DOCUMENT RESUME

ED 124 570 TM 005 323

AUTHOR Willens, Howard P.; And Others
TITLE Walter E. Washington, et al., Petitioners, v. Alfred E. Davis, et al., Respondents. Supreme Court of the United States, No. 74-1492, on Writ of Certiorari to the United States Court of Appeals for the District of Columbia Circuit; Motion for Leave to File Brief as Amicus Curiae and Brief Amicus Curiae for Educational Testing Service.
PUB DATE 29 Dec 75
NOTE 68p.

EDRS PRICE MF-$0.83 HC-$3.50 Plus Postage.
DESCRIPTORS Aptitude Tests; Court Cases; Cutting Scores; Employment Qualifications; Job Skills; *Personnel Selection; Police; Racial Discrimination; *Standardized Tests; *Supreme Court Litigation; *Test Bias; Test Interpretation; *Test Validity; Verbal Tests
IDENTIFIERS Educational Testing Service; *Washington v Davis

ABSTRACT

This brief, filed by Educational Testing Service (ETS), a non-profit educational corporation, deals with whether Test 21, a test of verbal ability developed by the U.S. Civil Service Commission and used by the Metropolitan Police Department of the District of Columbia to select police recruits, is sufficiently job-related to be legally permissible in selecting potential employees. As friend of the court, ETS presents a framework that utilizes technical evidence of test validity in an appropriate balance with the analytical judgments about fairness and rationality needed to implement properly the concept of job-relatedness under Title VII of the Civil Rights Act of 1964. (RC)

IN THE

Supreme Court of the United States

OCTOBER TERM, 1975

No. 74-1492

WALTER E. WASHINGTON, *et al.*,
Petitioners,
v.
ALFRED E. DAVIS, *et al.*,
Respondents.

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

MOTION FOR LEAVE TO FILE BRIEF AS AMICUS CURIAE

AND

BRIEF AMICUS CURIAE FOR EDUCATIONAL TESTING SERVICE

HOWARD P. WILLENS
DEANNE C. SIEMER
THEODORE S. SIMS
1666 K Street, N.W.
Washington, D. C. 20006
Attorneys for Amicus Curiae Educational Testing Service

Of Counsel:

JOHN S. KRAMER
General Counsel
Educational Testing Service
Princeton, New Jersey

WILMER, CUTLER & PICKERING
1666 K Street, N.W.
Washington, D. C. 20006

WILSON EPES PRINTING CO INC - RE 7-6002 WASHINGTON

INDEX

INDEX—Continued

INDEX—Continued

TABLE OF AUTHORITIES

CASES: Page

TABLE OF AUTHORITIES—Continued

TABLE OF AUTHORITIES—Continued

OTHER REFERENCES: Page

TABLE OF AUTHORITIES—Continued

IN THE

Supreme Court of the United States

OCTOBER TERM, 1975

No. 74-1492

WALTER E. WASHINGTON, *et al.*,
Petitioners,

v.

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 EDUCATIONAL TESTING SERVICE FOR LEAVE TO FILE BRIEF AS AMICUS CURIAE

Educational Testing Service (hereinafter "ETS") respectfully moves, pursuant to Rule 43(3) of the Rules of this Court, for leave to file the attached brief *amicus curiae* in this case.

Petitioners and the United States have consented to the filing of this brief by letters dated November 14, 1975 and November 25, 1975, respectively. Respondents, by letter dated December 4, 1975, refused to consent. ETS does not urge upon this Court the disposition sought by either petitioners or respondents. Therefore,

this motion and the attached brief, submitted before the date for completion of briefing by the parties, are timely presented to the Court. *See* Rule 42(3).

ETS was chartered in 1947 under the Education Law of the State of New York as a private, non-profit educational corporation.[1] ETS was organized by three major educational organizations then actively involved in testing, the American Council on Education, the College Entrance Examination Board, and the Carnegie Foundation for the Advancement of Teaching. These organizations combined their existing testing functions to create a single organization geared to meeting the increasing measurement needs in various fields of education.

The current program of ETS reflects both its initial charter and the substantial growth in the testing and measurement fields over the intervening decades. ETS develops and administers a wide variety of tests for use in employment selection and professional certification and licensing, as well as for use in education. In the educational field, ETS tests cover a wide variety of academic programs for elementary, secondary, college, and graduate levels.[2] In the certification and licensing field, ETS has developed nearly 30 different standardized examina-

---

[1] The principal offices of ETS are located in Princeton, New Jersey, and it maintains regional offices in Atlanta, Georgia; Berkeley, California; Los Angeles, California; Evanston, Illinois; Austin, Texas; San Juan, Puerto Rico; and Washington, D.C.

[2] For example, ETS developed and administers the Cooperative Tests for elementary and secondary school students, the College Board Tests (including the Scholastic Aptitude Test) on behalf of the College Entrance Examination Board, and the Graduate Record Examinations, Law School Admissions Test, and Admission Test for Graduate Study in Business, on behalf of the policy boards representing the respective graduate and professional schools, for prospective graduate students.

tions for professional and other occupational fields.[3] Included among these are two tests developed for use in selecting policemen.[4] ETS also offers advisory and instructional programs for educators, test users and the general public.

ETS has had one limited contact with this case in the proceedings below. Two current ETS employees, Dr. Albert P. Maslow and Dr. David M. Nolan, furnished affidavits that were submitted and relied on by petitioners in the District Court. At the time the affidavits were submitted, Dr. Nolan was Director of the ETS Washington Office, and Dr. Maslow was employed by the United States Civil Service Commission.[5] These affidavits expressed professional judgments on the validity study submitted by petitioners, principally as to the usefulness

---

[3] These professional fields include actuarial work, architecture, engineering, law, nursing, obstetrics, ophthalmology, podiatry, radiology, respiratory therapy, social work and teaching. The occupational fields include automobile repair, laboratory technology, banking, firefighting, foreign service, hospital financial management, merchant marine officer's work, insurance, furniture warehousing, stock brokerage, electrical contracting, optical work, and real estate sales and brokerage.

[4] One was recently developed at the request of the Department of Personnel of the City of Philadelphia. The test measures specific cognitive abilities that were related to performance of police work through job analyses by both measurement and police professionals. The test has been both content and concurrently validated and ETS has recommended further validation during operational administration. A second test was developed at the request of the International Association of Chiefs of Police for use by local police departments around the country. The test measures specific intellectual abilities that have been related to police work through job analyses by both measurement and police professionals and has been the subject of content validation as well as concurrent validation in four different locales. Neither of these tests is similar to the test at issue in this case. However, the analysis used by this Court to determine the "job-relatedness" of this test might affect the use of the police selection tests developed by ETS.

[5] Dr Maslow has since joined the ETS staff as Director of the Center for Occupational and Professional Assessment in Princeton, N.J.

of training success as a criterion for test validation and whether, in light of the findings of the validity study, the test at issue discriminates against blacks. These affidavits were made by the affiants in their individual capacities as measurement experts; Dr. Nolan did not submit his affidavit as a representative of ETS. ETS does not endorse the testimony of either affiant as reflecting its corporate position with respect to this case.

ETS' interest in this case arises out of its extensive experience in developing and validating standardized tests for employment selection and other purposes and its concern that the applicable legal standards be workable and fair. In particular, ETS is concerned with the analysis employed in determining whether Test 21, used by the Metropolitan Police Department of the District of Columbia in selecting among applicants for entry level positions, is "job-related." The requirement of job-relatedness is a legal standard of broad applicability to a variety of fact situations and, as such, it is important that considerable care be used in articulating its meaning and scope. ETS' interest in this legal issue is distinct from that of either of the parties, and the attached *amicus curiae* brief is limited to questions of law pertinent to this issue that may not be presented adequately by the parties.

As reflected in the papers filed below and by their initial brief to this Court, petitioners have an expansive view of the role of technical evidence of test validation in determining whether a particular test is job-related. They urge that a validity study of the test used by them confirms *both* the validity of the test under professional measurement standards *and* its job-relatedness under legal standards.

As indicated by their papers filed below, the respondents have a different but similarly expansive view of the

role of test validation. Respondents criticize the study offered by petitioners in numerous respects. Underlying their position is the view that technical validation data, using only one of several professionally accepted methods, constitute not only the preferred but virtually the only acceptable evidence of job-relatedness.

ETS has a different view—one which, perhaps paradoxically, assigns a less dispositive role to statistical validation evidence. ETS' experience in constructing, administering, scoring and evaluating a wide variety of tests related to employment licensing, certification and selection indicates that statistical evidence of test validity cannot be substituted for basic analytical judgments about fairness and rationality. ETS recognizes, however, that statistical and other evidence of test validity is important to both professional and legal judgments about the use of a particular test. ETS believes that this function can and should be defined with particularity. In the attached *amicus curiae* brief ETS presents a framework that utilizes technical evidence of test validity in an appropriate balance with the analytical judgments about fairness and rationality needed to implement properly the concept of job-relatedness under Title VII.

For the reasons stated above, ETS respectfully requests that this Motion for Leave to File Brief as Amicus Curiae be granted.

Respectfully submitted,

HOWARD P. WILLENS
DEANNE C. SIEMER
THEODORE S. SIMS
Attorneys for
Educational Testing Service

December 19, 1975.

IN THE

Supreme Court of the United States

OCTOBER TERM, 1975

No. 74-1492

WALTER E. WASHINGTON, *et al.*,
Petitioners,

v.

ALFRED E. DAVIS, *et al.*,
Respondents.

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

---

BRIEF AMICUS CURIAE FOR EDUCATIONAL TESTING SERVICE

QUESTIONS PRESENTED

Two questions are presented for review in this case. The first is whether a *prima facie* case of racial discrimination in the use of Test 21 has been established by the statistics presented by respondents. If so, the second question is whether Test 21 is sufficiently job-related to be legally permissible in selecting among potential employees.

This brief deals only with the second question. It does not discuss the holding of the Court of Appeals that the affirmative efforts of the Metropolitan Police Department of the District of Columbia (hereinafter the "Department") to recruit black officers and its success in so doing are irrelevant to respondents' showing that Test 21 had a disparate impact. ETS recognizes that this Court's disposition of the first question may render it unnecessary to reach the issue discussed in this brief.

STATEMENT OF THE CASE

The issue discussed in this brief arises on the following facts. Test 21, of which the record contains three versions, App. 209-31, 232-55, 256-78,[1] is a test of verbal ability developed by the United States Civil Service Commission. App. 187. It has been used by the Department in selecting applicants for many years and, for approximately 20 years, a score of 40 right answers out of 80 questions has been the minimum acceptable grade. App. 191. Whites achieve the requisite score on Test 21 in a greater proportion than blacks. App. 32-35.

A 1967 study conducted by David L. Futransky, a Civil Service Commission employee, in cooperation with the Department (the "Futransky Study"), inquired into the relationship between applicant scores on Test 21 and various aspects of subsequent performance. App. 99-109. It used a method referred to by psychologists as "criterion-related" validation. The criteria selected by Mr. Futransky were of two general types: recruit school performance criteria and on-the-job performance criteria. The only recruit school performance criterion available was the final average score achieved on eight subject-matter examinations. There were five on-the-job performance criteria available: supervisors' ratings, commendations, promotions, disciplinary proceedings and separa-

[1] References to the Appendix to Petitioners' Brief are designated "App."; those to the Appendix to the Petition for Writ of Certiorari are designated "CA."

tions. Mr. Futransky grouped commendations and promotions into a single criterion called positive incidents of performance; and also grouped low supervisory ratings, disciplinary proceedings and separations into a single criterion called negative incidents of performance. He performed the criterion-related validity study by comparing scores on Test 21 to each of these criteria to determine the statistical correlation with these various measures.

The Futransky Study suggested, with respect to white recruits, the existence of positive correlations between high scores on Test 21 and subsequent high job performance ratings, App. 106-07; positive correlations between scores on Test 21 and positive performance incidents, App. 109; negative correlations between scores on Test 21 and negative performance incidents, App. 108; and positive correlations between performance in the Department's training curriculum, as measured by final averages of scores on written examinations, and subsequent job performance ratings. App. 104-06. For black recruits, the Futransky Study did not disclose similar correlations. App. 104-09. Even for white recruits, the Futransky Study did not disclose, and it was not possible to infer from the study, whether the correlations for whites were based on sufficiently refined analysis to be meaningful, App. 181, or whether the job performance ratings used as a criterion themselves meaningfully reflected actual job performance. App. 50, 182, 207. Both courts below proceeded on the assumption that Test

---

[2] One obvious problem with the criterion of supervisors' ratings is that it included all positive ratings (+1 through +4) but excluded the 0 rating even though that rating represented "effective or competent" performance App 104. Inclusion of the 0 rating within an index of "successful" on-the-job performance might materially affect the statistical outcome of the study.

21 had not been shown by the Futransky Study to predict overall job performance. CA. 17a, 49a.[3]

The Futransky Study did establish, however, that high scorers on Test 21, both blacks and whites, were more likely to do well on written examinations administered during the course of the Department's training curriculum for new recruits (hereinafter "Recruit School"). At the time of the Futransky Study, Recruit School was of twelve weeks' duration. App. 102. However, the syllabus of the recruit training curriculum reflects that Recruit School had been expanded as of 1971 to seventeen weeks including two weeks of mid-curriculum on-the-job training. App. 112. Of the approximately 608 hours of total instruction during Recruit School, approximately 390 hours (or somewhat over 2 months) are devoted to classroom instruction. App. 112-13.

The two classroom-oriented portions of the curriculum are the "Police Operations Section" and "Laws and Documents Section." App. 114-15. Significant portions of the Police Operations Section are devoted to understanding the "Police Manual" (18 hours of classroom instruction) and to "Report Writing and Report Writing Review" (42 hours of classroom instruction). App. 122-31. Principal topics covered in the Laws and Documents portion of the curriculum include the District of Columbia Code (36 hours), Police Regulations (17 hours), Traffic Regulations (23 hours), Alcoholic Beverage Control Laws (11 hours), the Rules of Evidence (20 hours) and the Law of Arrest, Search and Seizure (40 hours). App. 135-55. Written examinations are administered

[3] The affidavits reflect some dispute, which neither court below regarded as pertinent, as to whether the Futransky Study established that Test 21 *did not* predict job performance ratings, positive performance incidents or negative performance incidents for black officers, App. 54-55, or whether it was simply inconclusive on this score. App. 181.

upon completion of instruction in each of these topics, App. 123, 131, 141, 144, 148, 155; *see* App. 186, and each trainee is required to score 75 percent or better. Those who fail an examination in a particular area are given such additional assistance or instruction as may be necessary for them to achieve a passing grade. App. 102. In this way, Recruit School failures are prevented.[4]

The Futransky Study examined the correlation between scores on Test 21 and final averages achieved during Recruit School. It found that, for both black and white recruits, those who scored higher on Test 21 were more likely to graduate from Recruit School with a final average of 85 or more.[5] The correlations were, from a statistical standpoint, significant for both blacks and whites. App. 190-91.

On these facts, respondents, intervening plaintiffs in the District Court, moved for partial summary judgment. In addition to an affidavit assessing the impact of Test 21, respondents' Motion was based on the affidavits of two testing professionals to the effect that scores on Test 21 did not correlate with ratings of overall job performance. App. 49-57. Petitioners responded with a Cross-Motion for Summary Judgment, based largely on affidavits of testing professionals to the effect that Test 21 had been validated as predicting relative ability to master the substance of the Department's Recruit School curriculum. App. 172-208. The District Court ruled in favor of petitioners. The Court of Appeals reversed.

[4] The Futransky Study, which observed without elucidation that "failure to complete Recruit School is, for all practical purposes, non-existent," App. 100, is the only evidence for this conclusion.

[5] App. 102-03. While one of respondents' experts suggested that the significance of the data disclosed by the Futransky Study was "confounded" by the absence of evidence of whether the final averages were, for recruits who had failed one or more of the written tests, based on the first or subsequent taking, the record discloses that they were in fact based on initial taking. App. 176, 181.

After holding that the respondents had made out a *prima facie* case of discrimination in the use of Test 21, it held that petitioners had not established the test to be "job-related," expressing "grave doubts" as to whether validation of an employment test against performance in a "training program" could ever satisfy the applicable legal requirements.

INTEREST OF AMICUS CURIAE

Educational Testing Service is a private, non-profit educational corporation engaged primarily in the development, administration, and scoring of standardized tests in many fields. ETS prepares tests to be used for employment selection purposes, including tests to be used for selection of policemen. The specific interest of ETS in this case is detailed in the accompanying Motion for Leave to File Brief As Amicus Curiae.

SUMMARY OF ARGUMENT

Analysis of the requirement of job-relatedness under Title VII of the Civil Rights Act of 1964, as amended, 42 U.S.C. § 2000e *et seq.*, has been confused by the failure to consider separately the two principal issues comprehended by this requirement and by the tendency to utilize technical evidence of test validation alone to measure compliance with basic legal requirements of fairness and rationality in classifications made for employment selection purposes.

Courts should consider first the issue of whether the knowledge, skills or other attributes used by an employer to classify prospective employees are fair and rational bases for selection decisions. Employers should be permitted to meet this threshold requirement of job-relatedness by demonstrating that the knowledge, skills or other attributes to be measured in prospective employees are a

fair sample of those required in overall job performance or that they are key elements of job performance, the absence of which will create a substantial risk of non-success on the job. If an employer elects to utilize key elements, that procedure should be permitted at each stage of a multi-stage employment process, such as one utilizing a training program, in the same fashion as it is permitted in a single-stage employment process. These judgments are largely independent of any test validation process, and should be made by assessing all the relevant evidence under traditional judicial standards of reasonableness and fairness.

After determining that a measure of job performance or selected key elements may properly be used as a basis for employment selection, the courts should then turn to the separate question of the fairness and rationality of any standardized tests used. Evidence of test validity may be useful in making these judgments, although there is no single measurement of test validity or preferred evidence by which the fairness and rationality of the use of a particular test can be assessed under all circumstances. The legal standards for determining the appropriateness of test use should avoid prescription of any one method of test validation and should be stated so as to permit the development of new techniques and appropriate combination of established techniques.

Such an analysis of the job-relatedness requirement is fully consistent with this Court's decisions and applicable legal and professional standards. Both the District Court and the Court of Appeals in this case failed to separate the two distinct issues involved in the determination of job-relatedness. Both courts over-generalized with respect to the applicable legal standards and failed to consider important evidentiary links in the chain of proof establishing job-relatedness in a multi-stage em-

ployment selection process. Further elucidation of the job-relatedness requirement by this Court is both desirable and appropriate on the basis of the record in this case.

ARGUMENT

I. THE ANALYSIS OF JOB-RELATEDNESS SHOULD DISTINGUISH BETWEEN THE APPROPRIATENESS OF THE KNOWLEDGE, SKILLS OR OTHER ATTRIBUTES USED BY AN EMPLOYER TO CLASSIFY PROSPECTIVE EMPLOYEES AND THE APPROPRIATENESS OF ANY STANDARDIZED TEST USED TO MEASURE THOSE ATTRIBUTES.

This case provides an opportunity for this Court to elaborate further the requirement of "job-relatedness" established in *Griggs* v. *Duke Power Co.*, 401 U.S. 424 (1971), and *Albemarle Paper Co.* v. *Moody*, 422 U.S. 405 (1975). Such elaboration is necessary to establish an analytical framework within which specific problems raised by individual cases can be coherently and consistently resolved.

Neither *Griggs* nor *Albemarle* required a detailed inquiry into the substance of the job-relatedness requirement. In *Griggs*, the employer used two verbal aptitude tests to select employees for the operating divisions of a steam generating plant. The only evidence offered by the employer to prove that the tests were work-related was a vague managerial judgment that greater verbal ability "generally would improve the overall quality of the work force." 401 U.S. at 431. This Court held that Title VII prohibits the use of "an employment practice which operates to exclude Negroes [and which] cannot be shown to be related to job performance." *Id.* Since there was no showing of any relationship of verbal ability to job performance, the Court needed to examine the job-relatedness requirement no further to conclude that, under the

circumstances disclosed by the record, the use of these verbal aptitude tests was improper under Title VII.

In *Albemarle*, the employer used general verbal aptitude tests in selecting applicants for employment in the operating divisions of a paper mill. After the decision in *Griggs*, the employer engaged an industrial psychologist who attempted to develop evidence of job-relatedness. The resultant study compared scores achieved on the tests by incumbent employees with supervisors' ratings of those employees as to overall on-the-job performance. The theory was that if high verbal aptitude scores correlated well with high supervisory ratings, the necessary job-relatedness would be established whether or not any particular level of verbal aptitude was actually required for the job. This Court found the validation study to be defective in a number of basic respects and concluded that, since the minimal relationship sought to be proved could not be supported by the study, and no other evidence was offered, the requirement of job-relatedness had not been met. On the facts in *Albemarle*, the Court was not required to examine the nature of the relationship sought to be proved or to consider evidence other than the test validation study.

As increasingly complex cases in the employment testing field come before the courts, it becomes highly desirable for this Court to articulate further the standards for determining whether the basic "job-relatedness" requirement has been met. Properly construed, the determination whether a particular test challenged under Title VII is "job-related" requires answers to two basic questions: (1) Has the employer selected an attribute or skill to be measured which is reasonably related to likely success (or lack of success) on the job? and (2) If so, is the test at issue properly designed and used to measure the particular attribute selected by the employer? Examination of these two questions will assist in sharp-

ening the relevant issues and in defining appropriate roles for the test validation expert and the court in making the required "job-relatedness" determination. Such an analysis is not only consistent with *Griggs* and *Albemarle* but is also supported by the relevant professional guidelines set out by the American Psychological Association[6] and the relevant legal guidelines set out by the Equal Employment Opportunity Commission.[7]

A. Employers Should Be Permitted to Meet the Threshold Requirement of Job-Relatedness by Demonstrating a Legitimate Interest in One or More Key Areas of Knowledge, Skill or Other Attributes.

The requirement of job-relatedness as first articulated by the courts understandably described job performance in general terms. Many employers assess their employees with respect to their overall competence in performing the tasks required by the job, frequently by reference to the observations of supervisors. A number of lower courts, in describing the job-relatedness requirement under Title VII, have used the phrase "overall job performance" as the standard to which any selection technique *must* be related.[8] Although it is plain that a

---

[6] AMERICAN PSYCHOLOGICAL ASSOCIATION, STANDARDS FOR EDUCATIONAL AND PSYCHOLOGICAL TESTS (1974) (hereinafter the "APA STANDARDS"). The APA STANDARDS are a revision of the STANDARDS FOR EDUCATIONAL AND PSYCHOLOGICAL TESTS AND MANUALS (1966), referred to at 29 C.F.R. § 1607.5(a).

[7] Equal Employment Opportunity Commission, "Guidelines on Employee Selection Procedures," 29 C.F.R. § 1607 *et seq.* (1971) (hereinafter the "EEOC Guidelines.").

[8] *See* United States v. City of Chicago, 385 F. Supp. 543, 555-56 (N.D. Ill. 1974); Vulcan Soc'y of N. Y. Fire Dept. v. C.S.C., 360 F. Supp. 1265, 1272-74 (S.D.N.Y.), *aff'd in relevant part*, 490 F.2d 387 (2d Cir. 1973). *Cf.* Smith v. City of E. Cleveland, 363 F. Supp. 1131, 1148 (N.D. Ohio 1973), *rev'd on other grounds sub nom.* Smith v. Troyan, 520 F.2d 492 (6th Cir. 1975). Many cases did not

measure of overall performance may provide an acceptable basis for establishing the necessary job-relatedness, any suggestion that such a measure is indispensable to establish job-relatedness should be rejected by this Court. The legitimate interests of employers and the existing legal precedents demonstrate that alternative means, relying on key elements of job performance rather than on overall assessments of success, are equally permissible and desirable under Title VII.

1. *There is a legitimate interest in focusing on one or more key elements of job performance.*

An employer's principal legitimate purpose in classifying persons at any point during the process of employment selection is to maximize the chances of selecting those persons who are likely to be successful at their jobs and to minimize the risk of selecting those who are not.[*] In many employment situations, attempting to maximize the chances of selecting successful employees involves a different, and more difficult, assessment than minimizing the risks of selecting those who are likely to be unsuccessful. Because of the difficulties in defining overall success, employers frequently find it useful to identify one or more *key* elements of job proficiency the absence of

---

reach this question simply because there was little or no evidence of job-relatedness at all. *E.g.*, Bridgeport Guardians, Inc. v. Members of Bridgeport C.S.C., 354 F. Supp. 778, 790-93 (D. Conn.), *aff'd in relevant part*, 482 F.2d 1333 (2d Cir. 1973); Chance v. Bd. of Examiners, 330 F. Supp. 203, 216-19 (S.D.N.Y. 1971), *aff'd*, 458 F.2d 1167, 1174 (2d Cir. 1972); Kirkland v. N.Y. State D.C.S., 374 F. Supp. 1361, 1372-76 (S.D.N.Y. 1974), *aff'd* 520 F.2d 420 (2d Cir. 1975).

[*] A classification can be fair if it provides some reasonable level of certainty that, as the standard is applied to a group of prospective candidates, the decisions made taken as a whole will serve legitimate employer objectives. No classification can ensure that *every* individual decision using the criterion will be correct in terms of the legitimate purpose.

which increases the risk that a prospective employee will not perform successfully.

Successful overall job performance is frequently difficult to define. Various combinations of individual strengths can overcome various combinations of individual weaknesses. Thus, it is difficult to agree on some meaningful measure of overall job success even for jobs requiring fairly simple skills:

> "For example, the success of production workers might be gauged in terms of the number of units of work turned out per day, the proportion of completed units that pass inspection, wastage, and maintenance of tools and equipment. Clerks might be rated by their superiors in terms of quantity of work, accuracy, and initiative. . . . [I]n most cases no single criterion can be taken to give a complete description of job success because each of them measures some important and pertinent phase of performance." [10]

Even for relatively simple jobs, the variety of ways in which proficiency is defined may depend on different and, more importantly, unrelated characteristics:

> "The fast typist may not necessarily be the most accurate, and the bus operator who has the fewest accidents may not be the one who best maintains his schedule. So questions arise as to whether such unrelated kinds of performance can be meaningfully

[10] *E.g.*, E. GHISELLI, THE VALIDITY OF OCCUPATIONAL APTITUDE TESTS 22-23 (1966) (hereinafter cited as "GHISELLI"). *See also* L CRONBACH, ESSENTIALS OF PSYCHOLOGICAL TESTING 413-16 (3rd ed 1971) (hereinafter cited as "CRONBACH"); A. ANASTASI, PSYCHOLOGICAL TESTING 417 (3rd ed. 1968) (hereinafter cited as "ANASTASI"); R. THORNDIKE & E. HAGEN, MEASUREMENT AND EVALUATION IN PSYCHOLOGY AND EDUCATION 234 (1955) (hereinafter cited as "THORNDIKE & HAGEN").

combined into a single over-all index of job success." [11]

It is even clearer for more complex jobs that:

> "Success is never unidimensional. Particularly in high-level positions, there are many patterns of success. Teachers, for example, may excel in different ways: one develops into a friend and counselor for youth, one stimulates independent thinking in the few brightest students, one overcomes the blockings that cause failure among weak students. To try to score these types of performance on a single scale is pointless—one loses information. . . ." [12]

Similar problems inhere in the implementation of measures of overall job success even when they are available. For example, supervisory ratings used to assess overall

---

[11] GHISELLI at 23. *See also* CRONBACH at 443-44.

[12] CRONBACH at 443. *See* Bridgeport Guardians, Inc. v. Members of Bridgeport C.S.C., 354 F. Supp. 778 (D. Conn.), *aff'd in relevant part*, 482 F.2d 1333 (2d Cir. 1973), where, in considering the job-relatedness of a test used to select police recruits, the Court observed that:

> "[W]ith jobs of more complexity, especially those requiring exercise of judgment, there is not likely to be agreement on which criteria are truly indicative of successful job performance, nor has the art of evaluation acquired such precision that measurements of job performance can always be readily or accurately made." 354 F. Supp. at 789.

See also Castro v. Beecher, 334 F. Supp. 930 (D. Mass. 1971), *aff'd in part and rev'd in part*, 459 F.2d 725 (1st Cir. 1972):

> "One might hazard a guess that reliable aptitude tests cannot be designed for candidates for the Presidency, Congress, the Supreme Court and many lesser public and private offices. The reasons include . . . the greater significance of integrity, strength of character, attractive personality, emotional stability, wisdom, judgment, sympathy, social imagination, intuition, common sense and good fortune." 334 F. Supp. at 942-43, n.9.

performance on many less complex jobs may frequently be biased and suspect.[1]

The risks of non-success, however, are easier to define conceptually and ultimately easier to measure. Almost every job involves one or more key elements—either particular knowledge, skills, abilities, personal characteristics, or other attributes—which, if absent in a prospective employee, increase substantially the risks of non-success. Examples are readily available: the prospective biology teacher who has an effective classroom manner, relates well to parents and keeps records efficiently but who lacks basic knowledge of biology; the prospective engineer who has the requisite knowledge of mathematics and engineering principles, is skilled in the handling of measurement and drafting tools, has the verbal ability to communicate with others, but who cannot read basic blueprints; the prospective bank teller who can do the computations required, operate the office machinery, and deal with the public but whose personality is such that gambling is an irresistibly attractive activity. In each of these cases, and in many others, the assessment of candidates in terms of an important dimension of work performance can assist an employer in identifying those for whom the risk of non-success is relatively high.[11]

---

[11] THORNDIKE & HAGEN at 234; CRONBACH at 571-74; Brito v. Zia Co., 478 F.2d 1200, 1206-07 (10th Cir. 1973); Vulcan Soc'y of N.Y. Fire Dept. v. C.S.C., 490 F.2d 387, 395 & n.10 (2d Cir. 1973).

[11] *See* Castro v. Beecher, 334 F. Supp. 930, 943 & n.9 (D. Mass. 1971), *aff'd in relevant part*, 459 F.2d 725 (1st Cir. 1972) ("It would not be astonishing to learn that the most that intelligence or aptitude tests can be counted upon to accomplish is to winnow out the obviously incompetent police candidates."); Pennsylvania v. O'Neill, 348 F. Supp. 1084, 1091 (E.D. Pa. 1972), *aff'd by an equally divided court*, 473 F.2d 1029 (3rd Cir. 1973) (*en banc*) ("[T]esting experts can reach judgments as to whether some applicants may be so obvi-

Tests may as a practical matter be more useful in assessing applicants in terms of particular characteristics for which an employer may have a legitimate concern. Tests can be related in a more meaningful way to particular work-related proficiencies [15] or aptitudes [16] than to overall job success. Ultimately it may be possible to combine a series of tests for particular job-related characteristics into batteries that may provide a more comprehensive appraisal of individual applicants.[17] But the development of an overall battery, for even a single job or class of jobs, may be an expensive and protracted process,[18] which even then has practical limitations which not all employers can afford to bear.[19]

Consequently, there is a practical need, particularly for employers that make hiring decisions frequently, to be able to appraise potential candidates in terms of

---

ously unqualified on the basis of their performance on the examination that their acceptance on the police force would not assist in the validation of a passing grade and/or would constitute an unacceptable risk.")

[15] *See, e.g.*, ANASTASI at 424-34.

[16] *See, e.g.*, CRONBACH at 417-21.

[17] *See* ANASTASI at 413-16; CRONBACH at 410-17. *See also* Buckner v. Goodyear Tire and Rubber Co., 339 F. Supp. 1108 (N.D Ala. 1972), *aff'd*, 476 F.2d 1287 (5th Cir. 1973).

[18] *E.g.*, CRONBACH at 414-21, 443-44.

[19] As one authority observed, in order to develop an appropriate test battery the U.S. Air Force;

> "went to the trouble of sending through training a 1300-man random sample of all eligible recruits, even though they knew in advance that the majority would fail. . . .
>
> "The Air Force took costly risks in sending an unselected group into training. Neither the employer nor the public wants to see unscreened men in responsible positions. The employer tied by seniority rules may properly refuse to hire low scorers that he will be unable to get rid of. All the costs, obvious and hidden, must be weighed in deciding on the proper scale and duration of a tryout." CRONBACH at 412.

employment-related characteristics more narrowly defined than overall job success. A classification of potential employees using rationally selected and defensible key elements of job performance serves the same legitimate purpose as a classification based on measures related to overall job performance and both methods should be equally acceptable under Title VII.

2. ***Use of key areas of knowledge, skill or other attributes as criteria for classifying employees is consistent with existing case precedent and guidelines.***

The foundation of this approach was recognized by this Court in *Albemarle Paper Co.* v. *Moody*, 422 U.S. 405, 433 (1975), when it faulted the employer's assessments of employee performance because there was no evidence that they were sufficiently related "to the Company's legitimate interest in *job-specific ability* . . ." (Emphasis added.)

The EEOC Guidelines also recognize the propriety of classifying potential employees on the basis of:

> "important elements of work behavior which comprise or are relevant to the job or jobs for which candidates are being evaluated." [20]

Elsewhere, the Guidelines explicitly provide that:

> "The work behaviors or other criteria of employee adequacy which the test is intended to predict or identify . . . may include measures other than actual work proficiency, such as training time, supervisory

[20] 29 C.F.R. § 1607.4(c). This provision of the EEOC Guidelines was endorsed by this Court in *Albemarle*, 422 U.S. at 431.

> ratings, regularity of attendance and tenure." 29 C.F.R. § 1607.5(b)(3).

It is reasonably clear that the Guidelines do not purport to exhaust the range of permissible work-related behaviors for which an employer may legitimately be concerned. Instead, they provide generally that:

> "Whatever criteria are used they must represent major or critical work behaviors as revealed by careful job analyses." 29 C.F.R. § 1607.5(b)(3).

Moreover, the Guidelines do not condition test use on the development of batteries designed to provide a comprehensive candidate evaluation. They provide only that a test must be a valid measure of "at least one relevant criterion" of work-related importance.[21]

The professional guidelines make a similar provision for selecting key elements:

> "Test users might define the performance domain . . . in terms of appropriately detailed and comprehensive job analyses . . . . The performance domain would need definition in terms of the objectives of measurement, restricted perhaps only to critical, most frequent, or prerequisite work behaviors." APA STANDARDS at 29.

This approach to the use of standardized tests is reflected in *United States* v. *North Carolina*, 400 F. Supp. 343 (E.D.N.C. 1975) (three-judge court), where the court considered a challenge to the use of tests in certifying prospective teachers. In that case the state had not

---

[21] 29 C.F.R § 1607.5(c)(1). The Guidelines caution, however, that where an employer uses "a single test as the sole selection device" use of the test "will be scrutinized closely when that test is valid against only one component of job performance." 29 C.F.R. § 1607.5(c)(1).

performed any job analysis for the job of teaching and had not attempted any assessment of candidates for certification that related to overall job performance. The court held that:

> "[I]t is beyond argument that the State of North Carolina has the right to [single out the key element of knowledge and] adopt academic requirements and written achievement tests designed and validated to disclose the minimum amount of knowledge necessary to effective teaching." [22]

The court distinguished between the employment selection and certification processes, pointing out that an employer may select only the best from among qualified candidates while a certifying authority must certify all those who are minimally competent. 400 F. Supp. at 350. However, the court pointed out that the requirement of a job-related criterion by which to make the classification is the same in both cases. 400 F. Supp. at 351 n.8.

The threshold question whether an employer has identified a sufficiently important work-related attribute is largely independent of technical questions of test validation. Indeed, the propriety of selecting employees on the basis of any particular attribute does not depend on whether that attribute is appraised by a test.

---

[22] 400 F. Supp. at 348; citing other cases in which particular knowledge had been singled out as the criterion on which a classification was based: Dent v. West Virginia, 129 U.S. 114 (1889) (medicine); Stephens v. Dennis, 293 F. Supp. 589 (N.D. Ala. 1968) (pharmacy); Lombardi v. Tauro, 470 F.2d 798 (1st Cir. 1972), *cert. denied*, 412 U.S. 919 (1973) (law); Graves v. Minnesota, 272 U.S. 425 (1926) (dentistry); Williamson v. Lee Optical Co., 348 U.S. 483 (1955) (optometry); England v. Louisiana Bd. of Medical Examiners, 246 F. Supp. 993 (E.D. La. 1965), *aff'd mem.*, 384 U.S. 885 (1966) (chiropractic medicine); and Milner v. Burson, 320 F. Supp. 706 (N.D. Ga. 1970) (driver training instruction).

This Court's reference to "job-specific ability" in *Albemarle* confirms that it is not possible to frame a comprehensive legal definition of what particular attributes are of legitimate concern for particular jobs. It is, moreover, reasonably clear that both the quantum and type of evidence necessary to identify the attribute will vary from situation to situation. Few would quarrel with the proposition that secretaries should know how to type or bank tellers how to add and subtract.[28] Equally few would dispute that social workers should have a basic knowledge of sociology, that nurses should have some essential understanding of relevant aspects of medical practice, or that air traffic controllers and airline pilots should be familiar with the pertinent FAA regulations. It may not be difficult for a prospective employer to demonstrate the propriety of such concerns by direct testimony of managers or supervisors familiar with the jobs in question.

In some contexts, however, establishing the legitimacy of an employer's concern for some work-related attributes may require more extensive evidence and more careful judicial review. This might be the case particularly with less concrete characteristics such as "thoroughness," "initiative" or "accuracy" and with less proficiency-oriented aspects of work behavior such as regularity of attendance and punctuality. In dealing with less concrete characteristics, demonstrating the relevance of the employer's concern might require use of careful job analyses carried out by professional industrial psychologists. *See* 29 C.F.R. § 1607.5(b)(3).

Express recognition of the necessity for case-by-case determination of this issue would be a useful elaboration of this Court's decisions in *Griggs* and *Albemarle*. By articulating the standards to be used in appraising an

[28] *See* APA STANDARDS at 29, 61.

employer's selection and use of key attributes in classifying employees, this Court would be recognizing the practical needs of the range of employers who are subject to the requirements of Title VII.

3. *The use of key elements related to job performance as classification criteria is permissible at each stage of multi-stage employment selection processes.*

Employers often must make employment selection decisions in stages, gradually separating out those who are unlikely to succeed. This may be required by the nature of the job, the relatively large investment to be made in new employees, or other considerations of employment efficiency. The legal standards with respect to fairness in employment selection should not work to the relative disadvantage of an employer who uses a multi-stage selection process. Such an employer should be permitted to demonstrate that each classification made is rationally related to his legitimate purpose in maximizing the chance of selecting successful employees or minimizing the risk of selecting unsuccessful employees. The use of one or more key areas of knowledge, skill, ability or other attributes should be subject to the same legal standards irrespective of the particular selection process used.

Multi-stage selection processes typically involve one or more training programs. Because of the specialized nature of the job, the employer may not be able to recruit potential employees who already have the precise attributes needed on the job.[24] If the absence of one or more

[24] A training program need not address every key element in a job. It can serve its legitimate purpose for the employer if it trains prospective employees in only one or several key elements of the job. There may be more effective ways to address the remaining key elements (not covered by the training program) such as completion of an academic program or on-the-job experience.

of these attributes is substantially related to the risk of non-success on the job, then the employer has a legitimate interest in constructing a training program to produce the desired attribute, and in selecting only those trainees who are likely to be able to acquire the desired attribute during the training program.[25]

Since a training program may be expensive to design and operate, an employer has a legitimate interest in obtaining the most effective use of the resources he has devoted to it.[26] The process of learning in training may require skills different than work on the job unless the job itself requires continual learning of new material. For this reason, the employer should be permitted to apply the key element analysis to the training program itself and to select skills or abilities without which there is a substantial risk of non-success in the training program. If the training program is itself fairly related to job performance, then the key elements needed to minimize the risk of non-success in the training program will necessarily be logically related to job performance.

Having narrowed his focus in this fashion, the employer should be permitted to set up a qualifying criterion or several qualifying criteria based on the key elements that relate to the risk of non-success in training.[27] This approach was recognized in *Griggs* when the Court found

---

[25] An important ancillary aspect of this analysis is that the method used to impart knowledge or skills in a training program be reasonable in light of the content of the training program. In so far as possible, the method should be no more academically oriented or abstract than is required by the job-related subject matter.

[26] If, in addition, training was protracted, an employer should be permitted to assert that the ability to complete training within the specified time was itself a work-related concern. 29 C.F.R. § 1607.5(b)(1).

[27] No one qualifying criterion need cover all key elements and typically different criteria such as academic course work, prior work experience, standardized tests, or demonstration of particular work skills will be used to measure the various key elements.

the employer had failed to sustain its burden of proof because neither test "was directed or intended to measure the *ability to learn* to perform a particular job or category of jobs." 401 U.S. at 428 (emphasis added). Qualifying criteria may properly be addressed to the ability to learn, if that learning process is reasonably necessary to job performance.

At each stage of such a multi-stage selection process, the employer must be able to demonstrate that the criteria utilized are related to legitimate employment objectives, either directly by reference to successful job performance or indirectly by reference to another stage in the selection process which itself bears the necessary relationship to job performance. However, if this requirement is satisfied at each stage of the process, there is no further overarching requirement with respect to the whole process.

The case law developed in the lower courts with respect to multi-stage selection processes is not inconsistent with this analysis.

Such an approach to evaluating training success as a work-related attribute is set out in *Buckner* v. *Goodyear Tire and Rubber Co.*, 339 F. Supp. 1108 (N.D. Ala. 1972), *aff'd*, 476 F.2d 1287 (5th Cir. 1973). In *Buckner* the court considered the job relatedness of a battery of employment tests used to select applicants for an industrial employer's four-year apprentice training program for skilled craft jobs. The court accepted evidence of the test battery's validity based on comparisons with training success. The court observed that accepting such evidence meant that:

> "[T]he relationship between the apprenticeship program and the filling of craft job vacancies then becomes also a matter for inquiry. . . ." 339 F. Supp. at 1114 n.4.

Thus, in a subsequent portion of its opinion the court examined the content of the training program to ascer-

tain whether it was "job-related." It found a number of the academic courses included in the training program to be insufficiently related to the job and directed their deletion from the training curriculum. It also directed removal from the battery of entrance tests those that had been validated only against success in the deleted courses. 339 F. Supp. at 1124. In doing so, the court cautioned, however, that it did not mean to imply that a training program is necessarily "subject to intensive validation studies such as outlined in 29 C.F.R. § 1607." 339 F. Supp. at 1124 n.19.

Other training program decisions to date involved a lack of evidence sufficient to determine whether the training program was intended to impart essential work-related knowledge or skills or that it was adequately designed to do so.[*] For example, in *Pennsylvania* v. *O'Neill*, 348 F. Supp. 1084 (E.D. Pa. 1972), *aff'd in relevant part by an equally divided court*, 473 F.2d 1029 (3d Cir. 1973) (*en banc*), the court considered evidence of job-relatedness based on a claim that an entrance test correlated with success in a police training program. In rejecting such evidence of job-relatedness, the court was influenced by an absence of the "showing of any correlation between success at the Police Academy and effective performance on the job." 348 F. Supp. at 1090. But it appears that the employer made no attempt to establish by competent evidence that knowledge of the relevant laws was essential for effective job performance. Similarly, there was no

[*] There are, of course, decisions which rejected training program validation, not on the ground that training success was an impermissible employer concern, but on the ground that the employer had not shown a test to predict training success. *See, e.g.*, Bridgeport Guardians, Inc. v. Members of Bridgeport C.S.C., 354 F. Supp. 778, 791 (D. Conn.), *aff'd in relevant part*, 482 F.2d 1333 (2d Cir. 1973); Officers for Justice v. C.S.C., 371 F. Supp. 1328, 1337 (N.D. Cal. 1973). *See also* Dozier v. Chupka, 395 F. Supp. 836, 845-46, 852-53 (S.D. Ohio 1975), in which one of two tests was simply not validated and the second was improperly combined with the results of a physical agility test for purposes of validation.

evidence that the training program curriculum was adequately designed to impart the essential knowledge or that passing grades on the training tests constituted a fair measure of mastery of the training curriculum. Indeed, the court pointed out that:

> "[T]he correlation made by [the employer's experts] may be misleading. There is no evidence as to the type of examination administered at the Academy. If in fact tests of the type contained on the entrance examination were used, a high correlation could be anticipated even if an individual had not mastered the training material." 348 F. Supp. at 1091.

The decision in *Harper* v. *Mayor and City Council of Baltimore*, 359 F. Supp. 1187 (D. Md.), *modified and aff'd sub nom. Harper* v. *Klosters*, 486 F.2d 1134 (4th Cir. 1973), involved an attempt to correlate scores on an entry test with success on a training program for fireman recruits. In *Harper* the court was confronted with the possibility that the tests used to assess mastery of the training curriculum were identical with the entrance test. 359 F. Supp. at 1202-03. However, there was, as in *Pennsylvania* v. *O'Neill*, no evidence that the recruit school was adequately designed to impart skills and knowledge which the fire department might legitimately consider to be work-related. Moreover, in rejecting the evidence of training school correlation, the court was influenced by the haphazard nature of the test itself.[29] Based upon its "own perusal of the quality of"

---

[29] The court observed that preparation of a number of the entry tests had been "the solitary work of a [Civil Service] Commission personnel technician" who "lacked professional training in test construction." He was, moreover:

> "as a Commission employee, unfamiliar with Fire Department work. Working principally on his own, he constructed a test that was used as an alternate to [another] exam from 1962 to 1969. . . . In 1971, the Commission employee was again given the job of constructing a Fire Department entry level test. This time he had the advice of a study panel which had been appointed by the fire board in an effort to lessen the

the exams in question, the court concluded "that if the tests were valid, it would only be by sheerest chance." 359 F. Supp. at 1187. Thus, the court never reached the question of the legitimacy of training success on the basis of a record demonstrating that the training curriculum was fairly designed to impart work-related knowledge.

One decision, *United States* v. *City of Chicago*, 385 F. Supp. 543, 556 (N.D. Ill. 1974), has suggested that employment selection tests must be related to actual job performance to the exclusion of any measure of training program success. The case involved an attempt to use training success for police recruits as an independent work-related consideration. But, as in *Pennsylvania* v. *O'Neill*, there was no evidence to establish either that the training curriculum was designed to impart knowledge essential to effective police performance or that acquisition of the relevant knowledge was fairly reflected in success on Police Academy exams.[30] Thus, the court's observation that overall job performance was the *only* legitimate employer concern was made in the context of a record that gave the court no opportunity to consider the independent relevance of training success.

---

racial impact of the entrance tests. Though he took their advice on many points it was not expert advice, and the actual choice of questions remained largely his own." 359 F. Supp. at 1197.

[30] The fact that the training examinations were themselves not in the record was suggested by the court's observation that "one would expect a correlation between performance on written exams," 385 F. Supp. 556, without considering, as in the *Harper* case, that the academy examinations might be essentially different from the employment selection tests.

B. Employers Should Be Permitted to Meet the Remaining Requirement of Job-Relatedness by Demonstrating that the Key Areas of Job Performance Selected as Criteria Are Measured by Valid and Appropriate Standardized Tests.

The first step in analyzing the fairness of an established selection process is to evaluate the legitimacy of the employer's concern for the knowledge, skill, ability or other attributes selected as a standard of likely success or non-success on the job. Once the key element has been identified and determined to be fairly related to job performance, there must be an evaluation of the means used to assess prospective employees in terms of the key element. The fairness of the measurement process is distinct from the legitimacy of the employer's concern for the key element to be measured, and should be separately considered.[31]

Standardized tests are only one measurement tool that employers can (and do) use. If particular knowledge is important, academic credentials are often used as a measure of its presence or absence. If prior work experience is central to an employer's objectives, references from former employers are often used as a measurement device. If personal attributes are important, interviews with the prospective employee are a likely means of assessment. The legal standards applicable to the measurement of key elements should be fair and workable for both test and non-test measurement devices.

---

[31] Plainly, a fair criterion can be unfairly applied. This would occur in a situation where possession of certain knowledge is a fair criterion, but only 30 percent of the questions on the test used to measure the criterion are devoted to this specified knowledge. Equally so, an unfair criterion can be fairly applied, as in a case where a police department establishes knowledge of Renaissance art as an entrance requirement, and the test adheres scrupulously to the specified subject.

When an employer has selected a standardized test as the measurement tool, courts have looked first to the technical evidence of test validation presented by the employer in determining the fairness of the measurement process. As the law has developed in the lower courts, there has been a tendency to rely almost exclusively on such evidence and to prefer evidence of criterion-related validity in making a judgment as to the appropriateness of a particular test use. Although evidence of test validation has an important role, ETS believes that there is no single measurement of test validity by which the fairness and rationality of a particular test use can be assessed in all circumstances. Judgments based on reason and all the relevant evidence are as appropriate and necessary where tests are used as a basis for the classification of individuals for employment purposes as in any other type of classification case which comes before this Court.

1. ***Evidence of validation is properly used to determine the measurement capability of a test.***

There are several professionally accepted techniques for ascertaining "*what* [a] test measures and *how well* it does so" [32] and confirming that "the test actually measures what it purports to measure." [33] These techniques are what measurement professionals refer to as validation. Of the methods reflected in the professional literature, three—"content," "criterion-related," and "construct" validation—have the widest degree of professional sanction. All three are specifically recognized by both the EEOC Guidelines, 29 C.F.R. § 1607.5(a), and the APA STANDARDS (at 25-31). Two—"content" and "cri-

[32] ANASTASI at 99 (emphasis in original); CRONBACH at 121-22.

[33] ANASTASI at 28; CRONBACH at 121-22; THORNDIKE & HAGEN at 108.

terion-related" validation—have the greatest practical relevance to the validation of tests for employment selection purposes given the current state of validation technology.[34]

Content validation is most commonly used with respect to proficiency or achievement tests "designed to measure how well an individual has mastered a specific skill or course of study."[35] This method of validation is an expert assessment of the design of the test to ensure that it adequately samples, for example, the topics of the course of study. Proper content validation requires the test developer to define the body of knowledge or catalogue the range of skills (commonly referred to as the "performance domain") to be assessed and to select the test questions so that, taken as a whole, they constitute a cross-

[34] "Construct" validity focuses "on a broader, more enduring, and more abstract kind of behavioral description" than either "content" or "criterion-related" validity. It is an attempt to ascertain "the extent to which [a] test may be said to measure a theoretical construct or trait." ANASTASI at 114-15; CRONBACH at 122-25. Thus, unlike "content" validity, which attempts to construct a test as a representative sample of some defined concrete area, or "criterion-related" validity which is the examination of correlations between test scores and specific external behaviors, "construct validity" is more concerned with ultimate traits such as "intelligence, mechanical comprehension, verbal fluency . . . neuroticism, and anxiety" reflected in scores on particular tests. ANASTASI at 111. It involves elements of both other forms of validation and holds perhaps the greatest interest for theoretical investigations into standardized tests.

Users of tests for employment selection purposes generally have more particularized needs and require more specific judgments regarding the measurement capability of a test than are available from construct validity under current measurement techniques. However, construct validity is an effort to combine reasoned judgments about fairness and rationality with statistical indices of correlation in a reliable framework for analysis and is ultimately likely to be more useful than either content or criterion-related validity standing alone.

[35] ANASTASI at 100; CRONBACH at 124-25; *see also* THORNDIKE & HAGEN at 109-10.

section or "representative sample" in both substance and difficulty of the defined performance domain.[36] For example, the validity of a test battery prepared to assess the outcomes of the training of medical school graduates would depend on the extent to which the battery sampled a cross-section of what, in the pooled judgment of experts, medical school curricula ought to impart.[37] In essence, content validity requires systematic examination both of "the test and the methods used in its preparation."[38]

Criterion-related validation is an inquiry, undertaken after a test has been formulated, as to whether it is possible "to infer from a test score an individual's most probable standing on some other variable called a criterion," and is most often used to validate aptitude tests. It is an attempt to ascertain the extent to which test performance correlates with performance in some external endeavor.[39] The process of criterion-related validation consists of gathering data on both test performance and criterion performance and, by a process of statistical analysis, as-

---

[36] ANASTASI at 100-02, CRONBACH at 122-23; THORNDIKE & HAGEN at 110-12. APA STANDARDS at 28.

[37] THORNDIKE & HAGEN at 111. *See* APA STANDARDS at 45-46.

[38] L. CRONBACH, ESSENTIALS OF PSYCHOLOGICAL TESTING 364 (2d ed. 1960). *See also* ANASTASI at 100; APA STANDARDS at 29.

[39] APA STANDARDS at 26; CRONBACH at 122, 126-27; ANASTASI at 105; *see also* THORNDIKE & HAGEN at 115-16.

Two generally recognized subcategories of criterion-related validation are predictive and concurrent validation. "Predictive validity" generally involves a comparison of the performance of the subjects of the study, first on the test and later on the criterion. "Concurrent validity" involves the correlation of the performance of the subjects of the study on the test and, more or less simultaneously, on the criterion. ANASTASI at 105-06; CRONBACH at 122; THORNDIKE & HAGEN at 115-16.

certaining whether the correlation between the two may be regarded as "significant." [40]

> 2. *There is no single method of evaluating test validity or preferred evidence by which the fairness and rationality of the use of a test can be assessed under all circumstances.*

The measurement profession's concerns for test validity are of almost equally long standing as the widespread use of standardized tests. The current APA STANDARDS have their roots in formal technical recommendations promulgated by that organization and others more than 20 years ago.[41] Textbooks and professional literature of a quarter-century and more ago reflect extensive inquiries into test validation.[42]

The purpose of the validation work done over the years is to provide information that will assist test developers to construct more useful tests and to assist test users in interpreting test scores, two principal measurement concerns. This validation work has not generally been concerned with developing standards that permit a decision as to *whether* a test may be used at all, the most

---

[40] "Significance" is used in the statistical sense, that is, the likelihood that the correlation could have occurred by chance. *See* n.49 *infra*.

[41] AMERICAN PSYCHOLOGICAL ASSOCIATION, TECHNICAL RECOMMENDATIONS FOR PSYCHOLOGICAL TESTS AND DIAGNOSTIC TECHNIQUES (1954); AMERICAN EDUCATIONAL RESEARCH ASSOCIATION & NATIONAL COUNCIL ON EDUCATIONAL MEASUREMENT, TECHNICAL RECOMMENDATIONS FOR ACHIEVEMENT TESTS (1955).

[42] *See, e.g.*, R. THORNDIKE, PERSONNEL SELECTION (1949); A. ANASTASI, PSYCHOLOGICAL TESTING, 120-51 (2d ed. 1954); F. FRIEMAN, THEORY AND PRACTICE OF PSYCHOLOGICAL TESTING, 26-41 (1955); THORNDIKE & HAGEN at 108-23, 256-60.

relevant legal concern.[10] To some extent, the interpretative assistance found in evidence of test validity assists in making the legal judgment. Such evidence may be necessary to the legal judgment; it should not be substituted for the legal judgment because it was not developed or intended for that purpose.

Yet a number of lower courts have been developing what appears to be an unrealistic reliance on evidence of test validity. Most of the lower court testing cases thus far have involved the use of aptitude tests. The APA STANDARDS reflect a preference, for measurement purposes, for criterion-related validation of aptitude tests. The EEOC Guidelines have adopted this preference for criterion-related validity as a legal standard for *all* types of tests. Section 1607.5(a) provides that employers may establish a test to be "job related" through the use of:

---

[9] As the APA STANDARDS recognize:

"Almost any test can be useful for some function and in some situations, but even the best test can have damaging consequences if used inappropriately. . . .

"This document is prepared as a technical guide for those within the sponsoring profession; 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 STANDARDS at 7-8.

[10] APA STANDARDS at 27-28. The distinction between aptitude tests and achievement tests is largely based on the use of a test rather than its intrinsic nature. A test of knowledge of chemistry, for example, might normally be used as an achievement test to measure how much chemistry the candidate had learned. However, the same test of chemistry might be used as an aptitude test to select candidates for admission to advanced placement or graduate study programs.

"Evidence of content or construct validity, as defined in [the APA STANDARDS], . . . where criterion-related validity is not feasible." [44]

This leap from a preference for criterion-related validity for measurement purposes with respect to some tests to a legal standard with respect to all tests has been reflected in some lower court decisions.[45] Such cases have usually involved aptitude tests and an evident lack of fairness and rationality such that the same result would probably have been reached under any analysis. Other courts, however, have been more sensitive to the need for flexibility in considering validation evidence.[46] Any

[44] The absence of full public scrutiny of and comment on the Guidelines before promulgation may have engendered some ambiguity. One such ambiguity involves the use in § 1607.5(a) of the term "feasible." The term "technically feasible," on the other hand, is used at a number of points in the Guidelines (*e.g.*, §§ 1607.4(a); 1607.5(b)(1)), and is defined with great rigor in § 1607.4(b). It is not altogether clear whether § 1607.5(a), in using the term "feasible," meant "technically feasible," as some courts have assumed to be the case. *E.g.*, Douglas v. Hampton, 512 F.2d 976, 986 (D.C. Cir. 1975); Harper v. Mayor and City Council of Baltimore, 359 F. Supp. 1187, 1202 (D. Md.), *modified and aff'd sub nom.* Harper v. Klosters, 486 F.2d 1134 (4th Cir. 1973). The indication in section 1607.5(a) itself that "evidence of content validity alone may be acceptable" suggests that some less rigorous connotation was intended by the use of "feasible" rather than "technically feasible."

[45] *E.g.*, Kirkland v. N.Y. State D.C.S., 520 F.2d 420, 426 (2d Cir. 1975); Rogers v. Int'l Paper Co., 510 F.2d 1340, 1349 (8th Cir. 1975); Chance v. Bd. of Examiners, 330 F. Supp. 203, 216 (S.D. N.Y. 1971), *aff'd*, 458 F.2d 1167, 1177 & n.16 (2d Cir. 1972); Harper v. Mayor and City Council of Baltimore, 359 F. Supp. 1187, 1201 (D. Md.), *modified and aff'd sub nom.* Harper v. Klosters, 486 F.2d 1134 (4th Cir. 1973); Fowler v. Schwarzwalder, 351 F. Supp. 721 (D. Minn. 1972); Pennsylvania v. O'Neill, 348 F. Supp. 1084, 1090-91 (E.D. Pa. 1972), *aff'd by an equally divided court*, 473 F.2d 1029 (3rd Cir. 1973) (*en banc*).

[46] *E.g.*, Vulcan Soc'y of N.Y. Fire Dept. v. C.S.C., 490 F.2d 387, 394-95 (2d Cir. 1973); Bridgeport Guardians, Inc. v. Members of Bridgeport C.S.C., 354 F. Supp. 778, 791 (D. Conn.), *aff'd in relevant part*, 482 F.2d 1333, 1337-38; Kirkland v. New York State D.C.S., 374 F. Supp. 1361 (S.D.N.Y. 1973), *rev'd in part on other*

standard which appears to rely exclusively on statistical evidence of criterion-related validity will be both illogical and unfair to employers who have used standardized tests in a rational and professionally accepted manner.

Criterion-related validity has inherent limitations as an exclusive basis for legal judgments about test fairness. To establish criterion-related validity it is necessary only to establish a moderately positive correlation between two sets of data for a given group of persons. One set of data is the score on a test; the other set of data is the "score" on some other performance criterion. The correlation range runs from +1 (perfect positive correlation where high scores on the test always coincide with high scores on the performance rating) to −1 (perfect negative correlation where high scores on the test always coincide with low scores on the performance rating).[48] If the correlation is 0.30 or above and the sample size is at least 45, then the measurement professional will conclude that criterion-related validity is established at a level that is statistically significant.[49]

---

grounds, 520 F 2d 420 (2d Cir. 1975); Officers for Justice v. C.S.C., 371 F. Supp. 1328, 1336 (N.D. Cal. 1973). *Cf.* Smith v. City of E. Cleveland, 363 F Supp. 1131, 1148-49 (N.D. Ohio 1973), *rev'd on other grounds sub nom.* Smith v. Troyan, 520 F.2d 492 (6th Cir. 1975); Coopersmith v. Roudebush, 517 F.2d 818, 824 (D.C. Cir. 1975).

[48] A correlation coefficient ranging within these limits is generally used in psychometric research. Other coefficients of correlation are occasionally used that have a different range.

[49] Statistical significance is ordinarily determined from mathematical tables and is stated in terms of a particular "level of confidence" A statement that a correlation is significant at the .05 level denotes that chances are no greater than five out of a hundred that the true correlation in the population is 0—in other words, the probability of such a correlation occurring by chance is no greater than five times in a hundred. R. FISCHER, STATISTICAL METHODS FOR RESEARCH WORKERS 209 (11th ed. 1950). Most psychometric research employs a 05 level of confidence. Occasionally a lower .10 level or a higher .01 level may be used.

But a closer look at such statistical evidence indicates its inherent limitations. First, for a correlation of 0.30 only 9% of the reason for the variation in scores on the performance criterion is attributable to whatever the standardized test measures. The remaining 91% is explained by some other factor. The 0.30 correlation does not provide any information about whether the remaining 91% is explained by factors more relevant than the factor measured by the standardized test or how vital the factor measured by the standardized test is for adequate job performance. Second, the 0.30 correlation does not help to make a judgment about what the standardized test actually measures or what the performance criterion actually measures. The standardized test may be titled "Verbal Aptitude" but it may in fact measure only spelling capability. The performance criterion may be labelled "supervisor's rating of overall job performance" but it may in fact measure only whether or not the supervisor likes the person rated. Therefore, the 0.30 correlation may be evidence only of the fact that good spelling correlates favorably with supervisors' personal preferences. These dangers of depending exclusively on correlational validity in evaluating a test have long been recognized by measurement experts.[50]

Plainly, for use as a legal standard of fairness and rationality, criterion-related validity should be supplemented by evidence to support at least two ancillary propositions: (1) what the standardized test actually measured; and (2) what the performance criterion actually measured. The actual subject matter tested by the test might be proved through a content validity study or, in the case of a simple test, through inspection.[51]

[50] *E.g.*, Gulliksen, *Intrinsic Validity*, 5 AMERICAN PSYCHOLOGIST 511, 514 (October, 1950).

[51] *Compare*, for example, decisions in which a court's judgment invalidating use of a test has been influenced by its own inspec-

Evidence with respect to the performance criterion might come from industrial psychologists or other measurement experts or from those experienced with the employment setting from which the performance data was collected. Once these fundamental requirements have been satisfied, there is still room for informed judgment as to whether the particular correlation reflected in the criterion-related validity study is appropriate in light of the employer's objective in using the test.

The use of criterion-related validity as "preferred evidence" also could lead to results that do not satisfy legal standards. For example, the "test" might consist of a series of multiple choice questions about family background and income level. The correlation between white middle class economic status and success on any given

---

tion of the test. Harper v. Mayor and City Council of Baltimore, 359 F. Supp. 1187, 1203 (D.Md.), *modified and aff'd sub nom.* Harper v. Klosters, 486 F.2d 1134 (4th Cir. 1973); Castro v. Beecher, 334 F. Supp. 930, 942 (D. Mass. 1971), *aff'd in relevant part*, 459 F.2d 725 (1st Cir. 1972); Bridgeport Guardians, Inc. v. Members of Bridgeport C.S.C., 354 F. Supp. 778, 791 (D. Conn.), *aff'd in relevant part*, 482 F.2d 1333 (2d Cir. 1973).

In making this determination, the nature of the performance criterion is particularly important. If the particular aptitude, skill or ability is of great importance, and the risk of non-success is high or the results of non-success extremely costly (either in social or monetary terms), then use of a test with a relatively low correlation might be rationally related to the legitimate purpose of minimizing that risk. On the other hand, where the risk is relatively low in terms of occurrence or cost, then the necessary rational relationship to an employer's legitimate purpose might require a higher correlation.

It is also important to consider the size of the applicant pool in relation to the number of candidates to be selected. If the size of the pool is large and the number of persons to be selected is small, then very low correlations may be practically useful because they provide greater efficiency than would a random selection of applicants.

employment performance criterion might be quite high.[1] But it would seem unfair to permit an employer to hire only white middle class persons on the basis of that statistical evidence of criterion-related validity. More subtly, a test may be labeled as an achievement or aptitude test with a specific subject matter content, but in fact be constructed in a way that measures highly esoteric material with which only a very narrow segment of the population would be familiar. Such a test might show statistical evidence of very high criterion-related validity but, again, should not be acceptable under a legal standard.

With these technical limitations in mind, ETS believes that the courts must look beyond the confines of measurement technology to make both of the fundamental judgments involved in the job-relatedness determination under Title VII. This is clearly so with respect to the employer's use of a key element of job performance as a selection criterion; evidence of criterion-related validity may assist in making this judgment, but cannot substitute for it. It may also be permissible, and under some circumstances it is obligatory, to look beyond the confines of measurement technology to make the second critical judgment whether the test fairly measures the key element of job performance. To do otherwise would result in an abdication of judicial responsibility and an excessive reliance on measurement technology which most professionals in the field would concede is not justified.

---

[1] In one much-cited study, generally positive and significant correlations were found between various measures of cranial capacity and grades received by freshman college students during their first semester. C. HALL, APTITUDE TESTING 134-36 (1928).

3. ***The legal standards for the measurement aspects of job-relatedness should permit development of new techniques and appropriate combination of recognized techniques.***

Properly used, standardized tests provide an objective measure that is open to inspection and easy to comprehend. From a standpoint of fairness, they are usually preferable to subjective measures because of their substantially higher reliability and potential for higher validity.[1] Therefore the legal standards applied to standardized tests should not be so much more stringent than those applied to subjective assessments that the use of tests is made economically or otherwise impractical.

Beyond these concerns is the need to state legal standards in a way that does not impede progress within the measurement profession or prevent employers from using new validation techniques as they are developed. Validation techniques have been revised and refined over the years. Various theories have obtained currency in the profession and then declined in acceptance as more persuasive theories were developed. The measurement profession has continued to search for more accurate ways to define what tests measure and describe how they measure it. "Experience teaches that the preferred method of today may be the rejected one of tomorrow." *Vulcan Society of N.Y. Fire Department* v. *C.S.C.*, 490 F.2d 387, 394 (2d Cir. 1973).

New validation methods are currently being introduced that use the judgments of large juries of persons who are experienced on the job about the propriety of the content of a test and the likely level of perform-

---

[1] Wilson, *A Second Look at Griggs v. Duke Power Company: Ruminations on Job Testing, Discrimination, and the Role of the Federal Courts*, 58 VA. L. REV. 844, 873 (1972); Note, *Employment Testing: the Aftermath of Griggs v. Duke Power Company*, 72 COLUM. L. REV. 900, 924 (1972).

ance on the test of a minimally qualified person. This method has evolved from theoretical discussions in the technical literature [35] and has been made possible by the availability of computers to manage and assess the data. It has substantial advantages over current techniques because the use of a large number of judgments eliminates the effect of personal bias that may be present even in experts, and the use of persons who are experienced on the job or in training persons for the job provides an additional touchstone of reality.

Such new developments would be adversely affected if the legal standards for job-relatedness are stated exclusively in terms of current validation techniques. A sounder approach would be for this Court to define the legal standards more generally in terms of the traditional requirement of rationality and fairness so that the necessary flexibility and room for experimentation can be preserved.

II. PROPER ANALYSIS OF THE JOB-RELATEDNESS ISSUE IN THIS CASE WILL FACILITATE ITS RESOLUTION AND PROVIDE NEEDED GUIDANCE FOR THE LOWER COURTS.

The *Davis* case is the first employment testing case to come before this Court which requires careful analysis of the requirement of "job-relatedness" in the context of a

---

[35] W. Crawford, Assessing Performance when the Stakes Are High, March, 1970 (Paper presented at the meeting of the American Educational Research Association, Minneapolis); J. Fremer, Criterion-Referenced Interpretations of Survey Achievement Tests, 1972 (Educational Testing Service, Test Development Memorandum 72-1), N. Luebke, A Practical Method of Determining a Criterion Score for Criterion-Referenced Measurement, February, 1973 (Paper presented at the meeting of the American Educational Research Association, New Orleans); K. Sparks, Memorandum on the Conference on the United States Department of State Foreign Service Examinations in French, Spanish, German, and Russian (July 12-14, 1957); Nedelsky, *Absolute Grading Standards for Objective Tests*, 14 EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT 319 (1954).

multi-stage employee selection process. Neither court below engaged in the necessary analysis; neither properly appraised the test validation evidence in the case; and neither resisted the temptation to offer excessively broad generalizations regarding the application of this Court's prior opinions to the case at hand. In many respects, full explication of the "job-relatedness" issues raised by the *Davis* case is handicapped by the limited record below and the summary posture of the case as it comes before this Court. Subject to this caveat, the *Davis* case provides an opportunity for this Court to establish the necessary parameters of the "job-relatedness" standard under Title VII.

A. Is Knowledge of Laws and Other Law Enforcement Subjects an Appropriate Key Attribute of Police Work which Can Be Used by the Department in Selecting Police Recruits?

Although neither court below addressed this central question in precisely these terms, their opinions appear to approach the issue from different perspectives and, not surprisingly, suggest different conclusions. Two basic questions are presented by *Davis*. First, does the knowledge identified by the Department bear the necessary relationship to successful police performance on the job? Second, is the Department's recruit training program reasonably designed to impart that knowledge? The District Court appears to have answered both questions in the affirmative; the Court of Appeals appears to have concluded that neither question need be addressed in the absence of convincing evidence validating either Test 21 or the recruit training program against overall performance on the job by policemen. Depending upon this Court's view of the adequacy of the evidentiary record, this Court could rule in favor of either petitioners or

respondents or remand the case for further proceedings below.[36]

1. *Knowledge as an appropriate key attribute of police work.*

In the absence of any correlation between performance on Test 21 and overall job performance,[37] the *Davis* case requires consideration of the propriety of identifying a key attribute—knowledge of certain law enforcement subjects—as a basis for selecting policemen.

Neither the Department's pleadings nor the proof below defined the knowledge considered by the Department to be a key element of effective police work. Nowhere is the risk assessed of the consequences to the public if policemen lack such knowledge. However, it may be reasonable to infer that the Department, based on its collective experience, has concluded that effective police work requires

---

[36] Because this case involves a governmental employer, the analysis used to determine whether the necessary rational relationships have been established will be the same regardless of which party has the burden of going forward. If this Court determines that adverse racial impact exists based on the statistical evidence offered by the respondents, then under Title VII the petitioners have the burden of going forward to establish that the test is being used properly. If the Court determines that no adverse racial impact exists, then under the Fifth and Fourteenth Amendments the respondents have the burden of going forward to establish that the test is being used improperly and that, accordingly, the classification based on its use is irrational.

[37] There appears to be no dispute that the Futransky Study disclosed no significant correlations between performance on Test 21 and overall job performance. App. 181, para. 8. However, it is appropriate to note that this conclusion does *not* support the proposition that a significant correlation between performance on Test 21 and overall job performance *cannot* be established. Futransky selected as a criterion of positive performance supervisors' ratings of +1 through +4. He did not include ratings of 0 although that rating means "effective or competent" performance. App. 105. If the 0 ratings had been included in the criterion against which Test 21 scores were compared, a positive correlation against job performance for all officers may have been established. *See* App. 181.

knowledge of (1) the laws and regulations to be enforced, and (2) certain rules, principles, and techniques with respect to law enforcement. It appears equally reasonable to infer that the risk of failure on the job in the absence of such knowledge—at least in terms of making improper arrests and failing to make proper arrests —would be substantial.

The District Court emphasized the "responsibilities and expertise required of modern police officers in a large metropolitan city . . .", and, after summarizing the syllabus of the recruit training course, concluded that the daily significance of police skills "demanding reasoning and verbal and literacy skills is borne out in the crucible of the criminal trial court." 348 F.Supp. at 17. Characterizing law enforcement as "a highly skilled professional service," the District Court obviously concluded that the knowledge taught in the training curriculum was sufficiently important and related to effective police work so as to permit its use by the Department in selecting among recruit candidates. On a very limited evidentiary record, in short, the District Court reasoned that it was not necessary to correlate performance on Test 21 with overall job performance in order to meet the requirement of "job-relatedness" under Title VII, and that the necessary relationship was evident.

The majority of the Court of Appeals disagreed. It emphasized that there were no correlations, at least for black officers, between recruit school performance and job-proficiency ratings. 512 F.2d at 963. The Court of Appeals, therefore, never considered the relationship emphasized by the District Court between the knowledge taught in the training program and the knowledge required for effective police work. Its failure to do so appears to reflect its view that "job-relatedness" can be demonstrated only on the basis of correlation with overall job performance, but it is possible that the Court of

Appeals merely concluded that the record in the *Davis* case could not support such a finding.[58]

Consistent with our analysis of the job-relatedness requirement, ETS submits that it would be entirely appropriate for the Department to identify knowledge of laws and enforcement rules as an important dimension of police performance on the basis of which its employees can be properly selected. The *presence* of such knowledge may not ensure successful police performance, which like most complex jobs requires a mix of personal and professional skills, but its *absence* probably will involve a substantial risk of poor police performance which the Department can properly seek to avoid. This is a determination as to which technical issues of test validation are largely irrelevant.[59] The principal question for this Court in considering this issue is whether the present record provides a sufficient basis for such a determination to be made.[60]

---

[58] Regardless of the Court's underlying rationale for not considering this question, its failure to do so led it to extensive further discussion of "trainability" and "training programs" which raise separate and equally important issues regarding the "job-relatedness" determination. *See* discussion *infra* at pp. 45-47.

[59] *See* Castro v. Beecher, 459 F.2d 725, 735 (1st Cir. 1972), in which the Court concluded that the requirement of a high school education was established to be job-related by the findings of the Report of the President's Commission on Law Enforcement and Administration of Justice, THE CHALLENGE OF CRIME IN A FREE SOCIETY (citing pp. 106-10 of the Report).

[60] ETS takes no position as to whether the Department has, on this record, adequately proved that certain knowledge is an appropriate key element of successful police work.

The Department has cited cases and other authorities for the proposition that a certain level of knowledge is important for successful police work. Petitioners' Brief at 18-22. The evidentiary problem is whether, if certain key elements of knowledge for successful police work *could* be identified, the Department can rely on the existence of this possibility whether or not, at the time the training program was established, the Department actually had identified such key elements of knowledge. Although not discussed

2. Adequacy and reasonableness of the recruit training program.

If the Court concludes that knowledge of particular subjects can properly be emphasized in the selection process by the Department, it is necessary to examine further whether the particular training program used by the Department is fairly and reasonably designed to impart the necessary knowledge.[61] The District Court, based upon its review of the training program syllabus, concluded that the program did teach subjects that were reasonably related to the Department's legitimate interest in policemen with the desired knowledge and expertise. 348 F. Supp. at 17. The Court of Appeals did not directly address the question. It did, however, indirectly cast doubt upon the Department's training program in the course of its discussion of the propriety of correlating any test with "trainability" or a "training program." In so doing, the Court of Appeals may have departed unnecessarily from the applicable precedents and the specific questions raised in this case.

The Court of Appeals seemed to suggest, in the absence of any correlation against overall job performance,

---

in these terms, the Department is plainly relying on the existence of a Recruit School syllabus containing subjects logically related to police work as the basis for an inference that at the time the Recruit School syllabus was devised, someone at the Department identified certain key elements of knowledge believed to be essential to successful police work.

[61] One test of reasonableness involves the design of the training program with respect to the nature of the subject matter knowledge to be imparted. An employer who has identified knowledge of certain mechanical assembly techniques as essential to successful job performance probably could not properly rely on a training program that utilized exclusively an abstract, academic or theoretical presentation of the subject.

that the only way in which the correlation between Test 21 scores and recruit school performance could be relevant was in terms of recruit "trainability." The Court of Appeals concluded, however, that Test 21 did not predict "trainability"; it stated:

> "[B]ecause of the departmental policy that nobody fail Recruit School, appellees have not shown that the admission of applicants who score below 40 on Test 21 into Recruit School would *necessitate expanded training time or produce Recruit School failures.*" 512 F.2d at 963 (emphasis added).

After holding on narrow grounds that the Futransky Study was inadequate, however, the Court stated further that the Department had not:

> "persuaded us that trainability could be a proper criterion for validating Test 21. . . . [W]e entertain grave doubts whether any of this type of evidence could be strengthened to the point of satisfying the heavy burden imposed by *Griggs*." 512 F.2d at 964-65.

Even more fundamentally, the Court of Appeals suggested that validation against a training program may be impermissible *per se*. According to the Court, the "ultimate issue in this controversy" is whether proof "that Test 21 is predictive of further progress" in the recruit training program "is an acceptable substitute for a demonstration of a direct relationship between performance on Test 21 and performance on the job." 512 F.2d at 962-63. These portions of the Court of Appeals' opinion deserve the careful attention of this Court.

First, the Court of Appeals may have failed to consider carefully the "job-relatedness" of "trainability" or success in training programs under the applicable precedents. The EEOC Guidelines provide that:

> "[W]ork behaviors or other criteria of employee adequacy which the test is intended to predict or iden-

> tify . . . may include measures other than actual work proficiency, such as training time. . . ." 29 C.F.R. § 1607.5(b)(3).

Further, in formulating the job-relatedness standard in *Griggs*, this Court observed that the fundamental flaw in the employer's proof was that neither test used in that case "was directed or intended to measure *the ability to learn* to perform a particular job or category of jobs." 401 U.S. at 428 (emphasis added). In *Davis* the Department has attempted to meet the very test set out in *Griggs*, by demonstrating that Test 21 *was* directed to or intended to measure "the ability to learn" certain subject matter necessary to perform the job of a policeman. The Court's finding that the Department had not proved that low scorers on Test 21 expanded training time or increased Recruit School failures may be consistent with the Guidelines and *Griggs*. The facts in the record do not compel that conclusion, however, and the competing inferences were never considered by either the District Court or the Court of Appeals.[42] In any event, the Court of Appeals should have defined its holding more

---

[42] The District Court did not address the question whether low scores on Test 21 correlated with longer training time or increased the likelihood of Recruit School failure. And, there was no direct evidence on that score. However, recruits were required to achieve a passing score on each of the eight Recruit School examinations and, although Recruit School failures were avoided, some candidates did not achieve passing scores the first time around. The Court of Appeals itself recognized that "if a particular candidate has difficulty" in doing so, "he is given assistance until he succeeds in passing the examinations." 512 F.2d at 963. Given the correlation between low scores on Test 21 and low final averages in Recruit School, it would be reasonable to infer that low Test 21 scorers were more likely to be those for whom additional assistance was necessary. The fact of no failures was also used to demonstrate a lack of validity of Test 21. However, since Test 21 was correlated with scores achieved *before* any assistance was provided, the fact of no failures is irrelevant to this issue.

precisely so as not to appear to preclude, under any circumstances, reliance on training program success as a basis for concluding that a test is job-related.

Second, the Court of Appeals never considered whether the Department's training program was reasonably designed to achieve proper employment objectives. *See* 512 F.2d at 965. To be sure, the record lacks the detailed evidence regarding the program on the basis of which such an evaluation should be made. In particular, evidence is lacking regarding those aspects of the training program singled out by the Court of Appeals as diminishing its value as a basis for validating the use of Test 21, *e.g.*, the fact that no recruits are dropped from the program for failure to pass one of the eight examinations. The Department did not offer any proof as to the reasons for this policy, although sufficiently rational reasons are immediately apparent and might have been the subject of an evidentiary showing.[63]

In the absence of such proof, this Court must decide whether the case is ripe for final disposition. Before validation of a test against a training program is either permitted or ruled out automatically, this Court may conclude that an employer should be required to relate the training program to the employment attribute the program is professed to impart and to demonstrate that his operation of the training program is reasonable, fair,

---

[63] The cost of extra assistance for a recruit who has nearly completed the training program may be substantially less than the cost of replacing that recruit with a new applicant who would begin the training program anew, thus the no-failure policy may be economically efficient. Further, the morale of the recruit class or the general atmosphere in which the training program is conducted might be adversely affected by the possibility of some recruits being dropped from the program causing the training program to lose efficiency with respect to the group as a whole, a result the Department may legitimately wish to avoid.

and suited to achieve that objective. If the Department cannot make such a showing in the *Davis* case, then its training program should not suffice as a basis for determining the "job-relatedness" of Test 21.

Third, the breadth of the Court of Appeals' statements regarding "training programs" is independently troublesome. As emphasized earlier, the validity of tests under the applicable legal and professional standards necessarily depends on the particular test in question and the circumstances of its use. Standardized tests used for purposes other than employment are sometimes validated against what might fall within an overly broad and ill-considered definition of a "training program." For example, academic entrance examinations used in assessing scholastic aptitude are most frequently validated through criterion-related studies, in which success in the curriculum for which the test is used in selecting is widely regarded as the most useful criterion. In another context, professional school curricula are the touchstone for validating tests in the certification and licensing of professionals such as teachers and lawyers. The distinctions between short pre-employment training programs carried on by individual employers and multi-year academic curricula seem obvious. Nevertheless, precision in analysis and language with respect to "training programs" seems particularly important in this Court's review of the lower Court's statements relating to "training programs" and "trainability."

3. *Verbal aptitude as an appropriate key attribute of recruit training.*

If the Recruit School program is adequately related to an important element of job performance, then the analysis must consider further whether verbal ability (regardless of how that attribute is measured) is itself

sufficiently related to the training program.[64] If that link is satisfactorily established, then verbal ability is necessarily related to job performance through the training program.

The affidavits of Drs. Mary L. Tenopyr and Diane E. Wilson offered by the petitioners contain conclusory observations in this regard. App. 174, 185-86. The affidavits offered by the respondents do not clearly dispute this conclusion. App. 49-57.

The District Court considered the relationship between verbal ability (independent of how measured) and the content of the training program as a link to job performance:

> "Study of the syllabus of the training course readily demonstrates the intricacy of police procedures, the emphasis on report writing, the need to differentiate elements of numerous offenses and legal rulings, and the subleties of training required in behavioral sciences and related disciplines. Daily the significance of these skills demanding reasoning and verbal and literacy skills is borne out in the crucible of the criminal trial court." 348 F. Supp. at 17.

The Court of Appeals did not find it necessary to consider this link in the evidentiary chain because of its view that verbal ability must be related directly to over-

[64] Two general avenues of proof are possible. The Department might offer proof as to the nature of the training program (dealing primarily with written materials, lectures, and report writing techniques) and expert testimony that this type of training program requires a certain level of verbal aptitude. The Department might also choose a different route and prove that (1) Test 21 actually measures verbal ability; (2) the eight Recruit School examinations actually measure the subject matter taught; and (3) there exists a statistically significant correlation between scores on Test 21 and scores on the Recruit School examinations. This chain of proof would permit an inference that verbal ability was a key element of recruit training.

all job performance. The question for this Court is whether there is a material issue of fact as to this point requiring the case to be remanded for the development of a more complete record at trial.

B. Are the Tests Used by the Department Proper and Valid Measures of the Knowledge Desired by the Department?

Further analysis of the issues raised by the *Davis* case is required if the Court concludes that (1) the Department has properly identified knowledge of certain subjects as a central attribute of police work, and (2) the Department's training program is a reasonable mechanism to impart that knowledge. At this point, the "job-relatedness" determination in this case becomes largely a matter of technical test validation considerations. Four specific issues are raised by the tests used in the *Davis* case.

1. ***Content validity of the Recruit School examinations.***

Under our analysis of the issues, the Department cannot prevail unless the eight examinations used to measure proficiency in the Recruit School are valid. In the absence of such a showing, any correlation between scores on Test 21 and scores on the eight examinations cannot be relied upon in making the required "job-relatedness" determination since the necessary link between the knowledge being taught in the program and the measurement tool does not exist. The required validity would be established through the technique of content validity, involving an analysis of the questions contained on the eight examinations and the curriculum actually taught in the training program. Neither court below addressed the need for this validation; and the record, which contains the curriculum syllabus but not the eight examinations, may not permit its resolution by this Court.

The District Court apparently assumed that the examinations given at the end of Recruit School validly assessed mastery of the topics taught. The District Court did not appear to consider that such a showing was required (and could not be assumed) or that technical learning on test validation might be of assistance in making the determination.[65] The Court of Appeals did not focus, as it might have, on the lack of content validity to support its conclusion that the correlation between scores on Test 21 and scores on Recruit School examinations was an insufficient basis for finding the necessary "job-relatedness."

2. *Content validity of Test 21.*

An apparently simple, but necessary, showing under our analysis is that Test 21 actually does measure verbal ability. This would be achieved in the same manner described above for the determination whether the Recruit School examinations actually measure the subject matter taught there. Both the District Court and the Court of Appeals seemed willing to assume that Test 21 did measure verbal ability, perhaps from a perusal of the versions of the test contained in the record.

3. *Correlation of Test 21 with Recruit School examinations.*

A separate issue raised by *Davis* is whether Test 21 shows a sufficient correlation with the results on the

---

[65] *Compare* Pennsylvania v. O'Neill, 348 F. Supp. 1084 (E.D. Pa. 1972), *aff'd by an equally divided court*, 473 F.2d 1029 (3d Cir. 1973 (*en banc*)), where the court rejected evidence of the job-relatedness of a test validated against a police department training program on a variety of grounds, one of which was that

> "There is no evidence as to the type of examination administered at the Academy. If in fact tests of the type contained on the entrance examination were used, a high correlation could be anticipated even if an individual had not mastered the training material." 348 F. Supp. at 1091.

eight examinations. Both the parties and the courts below seem to agree that Test 21 has been validated as a predictor through the technique of criterion-related validity—with the criterion being scores on the eight Recruit School examinations—and that the correlation was sufficient and statistically significant under appropriate professional standards.

The Court of Appeals did, however, inappropriately suggest that no correlation between two such examinations could ever suffice. The Court stated that the Department's validation data:

> "tends to prove nothing more than that a written aptitude test will accurately predict performance on a second round of written examinations, and nothing to counter this hypothesis has been presented to us." 512 F. 2d at 962 (footnote omitted).

In contrast, the APA STANDARDS recognize that "[f]or many employment . . . purposes" the "ideal criterion may be an achievement test. . . ." APA STANDARDS at 34. As a legal matter, the decisions thus far have suggested that test-to-test correlation is appropriately questioned only where the tests are of a similar type. *E.g.*, *Pennsylvania* v. *O'Neill*, 348 F. Supp. 1084 (E.D.Pa. 1972). The record in this case contains no suggestion that Test 21, an aptitude test, was essentially similar to, or even remotely resembled, the eight achievement exams used to assess curriculum mastery.[66]

---

[66] It is equally unclear, from a legal perspective, who bears the burden of showing that "test taking ability" accounted for a particular correlation. Nevertheless, in suggesting that "nothing to counter this hypothesis" that the correlations reflected test-taking abilities "has been presented," the Court of Appeals purported to dispose of this question, too. At least one decision has suggested the contrary. Harper v. Mayor and City Council of Baltimore, 359 F. Supp. 1187, 1203 (D.Md.), *modified and aff'd sub nom.* Harper v. Klosters, 486 F.2d 1134 (4th Cir. 1973).

4. Validity of the cut-off score on Test 21.

The fourth measurement issue raised by the *Davis* case is one not addressed by either court below. The record reveals that Test 21 was used by the Department with a cut-off score of 40; long use seems to have been the principal explanation for this particular cut-off score. App. 191. Respondents appear not to have challenged the use of a cut-off score but to have proposed a lower cut-off score of 35. 348 F. Supp. at 17.

Under professional standards, most cut-off scores when used for employment selection must be validated, usually through a process independent of the validation of the test itself. The cut-off score must be demonstrated to classify persons with a reasonable level of probability as to those likely to possess the desired level of verbal ability and those not likely to be so equipped.[67]

The record in *Davis* is silent on this issue and this Court may, therefore, wish to decide whether the Department's failure to justify the cut-off score provides grounds for affirming the Court of Appeals or for remanding for the taking of evidence on this and the other issues in the case which would profit from further proceedings below.

[67] There would be a similar issue with respect to the appropriateness of the cut-off score of 75 on the Recruit School examinations if that cut-off score were used for classification rather than advisory purposes.

CONCLUSION

For the reasons stated above, this Court should set out a single analytical framework to be used in determining the job-relatedness of employment selection techniques. That framework should be equally applicable to test and non-test methods of measuring employment selection criteria, should be designed to achieve basic objectives of rationality and fairness, and should be sufficiently flexible to permit a wide range of relevant evidence and to encourage continued experimentation and growth in measurement technology.

Respectfully submitted,

HOWARD P. WILLENS
DEANNE C. SIEMER
THEODORE S. SIMS
1666 K Street, N.W.
Washington, D. C. 20006
Attorneys for Amicus Curiae
Educational Testing Service

Of Counsel:

JOHN S. KRAMER
General Counsel
Educational Testing Service
Princeton, New Jersey

WILMER, CUTLER & PICKERING
1666 K Street, N.W.
Washington, D. C. 20006

Dated: December 19, 1975

CERTIFICATE OF SERVICE

I, Howard P. Willens, a member of the Bar of this Court and counsel for Educational Testing Service, hereby certify that I have this 19th day of December, 1975, caused three copies of the foregoing Motion for Leave to File Brief As Amicus Curiae and Brief Amicus Curiae for Educational Testing Service to be mailed, with first class postage thereon prepaid, to C. Francis Murphy, Esq., Corporation Counsel for the District of Columbia, counsel for petitioners Walter E. Washington, *et al.*; Robert H. Bork, Esq., Solicitor General of the United States; and Richard B. Sobol. Esq., counsel for respondents Alfred E. Davis, *et. al.* I further certify that all parties required to be served have been served.

HOWARD P. WILLENS