military organization.

---

[6] Technical Recommendations for Psychological Tests and Diagnostic Techniques, *Psychological Bulletin,* Supp. 1954, 1-38, 51. Technical Recommendations for Achievement Tests (American Research Association, Washington, D.C. (1955), reprinted in Buros, O.K., *Tests in Print,* 367-91 (1961).

[7] American Psychological Association, *Standards for Educational and Psychological Tests and Manuals,* Washington, D.C., 1966; American Psychological Association, *Standards for Educational and Psychological Tests,* Washington, D.C., 1974 (the latter hereinafter "APA *Standards*").

World War II, especially, involved literally hundreds of psychologists in the development and validation of selection procedures. From these efforts emerged a number of reports of the results of truly massive research efforts.[8] In addition, problems in the methodology of selection were given serious consideration and a number of innovations were reported.[9]

With the emergence of concern for "fair employment" sparked by the enactment of Title VII of the Civil Rights Act of 1964 (42 U.S.C. § 2000e, *et seq.*, hereinafter "Title VII"), interest in testing and other employment procedures took on an added dimension.[10] Industrial and other psychologists became very much concerned with the potential impact of "test bias" on the employment of protected groups, especially of racial minorities.[11] Both empirical research and theorizing on the subject were stimulated. A considerable literature has emerged in the last ten years and will continue to do so, reflecting new knowledge about and added insights into the effects of using tests and other selection procedures on the employment of protected groups.[12]

---

[8] *See, e.g.,* Flannagan, J., ed., *The Aviation Psychology Program in the Army Air Forces*: *Report No. 1.,* Washington, D.C., U.S. Gov't. Printing Off., 1948.

[9] A major innovation during World War II was the development and application by the Office of Strategic Services of assessment centers, subsequently used extensively by industrial and other organizations for selecting managers and other key personnel. See OSS Assessment Staff, *Assessment of Men,* N.Y., Rinehart, 1948.

[10] Even before the passage of Title VII, the "Plans for Progress" program initiated by the Federal Government and the decision of a hearing examiner for the Illinois Fair Employment Practice Commission in *Myart v. Motorola Co.* [reprinted at 110 Cong. Rec. 5662-64 (1964)], began to attract professional interest in the "fairness" of standardized tests to racial minorities.

[11] *See, e.g.,* Guion, R. M. Employment Tests and Discriminatory Hiring, 5 *Indus. Rel.* 20-37 (1966).

[12] Examples of books, monographs and articles by psychologists pertaining to the subject (research findings, methodological approaches, theoretical formulations, legal considerations, and/or policy concerns) are: APA Task Force on Employment Testing of

Early research efforts seemed to indicate that the hypothesized phenomenon of "differential validity"[13] would often be encountered in comparing the effectiveness by racial subgroup of employment tests;[14] subsequent recent research casts serious doubt upon its existence.[15] Today, the debate continues among industrial psychologists about whether differential validity is a theory deserving further major research efforts.[16]

Minority Groups, Job Testing and the Disadvantaged, 24 *Am. Psych.* No. 7 (1969); Boehm, V.R., Negro-White Differences in Validity of Employment and Training Selection Procedures: Summary of Research Evidence, 56 *J. Appl. Psych.* 33-39 (1972); Byham, W. & Spitzer, M., *The Law and Personnel Testing,* 128-46 (1971); Campbell, J.T. *et al., An Investigation of Sources of Bias in Job Prediction,* Princeton, N.J., Educational Testing Service, 1972; Darlington, R.B., Another Look at "Cultural Fairness", 8 *J. Educ. Measurement,* 71-82 (1971); Einhorn & Bass, Discrimination in Employment Testing, 75 *Psych. Bull.* (1971); Kirkpatrick, J.J., Ewen, R.B., Barrett, R.S., & Katzell, R.A., *Testing and Fair Employment,* N.Y., New York University Press, 1968; O'Leary, B.S., Farr, J.L., & Bartlett, C.J., *Ethnic Group Membership as a Moderator of Job Performance,* American Institutes for Research, 1970; Peterson, N.S. & Novick, M.R., An Evaluation of Some Models for Culture Fair Selection. *J. Educ. Measurement,* in press; Pothoff, R.L., *Statistical Aspects of the Problem of Biases in Psychological Tests,* Instit. of Statistics Mimeo Service No. 479, Chapel Hill, N.C., U. of N. Carolina, 1966; Thorndike, R.L., Concepts of Culture-Fairness, *J. Ed. Measurement,* 1971, v. 8., 63-80.

[13] In its *Georgia Power* brief (pp. A-15 and A-16) the Division 14 Executive Committee states that the term "differential validity" is ". . . used to describe differences in criterion-related validity for identifiable subgroups of applicants. For example, differential validity would be said to exist if a test has different validities for blacks and whites, for men and women, etc. This may refer to different *validity coefficients,* or different *slopes,* or *intercepts* on *regression lines."* For a more recent statement concerning the current debate, see Division 14 *Principles,* "A Comment on Fairness," p. 2.

[14] *See, e.g.,* Kirkpatrick, *et al., op. cit.*

[15] *See, e.g.,* Campbell *et al., op. cit.*

[16] *See, e.g.,* F. Schmidt *et al.,* Racial Differences in Validity of Employment Tests: Reality or Illusion?, 58 *J. Applied Psych.* 5 (1973).

Because of the continuing evolution of the discipline, psychologists have avoided setting rigid rules and regulations with respect to research on and use of tests and other selection procedures but have sought to establish principles for the guidance of those concerned with their development and use. The purpose of placing those principles in writing is summarized succinctly in the APA *Standards:*[17]

> This document is prepared as a technical guide for those within the sponsoring professions; it is *not* written as law. What is intended is a set of standards to be used in part for self-evaluation by test developers and test users. An evaluation of their competence does not rest on the literal satisfaction of every relevant provision of this document. The individual standards are statements of ideals or goals, some having priority over others. Instead, an evaluation of competence depends on the degree to which the intent of this document has been satisfied by the test developer or user.

Because of its continuing concern for the issues involving guidelines on employee selection procedures, both existing and proposed,[18] Division 14's Executive Committee decided in August, 1974 to have developed a set of principles which members of the Division concerned with validating employee selection procedures could follow in conducting research. A committee made up of specialists on selection procedures was established. Included on the committee were persons who have participated in the drafting of other guidelines and persons who have represented a full range of perspectives in legal proceedings involving testing. In addition, all members of Division 14 were given an opportunity to comment on a draft of the proposed principles, and many members took advantage of the opportunity to do so. Publication of the *Principles for the Validation and Use of Per-*

---

[17] APA *Standards,* p. 8; emphasis in original.

[18] *See infra,* pp. 13-18, for a discussion of the development of governmental guidelines affecting the profession.

sonnel Selection Procedures was authorized by the Executive Committee in June, 1975.[19]

Principles is not intended as a legal document, but rather as a guide for professional psychologists who are responsible for research on the validation and use of selection procedures. Furthermore, *Principles* is not intended as "minimal" standards, but as *ideals* for professionals to guide their performance in the conduct of validation studies.[20] *Principles* is designed to be consistent with the APA *Standards*. In that regard, the principal author of the APA *Standards* was also the co-author of Division 14's *Principles*. To ensure consistency between the two documents, Division 14's *Principles* was reviewed with the APA Committee on Psychological Tests and Assessment, which had been responsible for developing the APA *Standards*.

B. **Governmental guidelines are also of recent vintage, rely heavily upon professional opinions, and are continually being questioned and revised.**

In the ten years since the effective date of Title VII, governmental guidance concerning the legal requirements by which the work of industrial psychologists may be scrutinized has undergone substantial transformation from the first set of "guidelines" issued by the Equal Employment Opportunity Commission (hereinafter "EEOC") in 1966[21] to the current effort to draft a common set of guidelines for use by all federal agencies charged with fair employment responsibilities.[22]

---

[19] Reproduced in Appendix, hereinafter "Principles."

[20] Division 14 respectfully submits that the APA *Standards* were similarly intended as ideal goals rather than legal *sine qua nons*. See APA *Standards*, pp. 6, 8.

[21] Equal Employment Opportunity Commission, *Guidelines on Employment Testing*, Aug. 24, 1966.

[22] Discussed *infra*, pp. 16-17.

The 1966 EEOC guidelines on selection procedures were relatively simple and largely a statement of the need to validate paper-and-pencil selection procedures. Shortly thereafter, the Office of Federal Contract Compliance (hereinafter "OFCC"), published a much more comprehensive set of guidelines on selection procedures (1968).[23] In 1970, the EEOC published a new set of guidelines, which was an expansion of their earlier, much simpler guidelines and of the OFCC guidelines.[24] The OFCC published new guidelines in 1971 which were substantially similar, though not identical, to the 1970 EEOC guidelines.[25]

The 1970 EEOC guidelines were much more comprehensive, stringent and prescriptive in their requirements for demonstrating that tests are job-related.[26] Though Division 14 as an organi-

---

[23] Sept. 9, 1968 Order by Willard Wirtz "Validation of Employment Tests by Contractors and Subcontractors Subject to the Provisions of Executive Order 11246", 33 Federal Register 14392, published September 24, 1968.

[24] Equal Employment Opportunity Commission, *Guidelines on Employee Selection Procedures* (revised), 35 Fed. Reg. 12333, 29 CFR § 1607, *et seq.*

[25] "Testing and Selecting Employees by Government Contractors," 41 C.F.R. § 60-3.

[26] Further, the definition of "test" was broadened quite substantially to include:

> For the purpose of the guidelines in this part, the term "test" is defined as any paper-and-pencil or performance measure used as a basis for any employment decision. The guidelines in this part apply, for example, to ability tests which are designed to measure eligibility for hire, transfer, promotion, membership, training, referral or retention. This definition includes, but is not restricted to, measures of general intelligence, mental ability and learning ability; specific intellectual abilities; mechanical, clerical and other aptitudes; dexterity and coordination; knowledge and proficiency; occupational and other interest; and attitudes, personality or temperament. The term "test" includes all formal, scored, quantified or standardized techniques of assessing job suitability including, in addition to the above, specific qualifying or disqualifying personal history or background requirements, specific educational or work history requirements, scored interviews, biographical information blanks, interviewers' rating scales, scored application forms, etc.

29 C.F.R. § 1607.2. *Cf.* 41 C.F.R. §§ 60-3.2, -3.13.

zation was not involved in their development, various members of Division 14 as individuals participated in differing ways in the development of these various sets of guidelines. Those guidelines, *inter alia*, sought to impress upon employers the need to analyze their selection procedures by means of techniques found professionally acceptable by industrial psychologists. Indeed, the 1970 guidelines expressly cite[27] the APA's 1966 *Standards for Educational and Psychological Tests and Manuals* as a source for ascertaining what methods are professionally appropriate.

Based on their professional involvement with the issues raised in the EEOC and the OFCC guidelines, the members of Division 14 are vitally concerned with legal interpretations of the guidelines.[28] These guidelines have necessarily been framed in flexible and non-specific language. This allows employers to utilize those test validation procedures that are best suited for the specific characteristics of their situations. The EEOC guidelines have been given strong deference in court cases in recent years,[29] yet a wide latitude of possible interpretations has evolved in disparate interpretations and findings in different courts.[30] Division 14 members have a professional

---

[27] 29 C.F.R. § 1607.5(a). Curiously, the EEOC *Guidelines* have not yet been amended to reflect the APA's publication of substantially revised *Standards* in 1974. *See* p. 9, *supra.*

[28] Pursuant to this concern, for example, the Division 14 Executive Committee submitted an *amicus curiae* brief at the request of the Court in *United States v. Georgia Power Co.*, 474 F.2d 906 (5th Cir. 1973), which received favorable treatment. *Id.*, at 914-915.

[29] *See, e.g., Griggs v. Duke Power Co., supra; Albemarle Paper Co. v. Moody*, 422 U.S. 405, 95 S.Ct. 2362 (1975); *United States v. Georgia Power Co.*, 474 F.2d 906 (5th Cir. 1973).

[30] *Compare Kirkland v. New York State Dep't. of Correctional Services*, 520 F.2d 420, 426 (2d Cir. 1975) (content validation method held inappropriate when criterion-related method is technically feasible) *and Douglas v. Hampton*, 512 F.2d 976, 986 (D.C. Cir. 1975) (construct validation method held appropriate only after a showing that it was infeasible to undertake proof of empirical cri-

interest in aiding the courts in their establishment of legal definitions of technical procedures in personnel selection so that the procedures determined to be legally correct are also technically sound and professionally practical. They have accordingly been involved in recent governmental efforts to devise new guidelines to implement that objective.

The passage of the Equal Employment Opportunity Act of 1972, *inter alia,* created the Equal Employment Opportunity Coordinating Counsel (hereinafter "EEOCC") in an attempt to resolve the inefficiencies and conflicts of overlapping federal EEO responsibilities.[31] The EEOCC commendably decided to develop a set of uniform guidelines on employee selection procedures which would be used by all federal government agencies having responsibility for fairness in employee selection. There have been seriatim drafts of the proposed Uniform Guidelines and at least two of these have been made available for comment by interested parties, including the membership of Division 14. The Professional Affairs Committee[32] of Division 14 was given a major responsibility for reviewing the first of these drafts and had a major role in stimulating a public hearing which was jointly chaired by representatives of the APA and the EEOCC. The Committee also undertook a survey of the views of Division 14 members on the first of these drafts and made the views of Division 14 members available to the staff of the EEOCC. The Executive Committee of Division 14 responded formally for the Division to the request for views on both of the drafts

---

terion-related validity) *with Vulcan Society v. Civil Service Comm'n.,* 490 F.2d 387 (2d Cir. 1973) (either method held acceptable) *and Coopersmith v. Roudebush,* 517 F.2d 818, 824 (D.C. Cir. 1975) (content validity acceptable where lack of feasibility of criterion-related validity not shown).

[31] 42 U.S.C. § 2000e-14.

[32] Appointed by the President of Division 14 with the concurrence of the Executive Committee.

and offered suggestions for modifying many of their provisions. The Committee also offered to assist the EEOCC staff in its endeavors. Although the technology of industrial psychology is still relatively undeveloped, Division 14 believes that the effort to utilize the best techniques available to develop good selection procedures is an important one, for as the EEOC's *Guidelines* (1970) "Statement of Purpose" noted:[33]

> [T]he guidelines in this part are based on the belief that properly validated and standardized employee selection procedures can significantly contribute to the implementation of nondiscriminatory personnel policies, as required by Title VII. It is also recognized that professionally developed tests, when used in conjunction with other tools of personnel assessment and complemented by sound programs of job design, may significantly aid in the development and maintenance of an efficient work force and, indeed, aid in the utilization and conservation of human resources generally.

In summary, though the history of the development and use of psychological tests and other selection procedures spans nearly one hundred years, the science of individual differences on which such procedures are based and the technology of measuring those individual differences remain relatively rudimentary. Many guides have been furnished industrial psychologists (texts, articles, APA *Standards,* Division 14 *Principles*) for developing, validating and using employee selection procedures. In addition, the EEOC and other federal and state government agencies have prepared and issued guidelines on employee selection procedures for use by employers. Through the EEOCC, the preparation of such guidelines continues. It is in the application of such professional standards and federal government guidelines, especially those issued by

---

[33] 29 C.F.R. § 1607.1(a).

the EEOC, that the interpretations given them by federal courts have become a vital concern to industrial psychologists.

II. Judicial Opinions Too Often Fail to Recognize the Need for Flexibility and Reasonableness in the Application of Professional Standards Developed for Industrial Psychologists.

Judge Friendly has sounded an apt warning to the bench in its efforts to apply the art of the industrial psychologist to legal determination of whether a test or selection procedure illegally discriminates on the basis of race:

> The Fourteenth Amendment no more enacted a particular theory of psychological testing than it did Mr. Herbert Spencer's Social Statics. Experience teaches that the preferred method of today may be the rejected one of tomorrow. What is required is simply that an examination must be "shown to bear a demonstrable relationship to successful performance of the job for which it was used." *Griggs v. Duke Power Co.*, 401 U.S. 424, 431 . . . (1971); *McDonnell Douglas Corp. v. Green*, 411 U.S. 792, 802 n. 14 . . . (1973).[34]

Unfortunately, this salutary caution frequently has been ignored in the rush of judicial efforts to develop litmus tests to distinguish "good" from "bad" validation studies of selection devices.

From the perspective of an industrial psychologist, each study designed to determine whether tests or other selection procedures are job-related is unique. Carrying out a validation study is not comparable to following a recipe; rather,

---

[34] *Vulcan Society v. Civil Service Comm'n.*, 490 F.2d 387, 394 (2nd Cir. 1973).

each study is an applied research investigation. The methodology for the study varies with the nature of the tasks involved, including the human and economic risks incurred by inadequate performance, the type and extent of the training required to perform the job successfully, the effects of any historic practices or affirmative action procedures on the employer's applicant population, and overall labor force characteristics, including available relevant skill, experience and educational levels. The resources available to the employer and the appropriate performance criteria also create differences among validation studies which may, on the surface, appear to be addressing similar problems. These variables and distinctions should properly be taken into account in applying guidelines pertaining to tests and other selection procedures.

From the courts' perspective in litigation, once substantial adverse impact is ascertained (*see, infra,* pp. 32-35), the legal question of whether the employer can show by a preponderance of the evidence that the selection procedures involved in the studies are demonstrably job-related and thereby justify their use must be answered.[35] An accurate answer requires that guidelines issued by federal government agencies and court interpretations of them reflect the current state of the art in validating selection procedures. Such guidelines and their interpretations should require neither a technology that does not yet exist nor such literal and unrealistic compliance that they create expense which is not "cost-effective" or bar procedures of practical utility. Division 14's *Principles* are intended to aid industrial psychologists who carry out such validation studies to decide if they have conducted the studies in a manner entitling them to professional respect while providing the kind of flexibility in standards necessary in conducting applied research of this kind.

---

[35] *Griggs v. Duke Power Co., supra* at 431.

Several recent court opinions regrettably have translated the "great deference" to which the EEOC *Guidelines* have been held to be entitled as interpretations by the administrative agency charged with enforcing Title VII[36] into a mistaken belief that the *Guidelines* when coupled with APA *Standards* form a handy checklist against which a validation approach can and should be measured. They have lost sight of the need to analyze validity studies on a case-by-case basis appropriate to the uniqueness of each study. In the process they have tended to impose an increasing number of professionally unrealistic and effectively unattainable requirements as the bare minimum acceptable to the courts in establishing whether tests and other selection procedures are demonstrably job-related. Indeed, some of these decisions require such stringent technical compliance as to convert principles drafted and intended as idealistic goals to be striven for, but rarely attained *in toto,* into successive hurdles which must be surmounted for survival in the crucible of fair employment litigation. The inevitable result is that some significant and worthy validation efforts are being declared legally unacceptable,[37] along with those which have patently "earned' judicial opprobrium.[38]

---

[36] *See, e.g., Albemarle Paper Company v. Moody,* 422 U.S. 405, 95 S.Ct. 2362, 2378 (1975), and cases cited therein.

[37] *Cf.* Robert M. Guion, Recent EEO Court Decisions, *The Industrial-Organizational Psychologist,* Vol. 11, No. 2. Dr. William Enneis, who has been the senior industrial psychologist employed by the EEOC since 1966, and who was the principal draftsperson of the EEOC's 1970 Guidelines, has testified in at least three fair employment cases (*Banks v. Lockheed-Georgia Co.,* C.A. No. 11,675, N.D.Ga.; *EEOC v. Monsanto Co.,* C.A. No. 73-33-CIV-P, N.D. Fla.; and *Stallworth v. Monsanto Co.,* C.A. No. 73-45-CIV-P, N.D. Fla.), that he is aware of only three or four criterion-related validation studies which meet all of the requirements of 1970 EEOC Guidelines. The only one of these three or four whose cost he knew involved a sample of c.200 bus drivers and had required an expenditure of c.$400,000.00. Testimony in the *Monsanto* cases is reproduced in relevant part in an appendix to this brief. *See* pp. A-27-A-35.

[38] *See, e.g., Griggs v. Duke Power Co., supra; Albemarle Paper Co. v. Moody, supra.*

An excellent example is the question of whether the criterion-related method for assessing validity is mandated by Title VII, the EEOC *Guidelines*, or the professional standards of industrial psychology, as the legally required validation strategy *whenever* it is technically feasible. The debate has apparently originated in the inartfully drafted language of Section 1607.5 (a) of the EEOC *Guidelines*, 29 C.F.R. §1607.5(a). Several appellate courts have interpreted this section of the *Guidelines* to direct employers (and, in turn, industrial psychologists) to use only a criterion-related study if it is technically feasible. *See Douglas v. Hampton*, 512 F.2d 976, 984-85 (D.C. Cir. 1975); *Rogers v. International Paper Co.*, 510 F.2d 1340, 1349 (8th Cir. 1975), *vacated and remanded for further consideration in light of Albemarle, supra; Kirkland v. New York State Dept. of Correctional Services*, 520 F.2d 420, 426 (2d Cir. 1975); *but see Spurlock v. United Airlines, Inc.*, 475 F.2d 216 (10th Cir. 1972) (college degree and 500 flight hours held content valid for commercial airline pilot training program); *Hodgson v. Greyhound Lines, Inc.*, 499 F.2d 859 (7th Cir. 1974), *cert. denied*, 419 U.S. 1122 (1975) (maximum hiring age of 35 held content valid for intercity passenger bus drivers); *Vulcan Society v. Civil Service Comm'n., supra* (content validity held equally acceptable); *Coopersmith v. Roudebush, supra* (content validity acceptable where lack of feasibility of criterion-related validity not shown).

The irony of this particular controversy is that the chief author of the EEOC *Guidelines*, Dr. William Enneis, has consistently stated in other cases that content validity is quite acceptable in appropriate instances, even though a criterion-related study is feasible.[39] However, Dr. Enneis' testimony in this regard, which is strongly endorsed herein, as well as reflected in the APA

---

[39] Cases cited *supra*, n. 37; relevant portions of this testimony from the *Monsanto* cases is reproduced in an appendix filed with this brief, pp. A-27-A-35.

Standards and Division 14 *Principles,* was not in the record of any of the foregoing cases which reached a contrary conclusion.

Another particularly noteworthy area of substantial variation appears in the definitions of "business necessity," which have ranged from the quite rational one of being synonymous with "demonstrably job-related," endorsed by this Court in its seminal decision of *Griggs v. Duke Power Company, supra* at 431-32, to the "plant closing" concept of *Robinson v. Lorillard Corp.,* 444 F.2d 791, 797 (4th Cir.), *cert. dismissed,* 404 U.S. 1006 (1971), and its progeny.[40] From the standpoint of our profession, we regard a selection test as "demonstrably job-related" and its use as a "business necessity" if validation research (whether by the content, criterion-related, construct or other professionally appropriate method) establishes a relationship between successful performance on the test and successful performance on an appropriate criterion of job success. Division 14 thus reaffirms herein the position previously taken in the Division's *amicus curiae* brief filed in *United States v. Georgia Power Company,* 474 F.2d 906 (5th Cir. 1973) (hereinafter "*Georgia Power* brief"):[41]

> It should be noted that the Supreme Court's definition of business necessity in the context of employee selection as 'related' to job needs is distinguishable from the use of this term in other contexts where lower courts have spoken in terms of the employer's burden to demonstrate that 'no

---

[40] *Rogers v. International Paper Co.,* 510 F.2d 1340, 1347 (8th Cir. 1975), *vacated and remanded for further consideration in light of Albemarle Paper Co. v. Moody, supra; United States v. Bethlehem Steel Corp.,* 446 F.2d 652, 662 (2d Cir. 1971); *Palmer v. General Mills, Inc.,* 513 F.2d 1040, 1044 (6th Cir. 1975). The misconception in *Robinson* regarding the allocation of the burden of proving the existence of a "less restrictive alternative" to the selection procedure actually used by the employer has been overruled by this Court's decision in *Albemarle Paper Co. v. Moody, supra.*

[41] *Georgia Power* brief, pp. 11-12 and n. 19. For the Fifth Circuit's treatment of this brief, see 474 F.2d at 914-915.

> acceptable alternative policies or practices' are available.[19] To be sure, if an employer can demonstrate that his current selection procedures assess whether his new hires are able to absorb training, thereby reducing the time and cost of training, or increase productivity or reduce turnover, he has demonstrated that employment procedures are truly related to achieving his operational objectives. It is precisely this reasonable approach to determining the needs of a business which must be the guidepost for applying the business necessity *concept* in the context of employee selection.
>
> ---
>
> [19] *E.g., United States v. Bethlehem Steel Corp.*, 446 F.2d 652, 662 (2d Cir. 1971); *Robinson v. P. Lorillard Corp.*, 444 F.2d 791, 798 (4th Cir. 1971); *Local 189, United Papermakers & Paperworkers v. United States,* 416 F.2d 980, 989 (5th Cir. 1969), *cert. denied,* 397 U.S. 919 (1970). . . .

To the extent that the law may impose a further burden on the employer if the plaintiff offers evidence of the existence of some alternative selection procedure which would choose people as efficiently but with less adverse impact,[42] the courts have left the arena of industrial psychology and entered a realm in which our standards have no direct bearing.

We have similarly been disturbed by periodic references in opinions to certain standardized tests in terms such as "race-oriented," which suggest a presumption on the part of the court that these instruments are invariably illegal and invalid as predictors of job performance.[43] Such a presumption cannot be accepted in our profession, where each combination of job, employer, applicant population and test requires an individual analysis. The same test which has no bearing to one job may be an excellent selection device for another. Overly broad ju-

---

[42] *See Albemarle Paper Co. v. Moody, supra.*

[43] *Franks v. Bowman Transp. Co.*, 495 F.2d 398, 412 (5th Cir. 1974).

dicial declarations condemning the use *per se* of a selection procedure which has been misused by some employers tend to have an *in terrorem* effect on others, thereby discouraging the use of or further research about selection methods which might well be appropriate in another context. To the extent that courts have made such remarks, it suggests again that the bench has failed to appreciate the need for a case-by-case application of governmental guidelines and professional standards appropriate to the nature of an industrial psychologist's work.

It is against such a background of troublesome decisions that we have elected to file a brief as *amicus curiae* in this case. As discussed hereinafter, both the district court and court of appeals in the case at bar indulged in departures from the type of reasoned and reasonable analysis we advocate in matters affecting industrial psychology. This Court has yet to be confronted with a validity study which satisfies most professional requirements but was rejected by the lower courts because of a limited, highly technical failure to comply with each and every published guideline and technical standard. In candor, because the record below is incomplete regarding many important aspects of the study upon which petitioner relies, we cannot determine whether that situation could have arisen in this litigation. We do know, however, that only this Court's direction will cause the lower courts to return to a reasonable, case-by-case analysis of validation studies, and we believe that this Court's determination of some of the issues which are specifically raised in the case at bar will have that effect, if they reflect the method of analysis advocated here. Accordingly, we have taken the opportunity afforded by the grant of certiorari in this case to make known to the Court our general concerns about developments in the law which affect industrial psychology as a profession, as well as to discuss the specific aspects of the decisions below with which we disagree.

III. The Decision of the Court of Appeals Should Be Reversed and the Case Remanded to the District Court for Trial on All Issues Because:

A. **The courts below lacked an adequate evidentiary basis and failed to articulate a principled legal standard for determining whether the Police Department's selection process has had a substantial adverse effect or impact[44] on black persons as a class.**

In *Griggs v. Duke Power Company*, 401 U.S. 424, 91 S.Ct. 849 (1971), this Court held illegal under Title VII "requirements [which] operate to disqualify Negroes at a substantially higher rate" than their white counterparts (401 U.S. at 426, 91 S.Ct. at 851) . . . "unless they are demonstrably a reasonable measure of job performance." (401 U.S. at 435, 91 S.Ct. at 856)[45] This holding was reaffirmed and elaborated somewhat in *Albemarle Paper Company v. Moody*, 422 U.S. 405, —, 95 S.Ct. 2362, 2375 (1975):

> In *Griggs v. Duke Power Co.*, 401 U.S. 424, 91 S.Ct. 849, 28 L.Ed.2d 158, this Court unanimously held that Title VII forbids the use of employment tests that are discriminatory in effect unless the employer meets "the burden of showing that any given requirement [has] . . . a manifest relation to the employment in question." Id., at 432, 91 S. Ct. at 854. This burden arises, of course, only after the complaining party or class has made out a prima facie case of discrimination—has shown that the tests in question

---

[44] The terms "adverse effect" and "adverse impact" have been used interchangeably and are so treated herein.

[45] *Also see Hester v. Southern Ry. Co.*, 497 F.2d 1374, 1379 (5th Cir. 1974): "The initial inquiry that a court must make in evaluating employment testing is whether the tests 'operate to disqualify Negroes at a substantially higher rate than white[s] . . .,'" citing *Griggs, supra*, 401 U.S. at 424.

> select applicants for hire or promotion in a racial pattern significantly different from that of the pool of applicants. *See McDonnell Douglas Corp. v. Green,* 411 U.S. 792, 802, 93 S.Ct. 1817, 1824, 36 L.Ed.2d 668 (footnotes omitted).

However. this Court has not yet clarified what quantum of adverse effect or impact is sufficient to put an employer to the burden of showing "a manifest relationship to the employment in question."[46]

Lower courts have to date been predominantly occupied with determining what differences *did* constitute the requisite "adverse effect" and not with ratiocination of those which did not.[47] Consequently, this essential guidance has thus far been provided by various federal and state agencies,[48] or not at all.

---

[46] *Griggs, supra,* 401 U.S. at 431; cited and quoted in *Albemarle, supra* 422 U.S. at —, 95 S.Ct. at 2375.

[47] The narrowest differential to date held by an appellate court to constitute the requisite substantial adverse impact was the 1-1/2:1 present in *Chance v. Board of Examiners,* 458 F.2d 1167, 1171 (2d Cir. 1972); *but compare Castro v. Beecher,* 459 F.2d 725, 729 (1st Cir. 1972) (2.6:1); *Green v. Missouri Pacific R.R. Co.,* —F.2d —, 14 EPD ¶ 10,314 (8th Cir. 1975); (5.3%:2.23% or c. 2-1/2:1 does constitute the requisite adverse effect) *with Taylor v. Safeway Stores, Inc.,* — F.2d —, 10 EPD ¶ 10,410 (10th Cir. 1975) (4.27% to 2.01% not sufficient adverse impact to prove a prima facie case); *Smith v. Troyan,* 520 F.2d 492 (6th Cir. 1975) (33% vs. 29% disparity of insufficient magnitude to justify further scrutiny). *Cf. Stone v. F.C.C.,* 466 F.2d 316, 332 (D.C. Cir. 1972) (application for renewal of TV license in District of Columbia area; 7% black employees in 24% black geographical area held to be "within the zone of reasonableness"). It should be noted that some courts have indicated that a narrower differential may suffice to prove a prima facie case against a public as opposed to a private employer. *NAACP v. Allen,* 493 F.2d 614, 621 (5th Cir. 1974).

[48] *See* discussion *supra,* pp. 13-18.

1. **The nature of the industrial psychologist's concern with the definition of substantial adverse impact.**

Industrial psychologists advocate the validation (i.e., proof of job-relatedness) of all tests and other selection procedures[49] prior to their application in an employment process.[50] The obvious reason is that the current state of the art does not permit valid inferences regarding the utility of a selection procedure without supporting evidence. However, considerations of time, effort and expense, especially when combined with technical problems such as sample size and restriction in range which may make validation research impossible, will usually preclude an employer's validation of its entire personnel selection system. Moreover, the use of seniority as an important or controlling component in many industrial job decisions is indelibly inscribed in both law and practice without the past or future prospect of validation.[51]

Because an employer is legally required to validate only those selection procedures which operate to disqualify a protected group at a substantially higher rate,[52] the employer may

---

[49] Mere information gathering procedures should be distinguished.

[50] APA *Standards,* p. 25: "A test developer must provide evidence of the reliability and validity of his test. . . . [M]any test users should do similar research on their own application of the test." *See also* Division 14 *Principles,* p. 3: "Ideally, . . . [p]sychologists should demonstrate the validity of as many aspects of the decision-making process as feasible. . . . When it is impossible or infeasible to apply validation methods to a given part of the decision-making process, that part should have a relationship, discernible by a reasonable person, to appropriate purposes of the employer."

[51] *See, e.g.,* 42 U.S.C. § 2000e-2(h) regarding the legality of a bona fide seniority system under Title VII. *See also Watkins v. Steelworkers, Loc. 2369,* 516 F.2d 41 (5th Cir. 1975); *Jersey Power & Light Co. v. Loc. 327 (IBEW),* 508 F.2d 687 (3d Cir. 1975); *Waters v. Wisconsin Steel Works,* 502 F.2d 1309 (7th Cir. 1974).

[52] *Albemarle Paper Co. v. Moody, supra,* 422 U.S. at —, 95 S.Ct. at 2375; *Griggs v. Duke Power Co., supra,* 401 U.S. at 426; *Hester v. Southern Ry., supra,* 497 F.2d at 1379.

pragmatically determine to limit its validation efforts and the costs they entail to those situations in which requisite adverse impact has been shown or can reasonably be expected. In establishing validation priorities for an employer, it therefore becomes crucial to have a reliable means of ascertaining which of the selection procedures, if any, are producing a degree of adverse impact sufficient to trigger the legal validation obligation. Although as research scientists industrial psychologists prefer to conduct validation studies of each component of an employer's system for hiring and promoting employees, it must be conceded that it is manifestly inappropriate for this profession to impose its ideals upon employers charged with meeting legal requirements while maintaining a profitable business. A clear definition of substantial adverse effect or impact will thus be invaluable to both industrial psychologists and employers in setting priorities for validation efforts.

Another highly pragmatic reason for concern that a definition of adverse impact be developed as soon as possible is the very limited supply of professional industrial psychologists qualified to design and carry out validation research. Division 14 has approximately 1400 members, many of whom are fully employed academics, career military or civil service personnel, or otherwise not available to public and private employers. A fair estimate is that there are between 200 and 300 persons qualified and available to satisfy what is quickly becoming a truly overwhelming and virtually infinite demand.[53] While this is indeed embarrassingly ironic and anomalous in view of the profession's historic efforts to convince employers of the importance to their organizations of validating selection

---

[53] The prediction of one commentator that "[T]itle VII may become a full-employment act for industrial psychologists" has been fulfilled. *See* Blumrosen, A., Strangers in Paradise: *Griggs v. Duke Power Co.* and the Concept of Employment Discrimination, 71 *Mich.L.Rev.* 59, 104 (1972).

procedures, it underscores the critical importance of concentrating the limited professional resources available on those situations where legal and social considerations are paramount.

2. **Deficiencies in the record below.**

Neither the district court[54] nor the court of appeals[55] made findings concerning the relevant labor force from which the District of Columbia Police Department draws its officers and the concomitant relevant labor force parity[56] percentage for black officers. Several alternatives apparently were available to the courts, ranging from the eligible 20-29 age group in the District of Columbia (c.67% black),[57] to the eligible 20-29 age group in the District of Columbia Standard Metropolitan Statistical Area ("SMSA") (c. 24.7% black, according to the 1970 Census), to the eligible 20-29 age group in the geographic area within a 50-mile radius (assumed by the District Court to be c. 44% black, 348 F.Supp. at 16).[58]

There is also little exploration in the record of the consequences of the Police Department's affirmative action efforts on the characteristics by race of the applicant flow to the Department in recent years.[59]

---

[54] *See* 348 F.Supp. at 16.

[55] *See* 512 F.2d at 960 n. 24.

[56] By parity, we mean that, absent some selection procedure which produces adverse impact or other factor, one would expect that the percentage by race of persons selected would mirror the percentage by race in the relevant labor force.

[57] *See* Appendix, p. 72.

[58] This 44% black proportion of 20-29 age group within a 50-mile radius is apparently unsupported in the record and would appear from census data on the Washington D.C. SMSA to be erroneously large.

[59] The district court merely commended these efforts and concluded that they tended to negate the consequences of Test 21's ap-

From the industrial psychologist's perspective, it is quite important, if feasible, to determine if the group studied is representative of the applicants to whom a test will be given. The APA *Standards* regards as "essential" that:

> [a]ny selective factor determining the composition of the validation sample should be indicated in a manual or research report. The sample should be described in terms of those variables known as thought to affect validity, such as age, sex, socioeconomic status, ethnic origin, residential region, level of education, or other demographic or psychological characteristics.[60]

If there are differences among the applicants, the validation study should attempt to analyze the validity for subjects "who are of same age or in the same educational or vocational situation as the person for whom the test is recommended."[61]

By analogy, a similar situation may arise as a result of affirmative action efforts which produce an atypical black applicant flow, so that the bottom line impact on employment opportunities for blacks may be over or understated by the comparison of black applicant percentages and black hiring percentages. For example, affirmative action recruitment may have generated a

---

parent impact, 348 F.Supp. at 16, without analyzing the pertinent factors discussed hereinabove, while the court of appeals dismissed the efforts as "irrelevant," 512 F.2d at 960-61. The court of appeals relied for its conclusion upon a series of cases which uniformly deal with the question of whether good faith efforts are a defense. Cases cited 512 F.2d at 961 n.31. The court of appeals thereby confused *efforts* with *results*. None of the cited cases indicates that the effects of an employer's affirmative recruiting efforts upon a normal applicant flow cannot and should not be considered. Certainly, as described in text above, from the perspective of the industrial psychologist the representativeness or atypicality of the sample represented by the employer's applicant flow should be investigated.

[60] APA *Standards,* ¶ E6.1, p. 36; *see also* Division 14 *Principles,* No. 3, p. 6.

[61] APA *Standards,* ¶ 6.1.1, p. 37.

flow of black applicants whose average qualifications exceed those of white applicants in formal educational attainment. In light of the evidence in the record to date that verbal skills typically acquired in higher education are important to police officers,[62] it would not be unreasonable to infer that the actual impact on blacks in this case is greater than indicated, for as a more qualified group, their selection ratio should have been greater than the ratio among white applicants. On the other hand, white applicants may have tended to possess the equivalent of a junior college level of formal education, whereas affirmative recruitment may have resulted in a large group of black applicants who possess only the bare minimum high school education required (and that from an inferior segregated school). Comparative rejection rates between such groups would in that instance overstate the impact of Test 21.[63]

The need for analyzing factors which affect the composition of an employer's applicant flow have heretofore been recognized not only in the published professional standards of industrial psychologists, but also in the administrative guidelines furnished to employers to assist them in complying with fair employment laws and rules. Thus, the guidelines issued by the Fair Employment Practices Commission of the State of California take special note of the fact that the characteristics of the applicant

---

[62] *See, e.g.,* 348 F.Supp. at 17.

[63] With a relatively large and statistically adequate sample size and a black failure rate of 57% versus a white failure rate of 13%, simply to note as did the district court, that Test 21 produced a "relatively higher percentage of black test failures" (348 F.Supp. at 17) is, at best, an understatement. Assuming the accuracy of these figures, Test 21 clearly produced a substantial adverse impact upon the black sample to which it was administered. (The Court also held that Test 21 *per se* plays a "small" role in the Department's failure to hire more blacks. *Id.,* at 16-17.) Not adequately answered are the extent to which—if any—the Department's affirmative recruiting practices produced an atypical black sample, fewer of whom were qualified for police officer training than their white peers, and the degree to which any such black sample abnormality accounts for the substantially adverse test performance of the blacks tested.

population, especially when influenced by special recruiting efforts on the part of the employer, affect the inferences to be drawn from the raw statistics.[64] Similarly, the 1970 EEOC *Guidelines* caution that in a proper validation study the sample of subjects "must be representative of the normal or typical candidate group," assuming that the applicant sample is also "representative of the minority population available for the job or jobs in the local labor market."[65]

Absent findings concerning the effects of the undisputed affirmative recruiting efforts of the Police Department, we believe that both courts below were indulging in speculation when they attempted to reach a conclusion about the presence or absence of substantial adverse impact in this case.

3. **A suggested definition of substantial adverse impact.**

In the vacuum created by the courts' failure to date to provide a principled standard for ascertaining when impact is "substantial," administrative agencies have attempted the task. The State of California's long-established and well-respected Fair Employment Practices Commission ("FEPC")[66] in 1971 appointed a Technical Advisory Committee on Testing ("TACT") whose mandate was "to formalize and update their [the California FEPC's] Guidelines on employee selection procedures."[67] TACT recognized "something of a dilemma . . . [in that] [m]ost California employers were already expected to conform to two sets of very similar Federal standards[68] so that the addition

---

[64] California Fair Employment Practices Commission, Technical Advisory Committee on Testing, October, 1972; published in BNA, *Fair Employment Practices Manual,* p. 451-151.

[65] 29 C.F.R. § 1607.5(b)(1).

[66] Appointment pursuant to § 1414 of the California Fair Employment Practice Act.

[67] California FEPC Guidelines, "Introduction," BNA, *Fair Employment Practices Manual,* 451-145.

[68] EEOC Guidelines and OFCC Guidelines.

of a third set seemed somewhat burdensome . . . The solution . . . adopted is to add sufficient interpretive material to the Federal standards so that their meaning as they will be applied in California is clear."[69] TACT's efforts were adopted by the California FEPC and published in October, 1972.[70]

Among the provisions of the EEOC Guidelines and OFCC Guidelines for which interpretive material was provided was the term "discrimination" as used in § 1607.3 of the EEOC Guidelines and § 60-3.3 of the OFCC Guidelines. "Discrimination" in those contexts was synonymous with the "substantial adverse effect" of *Griggs v. Duke Power Company,* 401 U.S. at 426, 91 S. Ct. at 851. The definition of "discrimination" or "substantial adverse effect" adopted was a combination of the "bottom line concept"[71] and the "eighty percent rule."[72]

---

[69] BNA, *op.cit.,* 451-145. The "three basic reasons for the approach . . . taken" were: "1. The Federal Guidelines are basically acceptable standards, and there seemed to be little justification to add a third, unique set to the two which are already available. 2. This approach allows us to perform what we hope is a service to California employers in further explaining the intent of the Guidelines and discussing some of the interpretive issues that they pose. We cannot, of course, guarantee that the Federal agencies will agree with our interpretation in every case, but we will make every effort to minimize conflicting interpretations. 3. We felt a need to expand the Guidelines beyond the Federal wording because we have found that many testing professionals do not understand some of the implications of many of the sections. We have tried to provide a reasonable treatment which reflects current professional thinking." *Id.*

[70] BNA, *op.cit.,* 415-145 *et seq.*

[71] The "bottom line concept" refers to the total selection process for a job. If the total selection process has no substantial adverse impact, then the individual components (tests, interviews, etc.) need not be examined for substantial adverse impact. If the total selection process results in substantial adverse impact, then individual components of the process should be evaluated for substantial adverse impact. The "bottom line concept" thus bears directly on whether a protected group is getting its proportionate or "fair" share of employment opportunities for the job in question.

[72] The "eighty percent rule" is set forth in the first definitional paragraph below. The "eighty percent rule" definition of adverse

In its entirety it reads as follows:[73]

> Adverse effect refers to a total employment process which results in a significantly higher percentage of a protected group in the candidate population being rejected for employment, placement, or promotion. The difference between the rejection rates for a protected group and the remaining group must be statistically significant at the .05 level. In addition, if the acceptance rate of the protected group is greater than or equal to 80% of the acceptance rate of the remaining group, then adverse effect is said to be not present by definition. Groups which are defined by their race, religious creed, color, national origin, ancestry, or sex are the protected groups.
>
> If the total selection process has an adverse effect, then the components of the process, including tests, must be examined for adverse effect. Any component which has an adverse effect must be validated or the effect must be eliminated.
>
> Raw rejection rates are not necessarily sufficient to evaluate adverse effect. On many occasions the raw rejection rates must be compared with an expectation based on the characteristics of an employer's applicant population and any special recruiting efforts which might affect his applicant population. For example, when an employer is aggressively recruiting minority group members from

---

effect has also received the official sanction of the OFCC. OFCC Testing and Selection Order Guidance Memorandum No. 8, July 24, 1974, ¶ 3 (published in CCH *Employment Practices,* ¶ 4355); *also see* the OFCC's *Question-and-Answer Booklet on the Testing and Selection Order,* referenced in ¶ 3, *id.* In underscored language, the OFCC states in Memorandum No. 8, that, "when there is no adverse effect, there is no validation requirement." A modified form of the California FEPC's "eighty percent rule" also appears in the most recent draft of the EEOC's proposed uniform guidelines. *Cf.* 16-17, *supra.*

[73] BNA, *op.cit.,* 451-145.

among the "hard core" unemployed who have lower levels of education and experience than the general population, disproportionate rejection rates might not be judged evidence of adverse effect.

IF A TOTAL EMPLOYEE SELECTION PROCESS DOES NOT RESULT IN AN ADVERSE EFFECT, NO VALIDATION INFORMATION IS REQUIRED. (Emphasis in original.)

If this definition of "substantial adverse effect" or "discrimination" is viewed favorably by this Court, then the third paragraph (regarding use of "raw rejection rates") will obviously require consideration upon remand.[74] For reasons previously discussed,[75] such analysis of the applicant population is a professionally appropriate consideration for the industrial psychologist.[76]

The Executive Committee of Division 14 as *amicus curiae* respectfully submits that the foregoing definition of "substantial adverse impact" or "discrimination" within the meaning of fair employment laws is an appropriately stringent and eminently workable one worthy of this Court's endorsement. Above all, however, we urge the Court to develop a principled rule for defining "substantial adverse impact" so that employers and industrial psychologists may direct their energies and limited resources to analysis and validation of selection procedures which raise the most critical legal and social concerns.

---

[74] By so suggesting, no expression of opinion is intended as to the outcome based upon a complete, well-developed record in this case.

[75] *Supra,* pp. 30-32.

[76] This Court may well also desire to consider the public policy implications to ultimate fair employment goals of determining "substantial adverse impact" on the basis of raw rejection rates for employers who are making good faith and effective affirmative action recruitment efforts.

B. **The record below was inadequate for either the district court or the court of appeals to determine on summary judgment whether the Futransky study demonstrates that Test 21 is job-related.**

This Court has often cautioned that, "Rule 56 should be cautiously invoked to the end that parties may always be afforded a trial where there is a bona fide dispute of facts between them."[77] Justice, then Judge, Minton's conclusions about the proper role of Rule 56 have been affirmed by this Court:[78]

> The procedure for summary judgment was intended to expedite the settlement of litigation where it affirmatively appears upon the record that in the last analysis there is only a question of law as to whether the party should have judgment in accordance with the motion. . . . If there was any question of fact presented on the record in the proceedings for summary judgment, the motion could not be sustained.

The warning to avoid "trial by affidavit" has been deemed especially appropriate in complex litigation.[79] It is equally well-settled that the presence of cross-motions for summary judgment does not alter the standards to be applied to either party's motion.[80] When the record is inadequate in cases involving important issues, motions for summary judgment should be denied.[81]

We must part company with the conclusions regarding the Futransky study reached by the courts below because each failed

---

[77] *Associated Press v. United States,* 326 U.S. 1, 6 (1945).

[78] *Elgin, J. & E. Ry. v. Burley,* 140 F.2d 488, 490 (7th Cir. 1944), *aff'd,* 325 U.S. 711 (1945).

[79] *Poller v. Columbia Broadcasting System,* 368 U.S. 464, 472-473 (1962).

[80] *See* 6 Moore, *Federal Practice,* ¶ 56.13, cases cited p. 2247, n. 3.

[81] *Arenas v. United States,* 322 U.S. 419, n. 7 (1944).

to recognize these important legal principles, in light of the unresolved questions of industrial psychology which should be confronted before determining whether the study shows Test 21 to be job-related.

A crucial issue is whether, for the use to which the police department puts Test 21, it is legally permissible to validate it against a criterion of training performance, rather than of subsequent performance as a policeman. The district court summarily concluded that training success was an acceptable criterion in this case and that the Futransky study demonstrated a statistically significant relationship between Test 21 and such a criterion. In the court of appeals, on the same record, the court concluded that training success was an inappropriate criterion. In so ruling, the court of appeals went beyond the minimal issues posed by the record before it to express strong doubts that training success is ever a legally permissible criterion against which to endeavor to validate a test by means of a criterion-related study.[82]

1. **An industrial psychologist's standards for selecting a criterion and criterion measures.**

An industrial psychologist's first task in conducting a criterion-related validation study is to determine the purpose(s) for which

---

[82] *See, e.g.*, 512 F.2d at 958, 959, 961, 962-63, 964. It is unclear to what, if any, extent the court of appeals' skepticism about training performance as a criterion was influenced by its repeated stress that the employer has a "heavy" burden to demonstrate job-relatedness once substantial adverse effect is ascertained. 512 F.2d at 958, 959, 961, 965. This "heavy" burden imposition is attributed to this Court's seminal decision in *Griggs, supra.* 512 F.2d at 965. A careful reading of this Court's decisions in both *Griggs, supra,* and *Albermarle, supra,* indicates no more (or less) than the requirement of proving job-relatedness by the usual preponderance of the evidence standards, once plaintiffs have shown a prima facie case through demonstration of substantial adverse impact. Thus, the use of the term "heavy" in this context appears both erroneous and substantially unclear in its import.

the employer uses the selection procedure. Division 14's *Principles* explicitly notes the necessity for this determination:[83]

> Before any assessment procedure is considered or any validation effort planned, one should have a clear idea of what the assessment or validation is for. Any such statement of purpose logically must come from an understanding of the needs of the employing organization and of its present and prospective employees. As a general matter, a psychologist should develop clear objectives for an assessment procedure and design the validation effort to determine how well they have been achieved; these objectives should be consistent with professional, ethical, and legal responsibilities.

The need to establish purpose is further documented in the Division 14 *Principles* in the consideration of criteria for evaluating a selection procedure. The *Principles* states:[84]

> All criteria should be clearly related to the psychologist's purposes. Criteria should be chosen, not on the basis of availability, but on the basis of importance and relevance. This implies that the purposes (a) are clear, (b) are acceptable in the social and legal context in which the employing organization functions, and (c) are appropriate to the employing organization's purposes.
>
> All criteria should represent important work behaviors or work output, on the job or in training, as indicated by an appropriate review of information about the job. Criteria need not be all-inclusive, but there should be clear documentation of the reasoning determining what is, and what is not, included in a criterion. Criteria need not reflect actual job performance. Depending on the purpose of the

---

[83] Division 14 *Principles,* pp. 2-3.

[84] Division 14 *Principles,* Nos. 1 and 2, p. 3.

test user, various criteria, such as overall proficiency, training time, sales records, number of prospects called, and turnover may be used.

If the employer's primary purpose is to predict training perfomance, then the use of training performance as a criterion for evaluating the selection procedure is professionally appropriate. Training performance has long been recognized as an appropriate criterion for evaluating selection procedures.[85] The reasons for this are multiple and include:[86] (1) the cost of training to the organization, (2) the cost of possible failure in training to an individual, (3) the relevance of knowledge and skills obtained in training to performance on the job, (4) the fact that jobs require updating and retraining as a part of continuing job performance,[87] and (5) the human and/or economic costs of allowing poorly trained persons onto the job.[88] Failure in training is frequently very costly to the individual taking the training, as well as to the employer. This is particularly true where the training is lengthy and, therefore, the individual invests considerable time and effort, only to incur the frustration and career planning damage of a failure experience and the irreplaceable loss of the individual's personal time investment. Thus, from the standpoint of both employer and employee, it is highly desirable that people who enter the training be likely to succeed in it. This conclusion necessarily is predicated on training which has been developed from an analysis of job require-

---

[85] *See, e.g.,* Munsterberg, *op.cit.,* pp. 98-100.

[86] This list is by no means intended to be exhaustive.

[87] For jobs which require updating and retraining in order either to improve or to maintain job performance, it is relevant to use initial training performance as a criterion, because an applicant's ability to succeed in subsequent training experience is important.

[88] The economic costs may include not only the performance loss, but also the sometimes exponentially greater costs resulting from reduced productivity and poorer quality.

ments so that it represents skills and knowledge required by individuals to perform the job.[89]

Industrial psychologists who specialize in developing training programs usually rely on content validation procedures to demonstrate the job-relatedness of training. The first step in the process is to make a detailed task analysis (a form of job analysis) of the job in order to identify what employees do. From this information, which may be voluminous, the training program is developed. Measures of training performance are then developed which reflect the extent to which the knowledge and skills required on the job are being acquired by the trainee. Thus, the content of the training and performance in it directly reflect job requirements. Employees meeting training requirements are thereby equipped to perform the job to which they will subsequently be assigned.[90]

The measures used for evaluating training performance should reflect the purposes of the training. Where the purposes involve knowledge which is important to performance on the job, measures of job knowledge are pertinent. Such measures may very well include paper-and-pencil tests or other means for ascertaining whether the individual has assimilated the information presented in the training. Where skills are objectives in the training, then the appropriate measures to evaluate how well the person has mastered the skills should be skill measures of some sort. All such measures should be objective (to the extent reasonably practicable), relevant and reliable.

Precedents for using training performance as criteria for evaluating selection procedures abound. Probably the best

---

[89] The nature and mix of skills acquired on the job, if any, will vary from job-to-job.

[90] *See, e.g.,* Goldstein, Irwin I., *Training: Program Development and Evaluation,* Monterey, Cal., Brooks/Cole Publ. Co., 1974. It should be noted that methods for training employees have undergone major changes in the last 20 years.

known example is the use of training performance by the U. S. Army Air Force in World War II.[91] It was essential during the war to select people who could function effectively as pilots, navigators, bombardiers, and so forth. For each of these jobs, the psychologists responsible for conducting the research carefully developed measures of training performance, which were designed to duplicate as closely as possible the requirements of the job. The extensive research by Air Force psychologists on selection procedures using training performance criteria represent landmarks in selection research. The use of training performance as criteria for validating selection procedures is not, of course, restricted to World War II or the military. Many business firms and other organizations have used similar criteria for validating their selection procedures.[92] It is recognized that a major purpose for the use of tests by employing organizations is to select people who have the abilities to learn to perform the jobs for which they are employed. This is not to assert that other criteria of performance are not useful, but rather that in many situations training performance is an appropriate criterion for evaluating selection procedures.[93]

---

[91] Flannagan, J. (Ed.) *The Aviation Psychology Program in the Army Air Forces; Report No. 1,* Washington, D.C. U.S. Gov't. Printing Off., 1948 (esp. pp. 115-138).

[92] Many authorities cite training performance as appropriate criteria for validating tests and other selection procedures. *See, e.g.,* Dunnette, M.D. *Personal Selection and Placement,* 8-9, Belmont, Cal., Wadworth Printing Co., 1966; Ghiselli, E.E. *The Validity of Occupational Aptitude Tests,* 25, N.Y., Wiley, 1966; Guion, R.M., *Personnel Testing,* 93, N.Y., McGraw-Hill, 1965; Osburn, H.G. & Manise, W.R., *How to Install and Validate Employee Selection Techniques,* 33-35, Washington, D.C., American Petroleum Institute, 1971; Thorndike, *op. cit.,* p. 122 and pp. 136-137.

[93] Examples include craft-type jobs involving an apprenticeship program, virtually any training which is extensive in length, or which is costly either to give or to undertake, or in which the general applicant success rate is not high. A mere familiarization process would normally not be included.

Even though the industrial psychologist may find that training performance is an appropriate criterion and that it is positively correlated with the selection procedure, it is not unusual to be unable to find a statistical, correlation relationship of training performance to subsequent job performance, even though the training program is appropriate in light of job analysis and proper measures of training performance have been devised. Presumably, persons employed for a particular job must first have been trained for it and have demonstrated minimal competence in the training prior to being assigned to the job. Having obtained the necessary knowledge and skills to perform the job, subsequent job performance is likely to reflect other characteristics (e.g., motivation) which are not ordinarily incorporated in the training nor necessarily reflected in training performance. Furthermore, job performance reflects the impact of various factors in the work environment, such as the nature of assignments and relationships with supervisors, which vary for individual employees and may be beyond the control of the employee to influence.

The methods for measuring job performance also play a critical role in determining whether an empirical relationship between measures of training performance and measures of job performance is meaningful. Supervisory ratings, attendance records, disciplinary actions, etc., are measures of job performance which do not necessarily reflect aspects of performance also significant in training performance.

Industrial psychologists have recognized the relative independence of training performance and subsequent job performance. Along with it they have recognized that predicting training performance with ability tests may result in higher relationship than those obtained from predicting job performance with the same tests. Ghiselli, for example, states:[94]

---

[94] Ghiselli, *op. cit.*, p. 117.

> The generally higher predictiveness of training criteria as compared with proficiency criteria might well be anticipated. Since training programs are conducted under more controlled circumstances, measures of success in them would be expected to be more reliable and hence more predictable. It also seems reasonable to expect lower validity coefficients for proficiency criteria since there is quite possibly a smaller range of talent among workers on a job than among trainees. People low in ability are likely to fail during training and hence never appear among those who are working on the job itself.

In summary, training performance is recognized by a broad consensus of industrial psychologists as a legitimate criterion for validating tests and other selection procedures.[95] It is essential, however, that the content of the training reflect major job requirements and that measures used to evaluate performance in training do likewise. It is essential, also, that such measures be objective and reliable.

2. **Deficiencies in the record which preclude judgment about the merits of the Futransky study.**

In the current case, the record developed on the cross motions for summary judgment is too sketchy to determine whether, under the principles discussed above, performance in the Police Academy (i.e., training performance) is an appropriate criterion and, if it is, whether the Futransky study supports the conclu-

---

[95] Neither the Division 14 *Principles* nor the EEOC Guidelines discuss training performance or other kinds of criteria at any length. The *Principles* state, "Depending on the purpose of the test user, various criteria, such as overall proficiency, training time, sales records, number of prospects called and turnover may be used." *Id.*, p. 3. The EEOC *Guidelines* have a similar statement: "Such criteria may include measures other than actual work proficiency, such as training time, supervisory ratings, regularity of attendance and tenure." 29 C.F.R. § 1607.5(b)(3).

sion that Test 21 is related to successful performance in training.

At the outset it is unclear whether the employer was concerned with using a test to predict training performance, job performance, or both. The validation study cited by the employer[96] was carried out with measures of both as criteria. An evaluation of this study must take into account the intent of the employer in using the test. Because of the ambiguity of the employer's purpose as reflected in the record to date, it is not possible to discern the import of the Futransky study, for it is unclear whether the crucial concern is determining if a relationship exists between Test 21 and training performance or determining if there is a relationship between Test 21 and performance as a police officer. Without a clear understanding of purpose, the district court should not have ruled on summary judgment that Test 21 is demonstrably job-related, nor should the court of appeals so lightly have disregarded the evidence of a relationship between Test 21 and performance in the Police Academy.

Assuming for purposes of analysis that there was sufficient evidence that Test 21 is used primarily to assure that persons selected for the Police Academy are able to comprehend its curriculum, so that a training performance criterion would be appropriate,[97] summary judgment for either party is nevertheless

---

[96] Futransky, D.L. *Relations of D.C. Police Entrance Test Scores to Record School Performance and Job Performance of White and Negro Policemen.* Bur. of Policies and Standards, U.S. Civil Service Commission, Nov. 1967. In "Purpose of Study" (Appendix, p. 53), the author states, "This report is concerned with evaluating the effectiveness of a written test as part of the selection process."

[97] A seventeen-week program would generally be regarded by industrial psychologists as "training" rather than mere job familiarization, and thus appropriate as an independent source of criterion data. The district court's conclusion that training performance would be a proper criterion, however, rested in part on the entirely unsubstantiated view that, "In any event, so many factors affect a policeman's performance on the job it is doubtful that a written test could

inappropriate because the record is inadequate concerning whether the Police Academy's curriculum is job-related and is silent concerning whether the training performance measures selected for study are appropriate. The record does include a detailed description of the training program, but it fails to indicate on what basis the program was developed, i.e., no record of a job analysis or other rationale, though the detailed description of the curriculum[98] would suggest that such an analysis had been made. However, until further testimony or documentary information is furnished, there is not a professionally-acceptable basis for concluding that the curriculum in the Academy is job-related. Furthermore, there is very little descriptive material in the record on the examinations given in the Academy which are the criterion measures.[99] Consequently, no evaluation of their relevance, objectivity or reliability can be made at this time.

Similarly, there is little information in the record on the measures of job performance used.[100] Any attempt, there-

---

ever be devised that would prophesy performance accurately in advance." 348 F.Supp. at 17. In contrast, *see, e.g., Arnold v. Ballard,* 390 F. Supp. 723, 732, 737 (N.D. Ohio 1975), for a well-reasoned decision to the contrary based upon a complete record developed at a full evidentiary hearing. The case involved the Akron, Ohio Police Department and was not appealed by the plaintiffs; the defendants' appeal of the imposition of quota relief to the Sixth Circuit is pending. The opinion in *Arnold v. Ballard* is a commendable example of an appropriate effort to apply both professional standards and administrative guidelines, and illustrates the type of analysis which should be undertaken in this litigation after remand, upon a fully developed record.

[98] See "Recruit Training Curriculum," pp. 110-177 of Appendix.

[99] Futransky, *op. cit.* (p. 54a of Appendix), states only that, "Eight subject matter areas are covered in the training. Each of the subjects are tested and 70% right is required in each subject. The Recruit School Final Average is simply the average of the eight percentages on the tests."

[100] Futransky, *op. cit.* (p. 54a of Appendix): "The performance rating represents the last rating that appears in the policeman's official folder. It represents a summary rating over some nine individual rating items."

fore, to evaluate these measures or to interpret the relative lack of relationships between these measures and the measure of training performance or scores on Test 21 is presumptuous and premature.[101]

A further question whose answer must await the development of a complete record is whether the passing score of 40 on Test 21 is appropriate. The court of appeals correctly indicated concern about this cutting score[102] whereas the district court erroneously failed to consider the issue at all. On the other hand, the court of appeals erred in condemning the cutting score without having before it any evidence to indicate whether there was a rationale for it and, if so, whether the rationale is appropriate.[103]

---

[101] The *Principles* (p. 11) state, "Reports of validation research should include enough detail to enable a competent fellow psychologist to know precisely what was done and to draw independent conclusions in evaluating the work." The APA *Standards* (p. 33, E3) state, "All measures of criteria should be described completely and accurately. The manual or research report should comment on the adequacy of a criterion. Whenever feasible, it should draw attention to significant aspects of performance that the criterion measure does not reflect and to irrelevant factors likely to affect it." The EEOC *Guidelines* state, "The work behaviors or other criteria of employee adequacy which the test is intended to predict or identify must be fully described." 29 C.F.R. § 1607.5(b)(3). The ideal validation study from a professional standpoint is thus a detailed document, rather like an article prepared for a professional journal. For purposes of determining if a selection procedure is legal, however, the courts should be concerned with substance rather than form. Indeed, cost and time constraints often cause psychologists working in an industrial setting to prepare summary reports which can only be evaluated in light of supporting, supplementary information available in assorted business records or the personal knowledge of the psychologist and employer personnel.

[102] 512 F.2d at 963-64; *see also* Division 14 *Principles,* No. 11, p. 11; APA *Standards,* ¶ 14, p. 66 and EEOC *Guidelines,* 29 C.F.R. § 1607.6.

[103] 512 F.2d at 963-64. If an appropriate rationale exists for 40 as the cutting score, then from a professional standpoint, the court of appeals' concern with the fact that no applicants who scored below 40 had been admitted to the Police Academy for comparison purposes would be obviated.

Because the record to date is inadequate for a sound determination about the merits of the Futransky study, in light of the information an industrial psychologist would require in order to assess any such study, we believe that the case should be remanded for trial.[104] Confronted with a similarly deficient record in other cases, this Court has stated:[105]

> We consider it the part of good judicial administration to withhold decision of the ultimate question involved . . . until this or another record shall present a more solid basis of findings based on litigation or on a comprehensive set of agreed facts. While we might be able, on the present record, to reach a conclusion that would decide the case, it might well be found later to be lacking in the thoroughness that should precede judgment of this importance and which it is the purpose of the judicial process to provide.

That principle should be applied herein and the judgments below vacated so that the case may proceed to trial under proper direction from this Court.

---

[104] Indeed, except in cases where the absence of any adverse impact can be determined summarily, we believe it will be rare that summary judgment may be granted to either party on the issue of the job-relatedness of a selection device.

[105] *Kennedy v. Silas Mason Co.*, 334 U.S. 249, 257 (1948).

CONCLUSION

The *amicus curiae* urges the Court to reverse the court of appeals and remand the case to the district court for a full evidentiary hearing and determination under appropriate standards as set forth hereinabove.

Respectfully submitted,

R. LAWRENCE ASHE, JR.

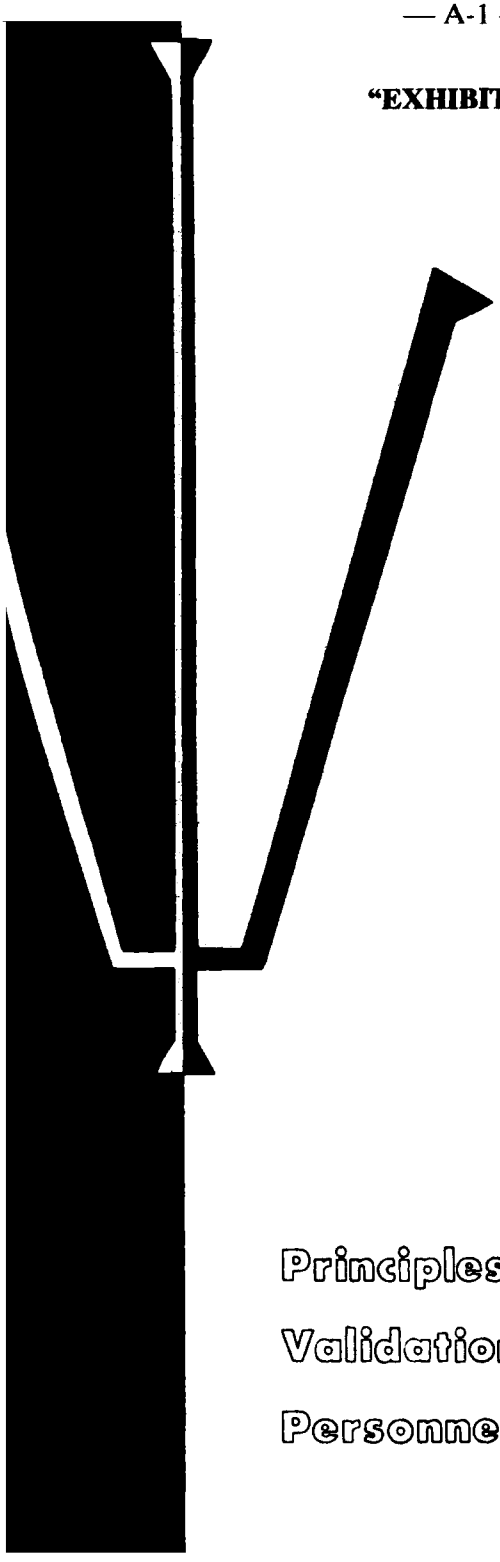
SUSAN A. CAHOON

WILLIAM A. WRIGHT
3100 Equitable Building
100 Peachtree Street
Atlanta, Georgia 30303
(404) 522-3100
Attorneys for *Amicus Curiae*
Executive Committee of Division 14,
American Psychological Association

Of Counsel:

KILPATRICK, CODY, ROGERS,
McCLATCHEY & REGENSTEIN

"EXHIBIT A"



Principles For The Validation And Use Of Personnel Selection Procedures

This document is an official policy statement of the Division of Industrial-Organizational Psychology, American Psychological Association. It does not, however, necessarily represent the policy of the Association. Copies are available at $1. each from the Secretary-Treasurer of the Division:

Dr. Mary L. Tenopyr
American Telephone & Telegraph Co.
195 Broadway, Room C 1620
New York, New York 10007

Published by **The Industrial-Organizational Psychologist**, Dayton, Ohio, and printed by the Hamilton Printing Company, Hamilton, Ohio.

Principles for the Validation and Use of Personnel Selection Procedures

Division
of
Industrial – Organizational Psychology
American Psychological Association
1975

Foreword

Because of growing concern over professional standards for employee selection research, I was instructed by the Division 14 Executive Committee at its August, 1974, meeting to appoint an ad hoc committee to develop an appropriate set of principles for the validation and use of personnel selection procedures. This document resulted. Its objective is to provide Division 14 members with professionally developed guidelines which they can follow in conducting validation research.

Robert M. Guion was appointed chairperson of the committee and **Mary L. Tenopyr** was named to work with him in drafting the document. Twenty-six Division members were appointed to review the various drafts and to advise on the content and style of the document. In addition, the April, 1975, issue of **The Industrial-Organizational Psychologist** announced the availability of the third draft of the **Principles** to members of the Division who might wish to review and comment on the document. A number of Division members took advantage of this opportunity. Furthermore, comments and suggestions regarding the **Principles** were solicited and obtained from the Committee on Psychological Tests and Assessment of the American Psychological Association.

Dr. Guion's qualifications as chairperson of the ad hoc committee are innumerable. Among many that could be cited are his roles as principal author of **Standards for Educational and Psychological Tests,** published by APA, and as a former member of the advisory committee on testing to the Office of Federal Contract Compliance. Dr. Tenopyr also served on the OFCC advisory committee and is widely recognized by the profession for her many contributions to psychological testing and its applications to employee selection.

At its May, 1975, meeting the Executive Committee of Division 14 reviewed the final draft of the **Principles** and authorized its publication. This is, therefore, an official document of the Division of Industrial and Organizational Psychology.

I extend the gratitude of our Division to Drs. Guion and Tenopyr for their effective efforts in writing the **Principles,** to the members of the ad hoc committee who advised on the document, to the Division 14 members who offered many useful comments and suggestions, to the members of the APA Committee on Psychological Tests and Assessment for their help, and to Dr. **Arthur C. MacKinney** who arranged for its publication.

Donald L. Grant
President

Executive Committee, Division 14

Michael Beer, Ph. D.
Corning Glass Works

John P. Campbell, Ph. D.
University of Minnesota

Edwin A. Fleishman, Ph. D.
American Institutes for Research

Frank Friedlander, Ph. D.
Case Western Reserve University

Donald L. Grant, Ph. D.
American Telephone & Telegraph Company

Edward E. Lawler, III, Ph. D.
University of Michigan

Lyman W. Porter, Ph. D.
University of California, Irvine

Virginia E. Schein, Ph. D.
Metropolitan Life Insurance Company

Mary L. Tenopyr, Ph. D.
American Telephone & Telegraph Company

Paul W. Thayer, Ph. D.
Life Insurance Marketing & Research Association

Victor H. Vroom, Ph. D.
Yale University

Ad Hoc Committee on Validation and Use of Personnel Selection Procedures, Division 14

Robert M. Guion, Ph. D. (Chair)
Bowling Green State University

Mary L. Tenopyr, Ph. D. (Chair)
American Telephone & Telegraph Company

Stanley R. Acker, Ph. D.
Olin Corporation

Lewis E. Albright, Ph. D.
Kaiser Aluminum & Chemical Corporation

Philip Ash, Ph. D.
University of Illinois, Chicago Circle

Richard S. Barrett, Ph. D.
Stevens Institute of Technology

C. J. Bartlett, Ph. D.
University of Maryland

Brent N. Baxter, Ph. D.
American Institutes for Research

William C. Burns
Pacific Gas & Electric Company

Joel T. Campbell, Ph. D.
Educational Testing Service

John C. Denton, Ph. D.
Psychological Business Research

Jerome E. Doppelt, Ph. D.
The Psychological Corporation

Robert D. Dugan, Ph. D.
International Telephone & Telegraph Corporation

Marvin D. Dunnette, Ph. D.
University of Minnesota

Edwin A. Fleishman, Ph. D.
American Institutes for Research

Donald L. Grant, Ph. D.
American Telephone & Telegraph Company

Raymond A. Katzell, Ph. D.
New York University

James J. Kirkpatrick, Ph. D.
California State College, Long Beach

Charles H. Lawshe, Ph. D.
Purdue University

Howard Lockwood
Lockheed Aircraft Corporation

William A. Owens, Ph. D.
University of Georgia

Erich P. Prien, Ph. D.
Memphis State University

Floyd L. Ruch, Ph. D.
Psychological Services, Inc.

Wayne W. Sorenson, Ph. D.
State Farm Insurance Company

Charles P. Sparks
Exxon Company

Paul W. Thayer, Ph. D.
Life Insurance Marketing & Research Association

Robert L. Thorndike, Ph. D.
Teachers College, Columbia University

Victor R. Tom, Ph. D.
Security Pacific National Bank

Principles for the Validation and Use of Personnel Selection Procedures

Statement of Purpose

This statement of principles has been adopted by the Executive Committee of the Division of Industrial—Organizational Psychology, American Psychological Association, as the official statement of the Division concerning procedures for validation research, personnel selection, and promotion. The purpose is to outline principles of good practice in the choice, development, and evaluation of personnel selection procedures. When using standardized tests or other selection procedures, the essential principle is that evidence be accumulated to show a relationship between decisions based on assessments made by a given procedure and subsequent criteria, such as job performance, training performance, permanence, advancement, and other job behavior.

This statement intends to provide:

(1) principles upon which personnel research may be based,

(2) guidance for practitioners conducting validation,

(3) principles of use of valid selection procedures, and

(4) information which may be interpreted for personnel managers and others who may be responsible for authorizing or implementing validation efforts.

The interests of some people will not be addressed by this statement. These principles are **not** intended to:

(1) be a technical translation of existing or anticipated legislation,

(2) substitute for textbooks outlining validation procedures,

(3) be exhaustive (they cover some of the more important aspects of validation), or

(4) freeze the field to prescribed practices, nor to limit creative endeavors.

The last point deserves emphasis. Traditional technology calls for a showing that (a) assessments made by a particular method (or combination of methods) are useful for predicting behavior in some aspect of employment and (b) that the predictions can be made within an acceptable allowance for error (usually expressed in terms of percentage of misclassification or correlation coefficients). Principles presented here are stated in the context of the traditional approach. Other approaches are not explicitly addressed here, e.g., the use of formal decision theory (Cronbach & Gleser, 1965; Dunnette, 1974) or various forms of synthetic validity (Guion, 1965; McCormick, 1959, in press; Primoff, 1959). The traditional approach is emphasized because its principles have been established through a long history. Other approaches may be equally good or even superior, but it seems premature to try to articulate formally the principles of their use.

These principles are meant to be consistent with the **Standards for Educational and Psychological Tests** (APA, 1974). They are intended to clarify the applicability of the **Standards** (written for measurement problems in general) to the specific problems of employee selection, placement, and promotion. Like the **Standards**, these guidelines present ideals toward which the members of this Division and other psychologists are expected to strive. Circumstances in any individual study will affect the applicability of any given principle. Psychologists should, however, consider very carefully any factors suggesting that a general principle is inapplicable or that its implementation is not feasible. It is appropriate to bear in mind the following statement from the **Standards**:

> A final caveat is necessary in the view of the prominence of testing issues in litigation. This document is prepared as a technical guide for those

within the sponsoring professions; it is **not** written as law. What is intended is a set of standards to be used, in part, for self-evaluation by test developers and test users. An evaluation of their competence does not rest on the literal satisfaction of every relevant provision of this document. The individual standards are statements of ideals or goals, some having priority over others. Instead, an evaluation of competence depends on the degree to which the intent of this document has been satisfied by the test developer or user (APA, 1974, p. 8).

Unlike the **Standards**, these guidelines will contain references for further reading. References to individual standards themselves will be by number.

There are many legitimate uses of tests within organizations which are not covered by these guidelines. For example, tests might be used solely for organizational analysis or for evaluation of training programs; these uses, although necessarily consistent with the **Standards**, are not covered here.

A Comment on "Fairness"

Social and legal influences have led to a concern, shared by psychologists, for fairness or equality in employment opportunity. A basic assumption of this statement of the principles of good practice is that those who follow them will also further the principle of fair employment. The interests of employers, applicants, and the public at large are best served when selection is made by the most valid means available. Bias in the use of employment procedures is ineffective for reaching both employer and job applicant objectives.

There are technical problems associated with the detection and reduction of bias. A simple difference between groups, whether in selection ratios, mean scores, or correlation coefficients, is **not** adequate evidence of bias in the use of an assessment procedure; bias is detected in reliable differences between groups in predictions or in the accuracy of predictions. This statement of principles does not choose between different statistical definitions of bias, some of which are essentially incompatible and have generated scientific controversy (Cleary, 1968; Cole, 1973; Darlington, 1971; Guion, 1966; Petersen & Novick, in press; Thorndike, 1971). The choice of a statistical definition depends on the psychologist's objectives.

These guidelines are technical in focus. They are principally concerned with validity. The maximization of opportunities for sub-group members can be most effective where validity enables one to attain the highest level of accuracy in prediction or assessment of qualifications.

Application of Principles

It is not likely that anyone will completely satisfy the ideal of every applicable principle. This probability raises the question of relative levels of stringency in adhering to the individual principles. The importance of a principle depends on the consequences of error. Will errors result in physical, psychological, or economical injury to people? Will the operating efficiency of the organization be impaired because of selection errors? If so, then the principles may need to be followed more rigorously than in less crucial situations.

Three axioms underlie the application of all these principles:

(1) Individuals differ in many ways.

(2) Individual differences in personal characteristics and backgrounds are often related to individual differences in behavior and satisfaction on the job.

(3) It is in the best interests of organizations and employees that information about relevant differences between people be developed and used in assigning people to jobs.

Definition of Purposes

Before any assessment procedure is considered or any validation effort planned, one should have a clear idea of what the assessment or validation is for. Any such statement of purpose logically must come from an un-

derstanding of the needs of the employing organization and of its present and prospective employees. As a general matter, a psychologist should develop clear objectives for an assessment procedure and design the validation effort to determine how well they have been achieved; those objectives should be consistent with professional, ethical, and legal responsibilities.

Ideally, all aspects of the decision-making process should make a valid contribution to achieving those purposes. Psychologists should demonstrate the validity of as many aspects of the decision-making process as feasible; generally, all assessment methods should be shown to be valid. When it is impossible or infeasible to apply validation methods to a given part of the decision-making process, that part should have a relationship, discernible by a reasonable person, to appropriate purposes of the employer.

The three aspects of validity described in the **Standards** are criterion-related validity (predictive or concurrent), construct validity, and content validity. Any of these may be emphasized in showing a relationship between various parts of the selection process and the objectives of the organization.

Criterion-Related Validity

In general, the purpose of employee testing is to predict future behavior, measured by a "criterion." The success or failure of the validation effort depends in large part on the adequacy of the criteria. The choice of measures to predict, and of measures to predict from, must be made thoughtfully and with great care.

In this section, the word "predictor" will be used in preference to "test." The competently developed and standardized ability test or personality inventory may be assumed as a mode. The term "predictor" should not be limited to such measures but should include, for example, biographical data and interview ratings as well. Insofar as technology and ingenuity permit, predictors should be standardized and should yield quantified "scores" or scale values amenable to psychometric analysis. The principles in this section apply to all predictors, although more easily to those most thoroughly standardized.

A. Criterion Development

1. **All criteria should be clearly related to the psychologist's purposes.** Criteria should be chosen, not on the basis of availability, but on the basis of importance and relevance. This implies that (a) the purposes are clear, (b) they are acceptable in the social and legal context in which the employing organization functions, and (c) are appropriate to the employing organization's purposes.

2. **All criteria should represent important work behaviors or work outputs, on the job or in training, as indicated by an appropriate review of information about the job.** Criteria need not be all-inclusive, but there should be clear documentation of the reasoning determining what is and what is not included in a criterion. Criteria need not reflect actual job performance. Depending on the purpose of the test user, various criteria such as overall proficiency, training time, sales records, number of prospects called, and turnover may be used (Wallace, 1965).

3. **If a criterion construct is not being satisfactorily measured, then substitute attempts to find another criterion measure should continue to focus on measurement of the same construct;** it is not acceptable practice to use substitutes measuring constructs not related to the psychologist's purposes. For example, if work output records prove unreliable or otherwise unsatisfactory, one should seek other measures of productivity and not shift to measures of employee conduct (such as absenteeism). One's purposes may, of course, call for the prediction of several constructs. However, one does not

choose a construct irrelevant to those purposes simply because it is predictable.

4. **The possibility of bias or other contamination should be considered.** Ethnic or sex bias, correlations between ratings and the length of acquaintance of the rater and ratee, or differences between day and night shifts are among variables which may bias the criterion measure. There is no clear path to truth in this matter. A difference between ethnic or sexual groups, older or younger employees, or day and night shifts may reflect bias in raters, equipment, or conditions — but it may also reflect genuine differences in performance. What is required is that the psychologist **consider** the possibility of bias, gather information relevant to that consideration, use his best judgment in evaluating the data, and be able to explain that judgment.

5. **If several criteria are combined to obtain a single measure, there should be a rationale to support the rules for combination.**

6. **Criterion measures should be reliable.** The reliability of a criterion measure should be estimated where feasible; estimates should be based on appropriate methods (Stanley, 1971). Reliability of a criterion measure need not be high, but there must be some reliability (Thorndike, 1949). Obviously, nothing will correlate with an unreliable criterion.

B. Choice of Predictor

1. **Predictor constructs should be chosen for which there is an empirical or logical foundation** (cf. Std. H1, H1.2). This principle does not call for elegance in the reasoning behind the choice of predictors so much as for having some reasoning; there should be some reason to suppose that a relationship exists between a predictor chosen and the behavior it is supposed to predict. For example, the research literature or the logic of development may provide the reason. One should consider alternative predictors for people of differing experience levels or probable approaches to the job (cf. Std. H1.2). This principle does not intend to rule out application of serendipitous findings, although such findings usually need verification.

2. **An investigator must be flexible in considering options in the choice of predictors,** i.e., not wedded to any favorite measurement technique nor prejudiced against any. No predictor should be ruled out at the stage of preliminary consideration except, of course, for technical or practical inadequacies or for legal or ethical reasons.

3. **Among predictors relevant to the purposes, those which minimize the effects of testers or testing situations should be chosen** (cf. Std. H4). The assessment of a candidate should not depend on who did the assessing. This is one reason why interviews are ordinarily less desirable than tests as predictors. Where non-test predictors are chosen, care should be taken to develop procedures which will minimize variance due to different users.

4. **Outcomes of decision strategies should be recognized as predictors.** Whether elements entering a decision have been quantified or not, the predictor having the effect is the decision reached. This fact applies to a range of decision strategies from test score composites to interviewers' judgments. If the decision strategy is to combine test and non-test data (reference checks, medical data, etc.) into a subjective judgment, the predictor in the final analysis is the judgment reached by the person who weighs all the information.

C. Design of Validation Research

1. **The feasibility of studying criterion-related validity should be considered carefully.** The ideal is that only predictors validated in well-designed studies should be used in making employment decisions. It is not always possible to do a well-designed or even reasonably competent study; a poor study is **not** better than none at all. Several questions must be considered in

deciding whether a study of criterion-related validity is feasible (cf. Stds., pp. 26-28).

First, one must consider the rate of change of job characteristics. The logic of validation research assumes that it is done under conditions representative of those in which the predictor will eventually be used. If technology, product, management practice, applicant populations, or other conditions are especially likely to change in ways that might affect validity, then a single criterion-related study is probably not useful, and one should consider alternative strategies, including a continuing program of research.

Second, a reasonably valid, uncontaminated, and reliable criterion is assumed in criterion-related validation. If such a criterion measure cannot be developed, criterion-related validation is not feasible. "Criterion-related validity studies based on the 'criterion at hand,' chosen more for availability than for a place in a carefully reasoned hypothesis, are to be deplored" (APA, 1974, p. 27).

Third, a competent criterion-related validation is based on a sample representative of the population of people and jobs to which the results are to be generalized. A wide variety of influences may distort actual samples: restriction of range, the use of existing employees rather than applicants, attrition, or population changes over time. Severe distortion from any source may render criterion-related validation infeasible.

Fourth, it is not useful to do a study where reliable results cannot be obtained. Of necessity, many employment test validations are done with numbers of cases too small to do the job as reliably as desired. There is a point in any practical situation where one says flatly that the **N** is too small. No firm minimum **N**, applicable to all situations, can be recommended, although a procedure can be. For example, the psychologist might, in a bivariate study, estimate the minimum validity coefficient he would accept in that situation. Then, on the basis of conventional statistics, he can determine the minimum **N** at which the validity coefficient would have a sufficiently narrow confidence interval. At the very least, the confidence interval should be narrow enough to exclude a value of zero.

If the above requirements cannot be met, the situation does not lend itself to the use of criterion-related validity for the evaluation of assessment, and the psychologist should so advise the client or employer.

2. **The appropriate validation model should be used.** Three methods have had some use for criterion-related validation: predictive, concurrent, and classificatory methods. Each of these is designed to answer a different research question; therefore, they are not ordinarily interchangeable. Moreover, any inferences the test user makes from a criterion-related study using a given method should be clearly stated within the context of the questions that method is designed to answer.

a. The predictive model, in which predictor information is obtained prior to placement of employees on a job and criterion information is obtained later, answers the most common employment question: whether the predictor does indeed have predictive value with respect to **later** job behavior. As such, the predictive model is, from the standpoint of scientific merit, to be preferred in most employee selection research. Its use properly begins with job candidates, not incumbents.

b. The concurrent method, in which both predictor and criterion information are obtained for present employees at approximately the same time, cannot be expected to answer questions of prediction. The concurrent method can only answer questions about the relationships of a given characteristic of preselected employees at a designated time. Often, the psychologist must choose between a concurrent study and no study at all. If concurrent validation is used, the psychologist should be par-

ticularly aware of restriction in range which may have occurred in criterion or predictor variables, effect of learning on the job on performance on these measures, and differences between the employee sample and the job candidate population on other relevant variables such as age or motivation. If any of these three considerations suggests that the results of a concurrent study would differ markedly from those of a predictive study, the psychologist is advised to use the predictive method or to declare criterion-related research infeasible.

c. The classificatory method (e.g., classifying scores as more like those either of people on a certain job or of people in general) is useful for answering questions about the degree to which persons not in a job compare with those holding the job. When answers to questions concerning probable performance or competence of job candidates are called for, the classificatory method is rarely appropriate.

3. **The sample upon which validation research is based should be reasonably representative of the population to which the predictor will be applied.** Because of the difficulty in defining the candidate population, this principle is difficult to follow, despite its importance. Among other things, it implies the superiority of predictive over concurrent designs, or, more accurately, of the use of an applicant sample rather than a presently employed sample in gathering predictor data.

When there is a substantial restriction of range in performance in the sample, a statistical correction for the effects of the restriction may be applied to a validity coefficient. However, it should be recognized that these apparently simple corrections may not be justified. Unless specific conditions exist, the results of such correction are subject to considerable error (Brewer & Hills, 1969). Any such correction should, of course, be based on data from the appropriate job candidate population.

4. **Validation research should ordinarily be directed to entry jobs, immediate promotions, or jobs likely to be attained.** Where a selection procedure is designed for a higher level job than that for which candidates are initially selected, that job may be considered an appropriate target job if the majority of the individuals who remain employed and available for advancement progress to the higher level within a reasonable period of time (cf. Std. E7.4.2). The point here is that predictability often diminishes over long time spans because of changes that occur in jobs and many other variables.

5. **Validation research should, in general, be based on samples large enough to yield reliable results.** However, the combination of data from dissimilar groups, e.g., persons in different job levels, for the sake of obtaining a large sample size is to be avoided. Moreover, if the sample size is extremely large, results may be statistically significant but of little practical use. An extremely large sample or cross validation is required before any credence can be placed in unusual findings, including but not limited to suppressor effects, moderator effects, nonlinear regression, results of configural scoring, or any other result which is likely to be affected by capitalization on chance effects. Partial and multiple correlation, in particular, require careful preplanning of data collection.

6. **Where traditional criterion-related validation is not feasible, the psychologist should consider alternative research strategies.** These may involve techniques not yet well understood or sufficiently studied. Examples might include the use of assessment centers, personal appraisals, or other judgments based on both person and job characteristics in which the predictor to be validated is a "final judgment." Other examples include "synthetic" validation or cooperative research plans such as industry-wide validation, consortia of small users, or generalization of validity. Such activities call for some pioneering research work, and they are not to be undertaken lightly.

7. **Procedures for test administration and scoring in validation research should be standardized and should be consistent with the standardization planned for operational use** (cf. Std. C1). Any operational characteristics (such as time limits, oral instructions, practice problems, answer sheets, scoring formulas, and physical conditions in the testing area) should be clearly defined and followed in validation research. Failure to follow the same standard procedures in operational use would suggest that validity of the predictor as used is in fact unknown, despite the prior research. The point of this principle is that the research should be consistent with procedures that will actually be used.

8. **Procedures of data collection should be consistent with the purposes of the study** (cf. Std. E7). If possible, predictors should be validated prior to the proposed use (cf. Std. E7.1). Some employers consider this principle difficult to follow because of the need to get on with the business of making employment decisions. Where there is external evidence which supports the probability of valid prediction (such as evidence of construct validity and of the appropriateness of the construct, carefully reasoned inferences from job descriptions, records of success of the same predictor in apparently similar situations, etc.), it may be feasible to utilize the predictors immediately. However, the psychologist must avoid situations that make it impossible or difficult to detect validity. For example, decisions should not be so highly selective that severe restriction of range results.

There is often no firm basis for the presumption of validity. In such cases, the psychologist must carefully judge whether the dangers of postponing the use of the predictor are greater or less than the dangers of using it prematurely.

9. **The collection of predictor data and collection of criterion measures should be operationally independent** (cf. Stds. E4.11, E7.3). A common example of non-independence is the collection of criterion ratings from supervisors who know the test scores. If a significant validity coefficient is obtained, it may be due either to a valid relationship or to manipulation of ratings to conform to scores. Such ambiguity should be avoided.

D. Data Analysis

1. **The type of statistical analysis to be used should be specified in planning the research.** The kinds of decisions to be made, and the way in which predictor variables are to be used in determining those decisions, should be considered in determining the method of analysis to be used. Although any standard method of analysis may be used, any new or unusual method should be clearly explained in the research report. (It is understood that conditions may arise requiring changes in plans.)

2. **Data analysis should yield complete information about the relationship between predictor and criterion measures** (cf. Stds. E8, E6.2). The analysis should provide information about the statistical significance of the relationship. Traditionally, a validity coefficient or similar statistic which has a probability of less than one in twenty of having occurred by chance may be considered valid for typical purposes. There may be exceptions to this rule; professional standards have never insisted on a specific level of confidence. However, departures from this traditional convention should be based on reasons which can be stated in advance (such as power functions, utility, economic necessity, etc.).

The analysis should provide information about the strength of the relationship. This is usually expressed in terms of correlation coefficients, but other methods (such as per cent of misclassification) are acceptable and even preferable in many situations. The analysis should also give information about the nature of the relationship and how it might be used in prediction. Regression equations or expectancy tables may be appropriate.

Complete information includes numbers of cases and measures of central value and variability of both predictor and criterion variables.

3. **Statistical corrections of correlations should be made when required for appropriate inferences about populations from which the samples are drawn** (cf. Stds. E6.2.2; E8.2). It is important to note that traditionally-developed confidence intervals appropriate for uncorrected coefficients of correlation do not apply to coefficients adjusted for the effects of restriction of range or unreliability. Usually, the corrected coefficient can serve only as a basis for a point estimate (not an interval) of the population coefficient.

If corrections are made, uncorrected **r's** should, of course, also be reported.

4. **The analysis should, when reasonable and feasible, investigate the possibility of moderating effects.** If the candidate sample can be divided into sub-groups where prior research or logical considerations (such as indications of gross differences in job duties) suggest different relationships between predictors and criteria, the analyses should be done separately for the sub-groups if technical and situational considerations permit (cf. Std. E9). The implementation of this principle may be approached through an investigation of such moderator variables as ethnic group, socio-economic status, age, sex, cognitive styles, etc., although this approach has generally not been found to improve validity appreciably. There are many difficulties in research with moderator variables. Technical and situational considerations often preclude the proper conduct of such studies. The moderator variable adds a variate and a multiplicative term to the regression model, and cross validation is therefore essential before the moderator can be used for selection decisions. Again, a poorly conducted study is **not** better than no study at all (cf. Std. E10.2).

5. **If a validity coefficient is to be adjusted for the unreliability of the criterion, the method of estimating criterion reliability should avoid spuriously low estimates.** A spuriously low estimate of criterion reliability produces a spuriously high estimate of the adjusted validity coefficient. Particularly inappropriate is the use of internal consistency reliability with a heterogeneous measure. No adjustment of a validity coefficient for unreliability of a predictor should be reported unless one clearly states that the resultant coefficient is purely theoretical and unachievable with the actual predictor involved.

Any estimates of reliability should be presented in reports in such a way that an employer or other reader will not confuse the reliability statement with a validity coefficient (cf. Std. J3).

6. **Where predictors are used in combination, the combination of intended use should be validated and the method of combination cross validated or replicated where technically feasible** (cf. Stds. E8.4, E8.5, E10, E10.1, E10.11, E10.2). Simple linear combinations are generally more reliable than complex non-linear combinations, simple weights are more reliable than complex weights, and confirming evidence from more than one sample is more reliable than evidence from a single sample. (This may be a useful place to reiterate the caveat that these principles are ideals. Clearly, one does not cross validate a regression equation where the **N** barely permits the original computation of regression weights. One does not divide a group of merely acceptable size into two small groups, from which only unreliable statistics may be derived, and call it replication. One does, however, make some judgments about the costs of using data that have not been cross validated relative to the cost of alternatives — which might be limited essentially to random selection. In short, one does the best one can.)

7. **Any method of analysis should be chosen with due consideration of the characteristics of the data and of the assumptions involved in the development of the method** (cf. Std. E8.1.3). Some violation of assumptions

may usually be tolerated with few ill effects; others may be grossly misleading. It is the responsibility of the investigator to know the assumptions of the method chosen and the consequences of violations of them.

8. **Data should be free from clerical error.** Coding and computational work should be checked carefully and thoroughly.

Construct Validity

The notion of construct validity, with its many optional procedures, may be extended to the point where it may be used to justify selection procedures. That justification requires that the construct be well defined, that the selection procedure considered is a measure of that construct, and that an appropriate criterion of job behavior involves that construct to more than a tangential degree.

In view of the lack of a substantial literature extending the concept of construct validity to employment practice, no principles for its use are presented here. Psychologists should, however, be aware that obtaining support for the relevance of a construct to a particular job, and of the validity with which a particular selection procedure measures that construct, is both an extensive and arduous undertaking, involving more than a single criterion-related validity study. It is, however, an undertaking that may pay great dividends in improving the scientific foundations for employment decisions.

Content Validity
(Content-Oriented Test Development and Use)

The content validity of a test is the degree to which scores on a test may be accepted as representative of performance within a specifically defined content domain of which the test is a sample. If the test is to be used for employment decisions, the relevant content domain is performance (or the knowledge or skill necessary for performance) on the job or on specified aspects of it. A test may be highly valid as a sample of a given content domain, but if that domain is not an important part of the job, the value of the test for employment purposes is negligible.

The distinction between a content domain and a psychological construct is not always a clean one. A useful distinction between content validity and construct validity can be made, however, in terms of the methodology associated with each. Construct validation is essentially an empirical, statistical process, involving more than a single validity coefficient. On the other hand, content validation has been primarily a judgmental process concerned with the adequacy of a test as a sample of specified activities. (Efforts to quantify such judgments have been proposed but have not yet had extensive trials.)

Content sampling is properly involved in any test construction, whether scores are to be interpreted as measures of achievement or as measures of an abstract construct. This discussion is limited, however, to situations where the assessment is evaluated solely in terms of content sampling. It is to be noted that content sampling is as useful in the construction and evaluation of criterion measures as it is for tests used for employment decisions.

In content sampling, any inferences about the usefulness of a score must be preceded by a set of inferences about the instrument itself based on the method of its construction (Messick, 1974). For that reason, the emphasis of this section and of its title is on the development of content-oriented assessment instruments rather than on the inferences from scores. Any evaluation of existing tests in terms of the adequacy of content sampling might follow parallel considerations.

A. **The job content domain to be sampled should be defined.** That definition should be based on an understanding of the job, organizational needs, labor markets, and other considerations leading to personnel specifications and relevant to the organization's purposes. The domain need

not be inclusive insofar as any larger domain is concerned. By this we mean that it does not have to cover the entire universe of duties of a particular job or of topics covered in a training course. In fact, there may be many domains in the total content universe for any given job. For what it **does** include, a job content domain should be completely defined and thoroughly described.

The **Standards** (p. 29) discuss the question of domain definition for employment tests. The following statements are intended to clarify that discussion:

1. Job content domains should be developed and defined by job analysis, which may be a formal investigation, or the pooled judgments of informed persons such as production engineers, job incumbents, their supervisors, or personnel specialists. The domain should be defined on the basis of competent information about job tasks and responsibilities.

2. Job content domains should be defined in terms of those things an employee is expected to do without training or experience on the job, i.e., the content should not cover knowledge or skills the employee will be expected to learn after placement on the job or in training for the job.

3. The definition may be restricted to "critical, most frequent, or prerequisite work behaviors" (p. 29). There is no virtue in measuring ability to handle trivial aspects of the work.

4. A test content domain may likewise be defined. Essentially, the content validity of an employment test should be seen as the degree to which a sample of elements from a **test** content domain matches the elements of a **job** content domain.

5. Once a specific job content domain has been defined, subject to the above constraints, an employer can justify the use of an employment test on the grounds of content validity if he can demonstrate that the content of the test is reasonably representative of important aspects of the job domain.

B. **A content domain should ordinarily be defined in terms of tasks, activities, or responsibilities.**

The principle here is that the domain be defined principally in terms of activities or consequences of activities which can either be observed or be reported by the job incumbent. One can add to this nucleus, without straining credulity, statements of specific items of knowledge, or specific job skills, prerequisite to effective activity. It is a much larger "inferential leap," however, to move from observation to inferences concerning underlying psychological constructs such as empathy, dominance, dexterity, leadership skill, spatial ability, etc. Such constructs suggest hypotheses to be tested in criterion-related or other empirical research. It is therefore inappropriate to define job domains in such terms if one's purpose is to develop and justify a test solely on the basis of that domain.

It follows that many tests developed for general use in a variety of situations are not representative samples of an appropriately defined domain of job content. Most such tests tap general constructs, not samples of specific knowledge or behavior required in a specific setting.

C. **Sampling of a job content domain should assure the inclusion in a measure of the major elements of the defined domain. Random sampling is probably inappropriate unless done within systematically sample areas or "subdomains."**

Sampling the job content domain **is** the process of constructing the selection procedure. The selection procedure should be developed according to accepted professional practices; however, once the domain and sampling procedure are defined, the actual construction of a test may not require the services of a psychologist.

Consider, for example, the content validity of assessments of performance during a new employee's probationary period. If random assignments include many trivial activities and exclude many crucial ones, the probationary period lacks adequate content validity. Probationary tasks should be systematically designed to include crucial assignments or assignments to activities which must be competently carried out before the worker is ready for further training.

D. **Panels of experts used in any aspect of the development of tests defended on grounds of content validity should be clearly qualified.** In developing evidence of content validity, the psychologist should document carefully every step of the procedure. Panels of experts (i.e., people with thorough knowledge of the job) may be used in defining domains, in writing test items, in developing simulation exercises, and in evaluating items or total tests. An important part of the documentation is a thorough statement of the qualifications and job knowledge of the people on such panels.

Implementation

Validation, discussed in the preceding sections, is the investigatory phase in the development of selection procedures. Whatever the outcome of such research, the psychologist should prepare a report of the findings; the importance of documentation in the form of such a report is especially great if the assessment procedure is to be adopted for operational use. Many valid testing programs fail at the point of their implementation. The following principles are intended to assure effective and proper use of measures found valid.

A. Research Reports and Procedures Manuals

1. **Whenever an assessment procedure is made available for use in employment decisions, one or more documents should be prepared to describe validation research and the standard procedures to be followed in using the results of that research** (cf. Std. A1). Reports of validation research should include enough detail to enable a competent fellow psychologist to know precisely what was done and to draw independent conclusions in evaluating the work. A basic principle in the preparation of such reports is that they should not be misleading. For example, studies which result in negative findings should not be omitted or buried since they may influence overall conclusions.

2. **Informational material distributed within the organization should be accurate, complete for its purposes, and written in language that is not misleading** (cf. Std. A1. 2.3). Memoranda should be worded to communicate as clearly and accurately as possible the information that readers need to know to carry out their responsibilities competently and faithfully. Care must be taken in preparing such memoranda to avoid giving others within the organization an impression that an assessment program is more useful than it really is. Too often such memoranda serve as sales documents; persuasion may be necessary to get cooperation, but it should not be misleading.

3. **Research reports and procedures manuals should be reviewed periodically and revised as needed; any changes in use or in research data that would make any statement in such documents incorrect or misleading should result in revision.**

4. **Research reports or procedures manuals should help readers make correct interpretations of test data and should warn them against common misuses of tests and test information** (cf. Std. B1).

5. **The reasoning underlying an assessment program should be clearly stated in a research report or procedures manual** (cf. Std. A2). The selection of criteria may involve certain inferences about the nature of the job. The kinds of traits considered predictive of such criteria were determined by some sort of reasoning. Measures of criteria and predictors have been chosen and

may have been empirically or logically evaluated. The reasoning in all of these processes should be stated clearly enough for readers to evaluate that reasoning from their own perspectives. The point of these illustrations is that another psychologist should be able to evaluate the research reported not only in terms of statistical evidence but in terms of how well it fits whatever psychological insights he may have (cf. Std. B3).

6. **Any special qualifications required to administer a test or to interpret the scores should be clearly stated in the research report and/or procedures manual.**

7. **Any claim made for any selection procedure should be supported in documentation with all available research evidence, including evidence that may be unfavorable to the conclusion** (cf. Stds. B5, B5.1, B5.2, B5.3, B5.4, B5.5 and sections C and E above).

8. **Any procedures manual for people who administer tests (or use other predictors) should specify the procedures to be followed and emphasize the necessity for standardization; these instructions should be clear enough that all persons concerned know precisely what they are supposed to do** (cf. Stds. C1, C1.1, C1.11). In one unit of one organization, a timed arithmetic test was given with no time limits because "we weren't getting enough people to qualify." One must be both insistent and persuasive to get people to understand both the nature of and the need for standardization. Where the psychologist cannot supervise directly the administration of tests or the use of other procedures, periodic seminars may be needed to reinforce the written instructions; observational checks or other quality control mechanisms should be built into the system. It should be made clear to everyone that failure to follow standardized procedures renders the research report irrelevant to some degree. There may be situations where research is based on data from operational studies where non-standardized procedures may have been used and where the results show no serious impairment of validity. In such situations the degree of standardization is shown to be relatively unimportant; this, however, should not be assumed without investigation.

9. **Any scoring or scaling procedures should be presented in the procedures manual with as much detail and clarity as possible to reduce clerical errors in scoring and to increase the reliability of any judgments required.** When keys must be kept confidential, this material should be made available **only** to people who do the actual scoring or scaling of responses.

10. **A research report should contain clear and prominent descriptions of the samples used in the research; such information should also be summarized on any accompanying report forms in which scores are given with normative interpretations such as centiles or expectancies of success.**

Too many people do not know that normative interpretations and estimates of validity and reliability are specific to the study that produced them. Almost magical properties are sometimes attributed to test scores.

Ordinarily, norm tables are less useful than expectancy charts for employment decisions. One should recognize, of course, that the expectancy chart is a normative interpretation of test scores; i.e., it indicates the proportion of a specific sample of candidates who reach a standard level of success. Norm tables may, therefore, be useful in identifying the effects of a cutting score, even if not in interpreting individual employment test scores.

11. **Any normative reporting should include measures of central tendency and variability and should clearly establish the nature of the normative data given,** i.e., centiles, standard scores, expectancies, predicted levels of attainment, etc. (cf. Std. D3).

12. **Any derived scale used for reporting scores should be carefully described in the research report or procedures manual.** Whether using standard derived scores (such as those described in general textbooks on measurement) or "home-grown" scales (such as "qualified," "marginal," or "unqualifed"), the psychologist should make clear their logical and psychometric foundations.

B. Validity Generalization

1. **Validity evidence obtained in one unit of a multiunit organization or in a consortium may be applied to other units where jobs and job settings are essentially similar.** Validity coefficients are obtained in specific situations. They apply only to those situations. A situation is defined by characteristics of the samples of people, of settings, of criteria, etc. Careful job and situational analyses are needed to determine whether characteristics of the site of the original research and those of other sites are sufficiently similar to make the inference of generalizability reasonable.

A pressing problem in employment psychology is that of determining how to generalize validities. Psychologists are strongly urged to engage in cooperative research ventures such as industry-wide validation studies, consortia of civil service jurisdictions, and the like. Until such time as such cooperative research results in an understanding of the limits of generalization, there will be few principles to observe in this area. The principle that one may apply validity evidence to essentially similar job units is an interim principle; it is not intended to discourage continued research to determine whether such application exceeds the legitimate boundaries of validity generalization.

2. **Assumptions of validity generalized from promotional literature, testimonial statements, or empirical studies in unrelated settings may not be used as evidence of the validity of the procedure in a specific situation.**

C. Use of Research Results

1. **It is the responsibility of the psychologist to recommend specific methods of score interpretation to the employer.** Although the employer usually reserves the final decision on whether to use a specific selection procedure, it is the responsibility of the psychologist to make recommendations on this question and on questions of how the procedure is to be used. The recommended use should be consistent with procedures with which validity was established.

2. **Validity of selection procedures should be assumed only for jobs for which validity has been situationally determined or for comparable jobs involving the same criteria to which validity may generalize.**

3. **The utility of a selection procedure should be considered in deciding whether to apply it operationally.** In reaching the decision, consideration should be given to relative costs and benefits to both the organization and its employees. It is not recommended that procedures with minimal usefulness be applied, but a procedure with at least some demonstrated usefulness is ordinarily preferable to one of unknown validity or usefulness.

4. **Selection standards may be set as high or as low as the purposes of the employer require, if they are based on valid predictors.** As in principle A.1 under criterion-related validity, this implies that (a) the purposes of selection are clear and (b) they are acceptable in the social and legal context in which the employing organization functions.

5. **Employers should provide reasonable opportunities for reconsidering candidates whenever alternative forms exist and reconsideration is technically feasible** (cf. Std. J7.2). Under at least some circumstances, employers should allow candidates to reapply. There might be any of several reasons for questioning the validity of prior assessment for any given person. Where there has been opportunity for new learning, retesting is usually a desirable practice.

In some situations, validity information does not generalize to the retest situation. In cases where biographical data or scored interview forms are the predictors, the validity of the retest is usually unknown; it may be negligible. The test user is expected to balance these opposing considerations.

6. **The use of a predictor should be accompanied by systematic procedures for developing additional data for continued research.** Changing social, economic, technical, or other factors may operate over time to alter or eliminate validity; continuing (or periodic) research is therefore necessary. A serious problem is that the operational use of a valid predictor may result in such severe restriction of range that its validity cannot be demonstrated in subsequent research (Peterson & Wallace, 1966). There is no well-established technology for checking validity of instruments in use; however, psychologists are urged to exercise their ingenuity to observe the principle that validity once demonstrated cannot be assumed to be eternal.

7. **All persons within the organization who have responsibilities related to the use of employment tests and related predictors should be qualified through appropriate training to carry out their responsibilities.** The psychologist or other person in charge of any testing program should know measurement principles and the limitations on the validities of interpretations of assessments (cf. Std. D1). That person should understand the literature relevant to the test use or employment problems. Other people in the organization may have some responsibilities related to the testing program. It is the psychologist's responsibility to see to it that such people have the training necessary to carry out those responsibilities competently (cf. Stds. G3, G3.1).

These considerations suggest the need for workshops, seminars, or other planned approaches to training technicians and managers involved in assessment procedures and in the interpretation of assessments.

8. **Psychologists should seek to avoid bias in choosing, administering, and interpreting tests; they should try to avoid even the appearance of discriminatory practice** (cf. Std. G4). This is another principle difficult to apply; it goes beyond data analysis. The appearance of bias may interfere with the effective performance of a candidate in the assessment situation. At the very least, a test user can create an environment that is responsive to the feelings of all candidates, insuring the dignity of persons.

9. **Psychologists should periodically review test use** (cf. Std. G5). Departures from established procedures often develop over time. New findings in psychological or psychometric theory, or new social criticisms, may be relevant to one or more of the assessment procedures in use. The principle is that it should not be left to chance to find examples of misuse or of obsolete data; some systematic plan for review should be followed.

10. **The psychologist is responsible for clerical accuracy in scoring, checking, coding, or recording test results** (cf. Std. I3). This principle applies to the psychologist and to any agent to whom he has delegated responsibility; the responsibility cannot be abandoned by purchasing services from an outside scoring service (cf. Std. I3.1).

11. **If cutting scores are used as a basis for decision (i.e., as rigid pass-fail points), the rationale or justification should be known to all users** (cf. Std. I4). This principle does not recommend cutting scores. Rather, "The intent is to recommend that test users avoid the practice of designating purely arbitrary cutting scores they can neither explain nor defend" (APA, 1974, p. 67). If cutting scores are to be established, some consideration should be given to the different effects of different cutting scores; e.g., the effects of the two kinds of error: selecting people who later prove unsatisfactory as opposed to rejecting people who would have been satisfactory if hired.

12. **The psychologist must make considered recommendations for the operational use of a predictor for which differential prediction is established.** A finding of differential prediction should not automatically lead to differences in predictor use for different groups. For example, if the study were based upon an extremely large sample, a finding of statistically significant differential prediction may have little practical impact. For another example, data apparently indicating differential prediction may be due to statistical artifacts or may suggest courses of action inconsistent with societal goals. In such situations, the reasonable course of action would be to recommend uniform operational use of the predictor for the different groups (or perhaps conduct further research).

Should a finding of differential prediction be compelling enough to warrant other action, possible approaches to dealing with it are (1) revising or replacing the tests involved, or (2) using the test operationally taking into account the differences in prediction systems.

Action under the second alternative should be in accordance with the definition of fairness upon which the study indicating differential prediction was based. (See "A Comment on Fairness," above.)

13. **The psychologist or other test user is responsible for maintaining test security** (cf. Std. I5). This means that all reasonable precautions should be taken to safeguard test materials and that decision makers should beware of basing decisions on scores obtained from insecure tests.

This principle is difficult to apply to non-test predictors such as judgments reached in an employment interview; nevertheless, the principle of senurity as a means for standardization may be applied to other variables as well. Reference checks, for example, should be held confidential as an illustration of the extension of this principle.

14.**In making interpretations of test scores the psychologist should be aware of situational variables introducing error** (cf. Std. J1). An individual test score may lead to invalid inferences because of unusual features of the testing situation (e.g., uncommon distractions), exceptional characteristics of the individual (e.g., a physical handicap) or the passage of time (e.g., new learning since testing occurred). Sometimes these may form a basis for retesting; they may suggest the consideration of other information. The principle is that some degree of judgment be retained in the interpretation of test scores obtained in circumstances differing from those in the validation research. Perhaps a better statement of the principle is that some degree of judgment should not be ruled out automatically in all situations.

15. **Test score information should not be available for use in personnel decisions when it is no longer valid.** It is recognized that some traits are more stable than others, but as a general principle, it is poor practice to retain test scores in personnel files long after the scores were obtained. Files should be purged of data, rendered invalid by new experience, aging, maturation, or other personal change — or by changes in jobs or in organizations — so that no one will base decisions on such invalid scores.

16. **When reporting test results, the psychologist should consider the level of knowledge of the person receiving the report; the report should be in terms likely to be interpreted correctly by persons at that level of knowledge.** Ordinarily, actual test scores should not be reported to candidates or to managerial personnel. If for any reason scores are reported, they should be explained carefully to be sure that interpretations are correct. In particular, one should not report actual scores to people who may later be asked to provide criterion ratings for validation.

17. **Scores on many tests developed for educational use are sometimes given in derived score form as I.Q.'s, grade-equivalent scores, or other terms not likely to be meaningful in the employment context; such terms are to be avoided** (cf. Std. D5.2.5; J4.2). Such terms are psychometrically and logically questionable as well as inappropriate for employment use. Even where they had legitimate psychometric significance historically, they have been so encrusted with spurious meaning that they lend themselves to misinterpretation.

Glossary

Assessment procedure: any method used to evaluate characteristics of persons.

Bias: any constant error; any systematic influence on measures or on statistical results irrelevant to the purpose of measurement.

Centile: commonly percentile; a point on a distribution. Centile rank or percentile rank refers to the percentage of all scores falling at or below a particular value.

Central value: an average of a set of measurements; e.g., mean, median.

Coefficient of correlation: an index number, which may be positive or negative, ranging from 0.00 to 1.00, indicating the extent to which two variables covary.

Construct: a formally articulated concept of a trait, i.e., an hypothesized property of people, objects, or events inferred from data.

Construct validity: the degree to which scores obtained through a specified test or other assessment procedure may be interpreted as measuring or reflecting a specified construct.

Content domain: a body of knowledge and/or a set of tasks or other behaviors defined so that given facts or behaviors may be classified as included or excluded. (See Job content domain, Job content universe.)

Content validity: the degree to which scores on a test may be accepted as representative of performance within a specifically defined content domain of which the test is a sample.

Correlation: the degree to which two or more sets of measurements vary together; e.g., a positive correlation exists when high values on one scale are associated with high values on another.

Criterion (pl., **criteria**): the behavior, performance level, or result of behavior, on the job or in training, to be predicted.

Criterion-related validity: the statistical statement of the existence of a relationship between scores on a predictor and scores on a criterion measure.

Cross validation: the application of a scoring system or set of weights empirically derived in one sample to a different sample (drawn from the same population) to investigate the stability of relationships based on the original weights.

Cutting score: a specified point in a predictor distribution below which candidates are rejected.

Derived scale: a scale of measurement using a system of standard units (based perhaps on standard deviations or centiles), to which obtained scores on any original scale may be transformed by appropriate numerical manipulation.

Expectancy table: a table or chart used for making convenient predictions of levels of criterion performance for specified intervals of predictor scores.

Feasible: capable of being done successfully; i.e., in criterion-related research, economically practical and technically possible without misleading or uninterpretable results.

Grade-equivalent score: a derived score which interprets a person's test performance in school grade norms; usually inappropriate for employment testing.

I.Q.: a derived score with a mean of 100 and, usually, a standard deviation of 15 or 16; used only with general mental ability or scholastic aptitude test; generally inappropriate for employment use.

Internal consistency: degree to which performance on a part of a test or other assessment procedure correlates with performance on other parts.

Job content domain: a defined segment or aspect of the job content universe regarding which inferences are to be made.

Job content universe: the total job; everything, known and unknown, which the incumbent does and must know in order to do it.

Linear combination: the sum of scores (whether weighted differentially or not) on different assessments to form a single composite score; distinguished from nonlinear combinations in which the different scores may, for example, be multiplied instead of added.

Moderator variable: theoretically, a variable which is related to the amount and type of relationship between two other variables; in practice, it is usually a basis for subdividing a sample into subgroups for independent analyses of the correlations of interest.

Normative: pertaining to norm groups, i.e., the sample of subjects from which were obtained descriptive statistics (e.g., measure of central tendency, variability, or correlation) or score interpretations (e.g., centiles or expectancies).

Objective: verifiable; in measurement, pertaining to scores obtained in a way that minimizes bias due to different observers or scorers.

Operational independence: gathering of data by methods that are different in procedure or source so that measurement of one variable, such as a criterion, is not influenced by the process of measuring another variable.

Predictor: a measurable characteristic used to predict criterion performance, e.g., scores on a test, judgments of interviewers, etc.

Psychometric: pertaining to the measurement of psychological characteristics such as aptitudes, personality traits, achievement, skill, knowledge, etc.

Regression equation: an algebraic equation which may be used to predict criterion performance from specific predictor scores.

Reliable: consistent or dependable; repeatable; reliability refers to the consistency of measurement.

Replication: a repetition of a research study designed to investigate the generality or stability of the results.

Restriction of range: a situation, varying in degree, in which the variability of data in a sample is less than the variability in the population from which the sample has been drawn.

Score: any specific number in a range of possible values describing the assessment of an individual; a generic term applied for convenience to such diverse kinds of measurement as tests, production counts, absence records, course grades, or ratings.

Standard deviation: a statistic used to describe the variability within a set of measurements, based on the differences between individual scores and the mean.

Standard score: a score which describes the location of a person's score within a set of scores in terms of distance from the mean in standard deviation units; may include scores on certain derived scales.

Synthetic validation: an approach to validation in which the validity of a test battery put together for a specific use may be inferred from prior research relating predictors to specified and relevant criterion elements.

Target job: the job in which performance is to be predicted at the time of the employment decision; not necessarily the initial job assignment.

Tester effect: bias in assessments attributable to differences between test administrators, test scorers, or others involved in the assessment procedure.

Validation: the process of investigation (i.e., research) through which the degree of validity of a predictor can be estimated. [Note: laymen often misinterpret the term as if it implied giving a stamp of approval; they should recognize that the result of the research might be zero validity.]

Validity: the degree to which certain specified inferences from scores on tests or other assessments may be considered justified or supported by research or by the method of test construction.

Validity coefficient: a correlation coefficient showing the strength of relationship between predictor and criterion.

Variability: the extent of individual differences in a particular variable.

Variance: a measure of variability; the square of the standard deviation.

References

American Psychological Association, American Educational Research Association, and National Council on Measurement in Education. **Standards for Educational and Psychological Tests.** Washington: American Psychological Association, 1974, 76 pages.

Brewer, J. K. and Hills, J. R. Univariate selection: The effects of size of correlation, degree of skew, and degree of restriction. **Psychometrika,** 1969, **34,** 347-361.

Cleary, T. A. Test bias: Prediction of grades of Negro and white students in integrated colleges. **Journal of Educational Measurement,** 1968, **5,** 115-124.

Cole, N. S. Bias in selection. **Journal of Educational Measurement,** 1973, **10,** 237-255.

Cronbach, L. J. and Gleser, G. C. **Psychological tests and personnel decisions** (2nd ed.). Urbana: University of Illinois Press, 1965.

Darlington, R. B. Another look at "cultural fairness." **Journal of Educational Measurement,** 1971, **8,** 71-82.

Dunnette, M. D. Personnel selection and job placement of disadvantaged and minority persons: Problems, issues, and suggestions. In H. L. Fromkin and J. J. Sherwood (Eds.), **Integrating the organizations.** New York: Free Press, 1974, 55-74.

Guion, R. M. Employment tests and discriminatory hiring. **Industrial Relations,** 1966, **5,** 20-37.

Guion, R. M. Synthetic validity in a small company: A demonstration. **Personnel Psychology,** 1965, **18,** 49-63. (b)

McCormick, E. J. Application of job analysis to indirect validity. **Personnel Psychology,** 1959, **12,** 402-413.

McCormick, E. J. Job and task analysis. In Dunnette, M. D. (Ed.), **Handbook of industrial and organizational psychology.** Chicago: Rand McNally, in press.

Messick, S. Meaning and values in measurement and evaluation. Address delivered at meeting of American Psychological Association, New Orleans, August, 1974.

Petersen, N. S. and Novick, M. R. An evaluation of some models for culture-fair selection. **Journal of Educational Measurement,** in press.

Peterson, D. A. and Wallace, S. R. Validation and revision of a test in use. **Journal of Applied Psychology,** 1966, **50,** 13-17.

Primoff, E. S. Empirical validations of the J-coefficient. **Personnel Psychology,** 1959, **12,** 413-418.

Stanley, J. C. Reliability. In Thorndike, R. L. (Ed.), **Educational Measurement** (2nd ed.). Washington: American Council on Education, 1971.

Thorndike, R. L. Concepts of culture-fairness. **Journal of Educational Measurement,** 1971, **8,** 63-70.

Wallace, S. R. Criteria for what? **American Psychologist,** 1965, **20,** 411-417.

EXHIBIT B

In the United States District Court
For the Northern District of Florida
Pensacola Division

Equal Employment Opportunity Commission, Plaintiff, vs. Monsanto Company, Defendant.	}	Civil Action No. 73-31 CIV-P
Eddie Stallworth, et al., Plaintiffs, vs. Monsanto Company, Defendant.	}	Civil Action No. 73-45 CIV-P

Washington, D. C.
September 17, 1974

Deposition of:

William H. Enneis

called for oral examination by counsel for the Defendant at the office of the Equal Employment Opportunity Commission, 1206 New Hampshire Avenue, Northwest, Washington, D. C. 20506, beginning at 10:00 a.m. on September 17, 1974, before Daniel R. Dotson, Jr., a Notary Public in and for the District of Columbia, when were present on behalf of the respective parties:

* * * * * * *

[*6] Q. Dr. Enneis, what is your residence address? A. I live in Washington, D. C.

* * * * * * *

Q. Where are you employed, sir? A. I work for the Equal Employment Opportunity Commission.

Q. Here in Washington? A. Yes.

Q. What is your job title and duties at the present time? A. My present title is Chief of the Research Studies Division and I am responsible for the analysis of sociological, phychological and economic data in the area of equal employment opportunity. And supervising people who do that analysis.

Q. Would you outline for us, please, your history of employment with the Commission from the time you first came to work for them, how your job titles and duties have changed, if in fact they have? A. I came to work for the Commission in September of 1966. My title was Staff Psychologist. In December of 1972, I was officially appointed Chief of the Research Studies Division. From approximately January of 1972 until December of that year I was in an acting capacity as the Chief of the Division, as Staff Psychologist, from 1966 until 1972.

[7] I was responsible for analyzing basically information and data related to validity of employee selection procedures, including psychological tests as they affected applicants insofar as they might have some indication or carry some indication of employment discrimination.

* * * * * * *

[10] Q. Am I correct, then, in thinking that from September of '66, until Dr. Safern came in 1972, that you were the only industrial psychologist employed at the EEOC? A. Employed in that capacity?

Q. Yes, sir. A. Correct.

---

* Numbers appearing in brackets in text indicate page numbers of original stenographic transcript of testimony.

Q. You were the only one with industrial psychology responsibilities at the Commission? A. That is correct. I would not say for a fact that there was no one else working for the entire agency that might have been trained in some aspect of industrial psychology. I was the only one working in that capacity.

Q. Would it be correct to say that Dr. Sharf, Dr. Taylor and yourself are now the three individuals with industrial psychological duties at the Commission and Dr. Sharf and Dr. Taylor report to you in terms of line responsibility? A. That's correct.

Q. So at the present time you are the senior man at the EEOC with industrial psychological duties, when I say "man," I mean person? A. That's correct.

Q. And you have always been such since September of 1966? A. That's right.

* * * * * * *

[15] Q. Did you participate in the drafting of the 1970 Guidelines? A. Yes, I did.

* * * * * * *

[16] Q. Could you summarize for us, please, your role in drafting these 1970 Guidelines? A. Yes. When then-Chairman William Brown of the Commission expressed an interest in looking into the feasibility of updating our 1966 Guidelines, I think really it was more than just merely up-dating—it was really a major rehaul, I consulted the OFCC Order of 1968 which dealt with the subject of employee selection and testing and I also consulted with attorneys in our office of General Counsel. We took several parts of the 1968 Order—Larry, I believe it is 1968.

Q. That's correct. A. And because the work on our 1970 Guidelines actually began in 1969, I took the various parts that I thought would be useful, some attorneys in the General Counsel's office drafted some additional language.

I discussed this project with some people at the Office of Federal Contract Compliance and also discussed it with their Advisory Committee on Testing, which they then had in effect.

And we came up with some language that we thought would be appropriate. And I might mention that also we eventually discussed the content of these Guidelines rather extensively with at least one person in the Solicitor's Office at the Department of Labor.

And it was fundamentally agreed, I think, between the **[17]** OFCC and ourselves, that is, the Commission, that the document fundamentally would serve the purpose of both agencies, and as you probably know, a year later the OFCC issued a document, which is very similar to our Guidelines.

I believe also in that discussion I was involved in talking to one attorney with the Justice Department, as well.

Q. And it has been your responsibility since they were issued to participate in their administration? A. That is correct.

Q. Are you the senior person that the Commission looks to for interpretation of these Guidelines? A. Well, without being immodest, I would say yes.

* * * * * * *

[19] Q. Am I correct in thinking that the EEOC, as a component part of the EEOCC, is presently engaged in revising the **[20]** Guidelines still another time? A. Well, that's one of the charges that the Coordinating Council—and if you don't mind, I will use that term from now on—the Equal Employment Opportunity Coordinating Council—it is too long to repeat every time. I don't want to say "EEOCC," it may be confused with this Commission. It is true the Council has taken upon itself, as one of its jobs, the production of what would be called a set of uniform guidelines on employee selection.

* * * * * * *

Q. What is the Coordinating Council? A. The Coordinating Council is composed of the heads of five Federal agencies. The Coordinating Council was established under the Equal Employment Opportunity Act of 1972. The agencies that are represented on the Council are the Equal Employment Opportunity

Commission, the U. S. Civil Service Commission, the U. S. Department of Labor, the U. S. Department [21] of Justice, and the U. S. Commission on Civil Rights.

* * * * * * *

Q. When, then, is the purpose of drafting a new set of guidelines? A. The purpose of the drafting—as I understand it—because you must understand I did not make the decision to engage in the writing of the guidelines, nor was I consulted in making of that decision.

As I understand it, though, the purpose is to have a uniform set of guidelines that the entire Federal Government could embrace. So that there will not be continual arguments, let's say, over what set of standards would apply to the various jurisdictions where employment discrimination might be a problem.

* * * * * * *

[23] Q. Am I correct in thinking that the 1966 APA Standards are incorporated by reference in the 1970 Guidelines? A. In the 1970 Guidelines?

Q. Yes. A. Yes, they are mentioned by reference, yes.

* * * * * * *

[24] Q. Now, those APA Standards have been revised this year, have they not? A. Yes, there are some new Standards.

Q. Would the Commission now follow the new Standards or continue to follow the old ones? A. Well, inasmuch as the 1966 Standards are referenced in the Guidelines, I would assume that the Commission would follow the 1966 Standards. Insofar as the administration of the 1970 Guidelines is concerned.

Q. Would it be fair to state that the 1974 APA Standards represent the more modern thinking of the APA and their profession? A. Well, I suppose by definition that they issued eight years later.

Q. Why would the EEOC not conform to this more modern thinking on the part of the Profession? A. Well, you must

recognize that the new Standards as published by the APA were only issued this year.

Q. The wheels of Government grind slowly? A. Possibly you can look at it that way.

Q. Do the 1966 APA Standards contain any preference for any one type of validation over another? A. I don't know that you could say that they state any real preference except that perhaps the Standards may refer to **[25]** appropriateness of various types of validity. I mean for what types of evidence are necessary to demonstrate certain types of validity, let's put it that way.

Q. In other words, they state the circumstances under which various types of validity are appropriate? A. Can be justified.

Q. Yes. But they don't—assuming that the appropriate circumstances are present they don't state any preference for one type of validity over another. A. Those Standards do not say a particular type of validity must be demonstrated in a particular type of context.

* * * * * * *

[36] Q. Dr. Enneis, did the EEOC Guidelines of 1970 permit the use solely of content validity in some conditions even though it might have been possible to do criterion-related validity studies? A. Yes. In fact I can quote or cite to you part of the 1970 Guidelines, where it says that. This is in Section 1607.5, entitled "Minimum Standards for Validation."

* * * * * * *

[37] A. "Evidence of content validity alone may be acceptable for well-developed tests that consist of suitable samples of the essential knowledge, skills or behaviors composing the job in question. The types of knowledge, skills or behaviors contemplated here do not include those which can be acquired in a brief orientation to the job."

* * * * * * *

[49] A. "Can you comply with the 1970 Guidelines and be in **[50]** non-compliance with the APA Standards?" Answer: "I think it would be possible to comply with our Guidelines and not be in compliance with the APA Standards, absolutely." That's it.

Q. Would you agree with that statement today? A. I think that it is possible certainly for an employer to meet the technical requirements of our Guidelines and certainly not meet all of the requirements that are in the APA Standards.

* * * * * * *

[62] Q. Is it possible, then, to be in general compliance with the Guidelines without meeting all of the technical requirements of the 1970 Guidelines? A. Yes.

* * * * * * *

[64] Q. How many criterion-related validation studies have you seen in the course of your duties with the EEOC that in your opinion meet all the requirements of the 1970 Guidelines? A. How many criterion-related validation studies?

Q. Yes. **[65]** A. Well, again, I want to say that it's not necessary for an employer to meet, under certain circumstances, all of the requirements. I have seen maybe two studies that do, I guess, really meet all of the requirements. Even though it may not have been necessary in those particular circumstances to have met all of the requirements, certainly I would say that certanly there were two studies that I would say met all of the requirements, every single one.

Q. All right. In the case of those two studies, did they also meet all of the requirements of the referenced 1966 Standards? A. I am going to take that back. I've actually seen three studies that I would be willing to state. I have forgotten one of them, and there may be more. I'm just trying to do a mental recollection here.

Q. Of the three studies—— A. Now, that's with respect to criterion-related——

Q. I understand. Of the three studies, did each of them also meet all of the referenced requirements of the 1966 APA Standards? A. I don't know.

Q. It is possible, then, that none of the three met the 1966 APA Standards as well? A. Again, I don't know.

Q. You are just not in a position to state one way or **[66]** another? A. Correct.

Q. What were the nature of the jobs for which these three jobs were conducted, if you recall, sir? A. One of them was for sales positions. One of them was for transit operators, actually a bus driver to be more specific, another one was for a police officer.

Q. What were the criterion used in each instance? A. Well, in one it was supervisory evaluations.

Q. Which one was that? A. That was for sales positions. And in terms of the bus driver, there were many, many criteria that were used but among those were punctuality, attendance, turnover, supervisory ratings, accident rating.

Among the patrol officers, again, supervisory ratings and some other objective criteria. And also complaints, that is, commendations and complaints in terms of the police officers. I am not saying that is all of the criteria. That is all I can remember right now.

* * * * * * *

[99] Q. Am I correct in thinking if there is no adverse impact of substantial nature then the requirement for job validation is not even triggered by the 1970 Guidelines? A. You mean test validation?

Q. Yes. A. That's correct.

* * * * * * *

[107] Q. Has EEOC taken the position certain inherent tests were discriminatory and should not be used? A. No.

Q. That would include the Wunderlich test, would it not? **[108]** A. That is correct. I would say I certainly have never taken that position. An individual employee of the EEOC might have said that to somebody at one time. That is not official Commission policy.

* * * * * * *

[131] Q. Are you aware of the cost of performing any of the studies which you have identified as being acceptable under the 1970 Guidelines? **[132]** A. Only one of them.

Q. Which one was that? A. That was—I'll tell him in the general sense what job category it was. I can say it about the job category. It was the one concerning validation of selection procedures of the bus drivers.

Q. How many employees did it involve? A. Several hundred.

Q. And what was the cost of performing that study? A. It was conducted in several different locations.

Q. What was the cost of performing that study? A. This was a nation-wide study and it—I think it ran in the neighborhood of $400,000.

Q. When was it done? A. When was it done?

Q. Yes, sir. A. It was—it's still going on.

Q. When was it sufficiently final for you to have reviewed the study and found it acceptable? A. Oh, about maybe approximately a year ago. When I say it is still going on, that is there is some follow-up being collected. But I think basically and fundamentally it is closed now.

* * * * * * *