INDEX

212-451—75——1

CITATIONS

Miscellaneous—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.

ON WRIT OF CERTIORARI TO THE UNITED STATES COURT OF APPEALS FOR THE DISTRICT OF COLUMBIA CIRCUIT

BRIEF FOR THE FEDERAL RESPONDENTS

This brief is submitted on behalf of the Commissioners of the Civil Service Commission as respondents (see Rule 21(4) of the Rules of this Court).

QUESTIONS PRESENTED

1. Whether the procedures used by petitioners to select applicants for appointment to the District of Columbia Metropolitan Police Department have been shown to have a significant adverse impact on black applicants.

2. Whether the written entrance examination administered to applicants for appointment to the Metropolitan Police Department has been shown to be substantially job-related.

STATEMENT

1. To be eligible for appointment to the District of Columbia Metropolitan Police Department, an applicant must satisfy specified age and physical standards, must have a high school education (or must have worked for at least one year in the police department of a large city), must undergo a psychiatric evaluation and a character investigation, and must pass a written entrance examination (known as Test 21) (A. 101). Test 21—developed by the Civil Service Commission for use generally in the federal civil service (Pet. App. 48a)—was designed to measure ability to read and comprehend the written language. Its 80 questions include vocabulary, reading comprehension, reading interpretation, and general information items (A. 202).[1] An applicant must answer at least 40 of the 80 questions correctly in order to pass the test (A. 101).

Appointments to the police department are required by law to be made in the same manner as appointments to the federal classified civil service (D.C. Code § 4–103). The Civil Service Commission, which administers Test 21 to applicants for appointment to the police department, maintains a register containing the names of applicants who have passed Test 21 in the order of their scores on the test (see 5 C.F.R. 332.401). When the police department needs recruits, it requests the Commission to prepare a certification of eligibles. The certificate contains a sufficient number of names, drawn from the top of the register, to permit the appointing

[1] Three series of the test are reproduced at A. 209–278.

officer to consider three eligibles for each vacancy (5 C.F.R. 332.402). The appointing officer, "with sole regard to merit and fitness" (5 C.F.R. 332.404), must fill each vacancy from the highest three eligibles on the certificate who are available for appointment.[2]

An applicant who is appointed to the police department is assigned to the Recruit School for 17 weeks of training (Pet. App. 48a).[3] Eight broad subjects are covered in the school, and recruits are given written examinations in each subject. A recruit must score 70 percent in each subject in order to complete the training program successfully (Pet. App. 13a).[4] In 1967, any recruit who was unable to pass a Recruit School test was given additional help and was allowed to repeat the test (A. 102). Thus, "for all practical purposes [there were] no failures in Recruit School" at that time (*ibid.*; emphasis omitted).

---

[2] It appears from the record that, in the years 1968 through 1971, approximately 40 percent of the applicants who passed Test 21 were ultimately appointed to the police department (A. 34–35). The same affidavit also indicates, however, that, apparently as of February 1972, there was "no waiting list" for appointment (A. 35). It is not clear whether that statement implies that every applicant who passed Test 21 and satisfied other qualification standards was appointed to the police department.

[3] The detailed curriculum of the Recruit School as of November 1971 is reproduced at A. 110–171. In November 1967, the Recruit School apparently was a 12-week program (A. 102).

[4] There is some ambiguity in the record concerning the scores required for successful completion of Recruit School. The court of appeals used the figure 70 percent (Pet. App. 13a). That figure also appears in the version of the "Futransky study" (see p. 5, *infra*) that is reproduced in the appendix to the petition for a writ of certiorari (p. 54a). A slightly different version of the same study, reproduced in the separate Appendix, uses the figure 75 percent (A. 102).

2. In September 1970, respondent Harley (joined later by respondent Sellers) intervened in an action pending in the United States District Court for the District of Columbia involving a challenge to the Metropolitan Police Department's promotion practices.[5] The complaint in intervention named as defendants the Mayor-Commissioner of the District of Columbia, the Chief of Police of the Metropolitan Police Department, and the Commissioners of the United States Civil Service Commission (A. 25). It alleged that respondents, black residents of Washington, D.C., applied for appointment to the Metropolitan Police Department, were given Test 21, and were told that they did not pass it (A. 26). The complaint alleged that Test 21 excludes a disproportionately high number of black applicants and does not aid in predicting the performance of black police officers (*ibid.*).

Respondents requested that the court certify a class composed of "all black applicants who, since the beginning of 1968, have unsuccessfully sought appointment to the Police Department" (A. 28). They sought declaratory and injunctive relief against the use of Test 21 and an award of back pay to members of the class (A. 28–29).

The district court, on the basis of cross-motions for summary judgment and supporting affidavits, entered summary judgment for the defendants (Pet. App.

[5] The issues raised by the original action were resolved favorably to the defendants. *Davis* v. *Washington*, 352 F. Supp. 187 (D. D.C.). The decision was not appealed.

48a–52a).[6] The court found that the plaintiffs had shown (*id.* at 49a):

> (a) The number of black police officers, while substantial, is not proportionate to the population mix of the city.
>
> (b) A higher percentage of blacks fail the Test than whites.
>
> (c) The Test has not been validated to establish its reliability for measuring subsequent job performance.

The court accepted that "minimal" showing as "sufficient to shift the burden of the inquiry to defendants" (*ibid.*), and it accordingly turned to a consideration of the job-relatedness of Test 21.

The test is used by the Metropolitan Police Department to predict whether an applicant is likely to succeed in the Department's 17-week Recruit School (A. 102). A study of the relationship between Test 21 scores and Recruit School performance was prepared for the Civil Service Commission in 1967 by David L. Futransky (A. 99–109). The study found a significant positive relationship, for both black and white recruits, between scores on Test 21 and the Recruit School Final Averages (A. 102–103).[7] "The higher the score on [Test] 21 the more likely the trainee will have a Recruit School Average of 85% or higher" (A. 103). According to the Futransky

[6] The court also dismissed the complaint with respect to the Commissioners of the Civil Service Commission (Pet. App. 52a).

[7] A trainee's Recruit School Final Average is the average of his scores on the examinations given during the training program (A. 102).

212-451—75——2

study, "[t]his finding supports the conclusion that [Test] 21 is effective in selecting trainees who can learn the material that is taught at the Recruit School" (A. 103; emphasis omitted).[8]

The Civil Service Commission submitted in the district court the affidavits of personnel research psychologists and other professionals who expressed the opinion that the Futransky study established the job-relatedness of Test 21 as a professionally validated predictor of a recruit's success in the Department's training program (see A. 172–208). In particular, the affidavits stated that success in training is a professionally acceptable criterion for the purpose of establishing the validity of an entrance examination (A. 174–175, 179, 181, 188, 192–193, 203, 207),[9] that verbal

[8] The study also examined the relationship between Test 21 scores and certain measures of job performance. It found a positive but low relationship for white officers and no significant relationship for black officers (A. 99, 106–109). The study also found a low positive relationship between Recruit School Final Averages and job performance ratings for white officers and no significant relationship for black officers (A. 99, 104–106).

[9] Indeed, Dr. Owens stated that "the criterion of Recruit School Average is very likely substantially superior, as a criterion, to the criterion of job performance" in the context of this case (A. 207). Dr. Maslow expressed the view that, because a police officer's job performance depends in part on what he has learned in training school and also on a variety of unpredictable factors (such as the nature of his supervision and assignments), it would not be feasible, "*before* entry into training, to measure accurately the probable behavior of a trained policeman in future situations which are difficult to anticipate at the time of selection for training" (A. 195; emphasis in original). For that reason, and because "applicants who do not meet recruit school standards will not become policemen," Dr. Maslow concluded that "it is completely reasonable to aim the selection test directly at applicants in terms of likelihood to succeed in recruit school" (*ibid.*).

ability is a critical factor in training success (A. 185, 192), and that "the level of verbal ability measured by Test 21 is, at most, no higher than that required for successful completion of recruit school training as determined by the final grade average attained by the recruit" (A. 185). The affidavits stated that "Test 21 has a significant positive correlation with success in * * * Recruit School for both blacks and whites" (A. 179; see also A. 188, 202, 206).[10]

Respondents submitted the affidavits of two psychologists. Dr. Barrett stated that the Futransky study "shows no benefit for using [Test 21] in selection of Negroes" (A. 50). Although "[t]he study indicates that test scores of both white and Negro candidates are positively correlated with grades in Recruit School," the significant fact, in Dr. Barrett's view, was that "no relationship is shown between test score and [post-training] job performance" for black applicants (*ibid.*). In his opinion, the relationship between Test 21 scores and Recruit School Final Averages "is of no significance" (A. 51), because recruits are given extra help if they need it to ensure successful completion of training and because Recruit School

[10] Dr. Schwartz's affidavit stated that there was insufficient data to determine the relationship between Test 21 scores and job performance after training, or between Recruit School Final Averages and subsequent job performance, but that the demonstrated relationship between Test 21 scores and Recruit School Final Averages was itself sufficient to establish the job-relatedness of Test 21 (A. 183). Dr. Owens similarly stated that the lack of significant correlation between Test 21 and ratings of job performance after training "is not a concern" (A. 208) if the Test 21 scores are used to predict trainability *for* the job rather than proficiency *on* the job (A. 206–208).

grades "do not correlate with later performance" (*ibid.*). The other psychologist expressed the opinion that "to attempt to establish validity of an entry test solely on the basis of an achievement test is a somewhat questionable procedure in personnel research" (A. 55). Both of respondents' psychologists questioned the use of 40 correct answers as the passing score for Test 21 (A. 51, 56).

On the basis of this evidence, the district court found that Test 21 is job-related (Pet. App. 50a–51a):

> Plaintiffs and their expert affiants have misconceived the responsibilities and expertise required of modern police officers in a large metropolitan city such as the Nation's Capital. 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 subtleties 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. Law enforcement is a highly skilled professional service. The ability to swing a nightstick no longer measures a policeman's competency for his exacting role in this city. * * * The Court is satisfied that the undisputable facts prove the test to be reasonably and directly related to the requirements of the police recruit training program and that it is neither so designed nor operates to discriminate against otherwise qualified blacks.

3. The court of appeals, with one judge dissenting, reversed the judgment of the district court and re-

manded with directions to enter summary judgment for respondents (Pet. App. 1a–47a). The majority held that the plaintiffs "demonstrated on the record that Test 21 has a racially disproportionate impact," and that the defendants failed to meet "their heavy burden of showing that the test is related to job performance" (*id.* at 3a–4a).

On the issue of adverse impact, the majority ruled that "evidence establishing that significantly more blacks than whites fail a written entrance examination given to all applicants is sufficient, as a matter of law, to show the racially disproportionate impact of the examination" (*id,* at 6a–7a).[11] The court declined to consider evidence relating to the Metropolitan Police Department's active efforts to recruit black officers, because "efforts to recruit minority members have no bearing on a showing that an employment practice has a racially disproportionate impact" (*id.* at 9a).

On the issue of job-relatedness, the majority assumed, "despite serious doubt" (*id.* at 12a),[12] that

---

[11] The record shows that, from 1968 through 1971, the failure rate for locally-tested black applicants was 57 percent, while for other locally-tested applicants the rate was 13 percent (A. 34, 59). The failure rates for field-tested applicants in 1970 and 1971 were 47 percent for blacks and 12 percent for others (A. 34, 59).

[12] The majority hypothesized that the demonstrated correlation between Test 21 scores and the Recruit School Final Averages "tends to prove nothing more than that a written aptitude test will accurately predict performance on a second round of written examinations" (Pet. App. 12a). It noted that "[t]he Recruit School examinations are not in the record, and there is no other basis in the record for confirming or disputing the hypothesis" (*id.* at 12a n. 40).

"Test 21 is predictive of further progress in Recruit School" (*id.* at 12a–13a). The "ultimate issue," the court stated, is "whether that kind of proof is an acceptable substitute for a demonstration of a direct relationship between performance on Test 21 and performance on the job" (*id.* at 13a). The majority ruled that it was not (*id.* at 13a–14a; footnotes omitted):

> Appellees assert that their validity study establishes that Test 21 is predictive of "trainability," and that therefore the examination survives the *Griggs* standard. Appellants, on the other hand, have convincingly argued that the record evidence does not demonstrate a sufficient relationship between Test 21 scores and trainability. All entrants into Recruit School pass the final examinations with a grade of 70 or above; if a particular candidate has difficulty, he is given assistance until he succeeds in passing the examinations. The validity study revealed that persons with high Test 21 scores are more likely to achieve a final average exceeding 85 in Recruit School, but there is no evidence to support the proposition that a candidate with an average below 85 is more difficult to train or will not be as good a police officer as a candidate with an average over 85. Moreover, since applicants who scored below 40 on Test 21 have never been admitted to Recruit School, the validity study expressed no conclusion regarding the likely performance in Recruit School of Test 21 failures. For these reasons, and because 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. We might add that the Recruit School averages apparently have not been used by the Department for any purpose other than the attempt to validate Test 21 in this case. The *Griggs* standard does not, in our opinion, permit validation by a criterion that the employer itself does not believe sufficiently job related.

Judge Robb dissented (*id.* at 19a–21a). He concluded, as had the district court, that Test 21 "on its face is a fair and reasonable test of the ability of a police recruit to measure up to the qualifications" required of a metropolitan police officer—including "a basic understanding of the English language and the meaning of words and the ability to perceive the import of written sentences" (*id.* at 20a).

ARGUMENT

I. INTRODUCTION AND SUMMARY

Section 717(a) of the Civil Rights Act of 1964, as added by Section 11 of the Equal Employment Opportunity Act of 1972, 86 Stat. 111, 42 U.S.C. (Supp. III) 2000e–16(a), provides that "[a]ll personnel actions affecting employees or applicants for employment * * * in those units of the Government of the District of Columbia having positions in the competitive service * * * shall be made free from any discrimination based on race, color, religion, sex, or national

origin."[13] "In general, it may be said that the substantive anti-discrimination law embraced in Title VII was carried over and applied to the Federal Government" by the 1972 Act. *Morton* v. *Mancari,* 417 U.S. 535, 547.

Last Term, in *Albemarle Paper Co.* v. *Moody,* 422 U.S. 405, this Court reiterated the holding in *Griggs* v. *Duke Power Co.,* 401 U.S. 424, that Title VII prohibits the use of tests or other employee selection procedures that operate to exclude members of minority groups, unless the employer demonstrates that the procedures are substantially related to job performance—*i.e.*, that they reliably measure capability for, or accurately predict successful performance of, the jobs for which they are used.[14]

The plaintiffs bear the threshold burden of showing that a challenged procedure "select[s] applicants for hire or promotion in a racial pattern significantly different from that of the pool of applicants" (*Albemarle,* 422 U.S. at 425). The burden then shifts to the employer to demonstrate that the tests (or other

[13] Section 717(b) of the Act gives the Civil Service Commission authority to enforce the provisions of subsection (a) and to "issue such rules, regulations, orders and instructions as it deems necessary and appropriate to carry out its responsibilities under this section."

[14] Even apart from any disproportionate impact on minority group members, it has long been a statutory objective of the Civil Service Commission to ensure that its competitive examinations "are practical in character and as far as possible relate to matters that fairly test the relative capacity and fitness of the applicants for the appointment sought" (5 U.S.C. 3304(a)(1)). No issue is presented here concerning the applicability of that provision to this case (see Pet. App. 11a, n. 37).

challenged elements of the selection procedure) "have a manifest relationship to the employment in question" (*Griggs,* 401 U.S. at 432).[15]

The issues in this case are whether respondents met their threshold burden of showing that the Metropolitan Police Department's employee selection procedures have an adverse impact on black applicants and, if so, whether petitioners demonstrated that the challenged entrance examination is job-related. In our view, the present record, consisting of affidavits and exhibits submitted on cross-motions for summary judgment, is incomplete with respect to both issues.

1. Statistics in the record appear to show that both the total employee selection process used by the Metropolitan Police Department and the entrance examination administered as part of that process have a substantial adverse impact on black applicants. Petitioners could conceivably account for those raw statistics (and therefore rebut respondents' *prima facie* showing of adverse impact) by proof that the police department's special recruiting program for blacks caused the pool of black applicants to be atypical. Although petitioners do not suggest that the statistics can be explained in that manner, and although we doubt that so large a racial disparity in the total se-

[15] "If an employer does then meet the burden of proving that its tests are 'job-related,' it remains open to the complaining party to show that other tests or selection devices, without a similarly undesirable racial effect, would also serve the employer's legitimate interest in 'efficient and trustworthy workmanship.' * * * Such a showing would be evidence that the employer was using its tests merely as a 'pretext' for discrimination" (*Albemarle,* 422 U.S. at 425).

212-451—75——3

lection rate and the examination pass-fail rate could be accounted for by peculiarities in the applicant population, we believe that the issue should remain open for inquiry in the district court if, as we suggest, a remand is ordered with respect to other issues in the case.

2. The record does not adequately address the question of job-relatedness. Under Civil Service Commission regulations and current professional standards governing criterion-related test validation procedures, the job-relatedness of an entrance examination may be demonstrated by proof that scores on the examination predict properly measured success in job-relevant training (regardless of whether they predict success on the job itself).[16]

The documentary evidence submitted in the district court demonstrates that scores on Test 21 are predictive of Recruit School Final Averages. There is little

[16] The validity of a test (or other formal or quantified selection procedure) can be evaluated in several ways. A criterion-related validity study examines the extent to which scores on the test correlate with ratings on some other performance measure (called a criterion), such as training success or productivity. A construct validity study examines the extent to which the test measures a specified ability or trait (called a construct), such as clerical aptitude or general intellectual capacity. A content validity study examines the extent to which the test requires the demonstration of behaviors that are representative of the job itself (for example, a typing test for a stenographer) or measures mastery of a skill or course of study (for example, a law school examination). See American Psychological Association, *Standards for Educational and Psychological Tests*, pp. 25–31 (1974); Division of Industrial-Organizational Psychology of the American Psychological Association, *Principles for the Validation and Use of Personnel Selection Procedures*, p. 16 (1975).

evidence, however, concerning the relationship between the Recruit School tests and the substance of the training program, and between the substance of the training program and the post-training job of a police officer. It cannot be determined, therefore, whether the Recruit School Final Averages are a proper measure of success in training and whether the training program is job-relevant.

This Court's decision in *Albemarle* indicates that, to demonstrate a test's job-relatedness by a criterion-related validity study, the employer must present evidence showing that the performance ratings with which the test scores correlate are "sufficiently related to the [employer's] legitimate interest in job-specific ability to justify a testing system with a racially discriminatory impact" (422 U.S. at 433). That suggests—and current professional standards appear to support the suggestion—that it is for the defendants in this case to show that the Recruit School Final Averages are an appropriate measure of success in training and that the training program imparts knowledge, skills, and abilities required for performance of the post-training job.

That the existing record does not clearly reflect such evidence does not mean that Test 21 should be held unlawful. The applicable legal and professional standards in this technical area have undergone significant evolution since the validation study in this case was performed and the district court litigation was conducted, and it is therefore understandable why the parties did not focus in the district court on what now seem to be the critical issues. The plaintiffs did

not claim that the recruit School Final Averages were an inappropriate measure of training success or that the training program was not sufficiently related to the post training job, and the defendants' affidavits did not specifically speak to these points.

Thus, while it might be appropriate in other circumstances to decide a question of job-relatedness on the basis of an incomplete documentary record compiled on motions for summary judgment, the appropriate disposition here is to remand the case to the district court to afford the parties an opportunity to present evidence bearing on the issues that have now been identified as critical.

II. THE EVIDENCE SHOWS *PRIMA FACIE* THAT THE METROPOLITAN POLICE DEPARTMENT'S OVERALL PROCESS FOR SELECTING APPLICANTS, AND THE ENTRANCE EXAMINATION ADMINISTERED AS PART OF THAT PROCESS, HAVE A SIGNIFICANT ADVERSE IMPACT ON BLACK APPLICANTS

The entrance examination here, as in most situations, is one of several components of the total employee selection process used by the Metropolitan Police Department. The threshold inquiry in a case like this one, therefore, is whether the total selection process has a significant adverse impact on the employment opportunities of members of an identifiable racial, ethnic, or sex group.

If it does not, the employer should not be required under Title VII to demonstrate the job-relatedness of the individual components of the selection process, regardless of any adverse impact that one or another of those components might have if it were used separately. So long as no racial, ethnic, or sex group is

significantly disfavored by the employer's overall hiring process, the job-relatedness of the elements of that process should not ordinarily be of concern under Title VII.[17]

If the overall selection process does have such an adverse impact, however, then it is appropriate to examine challenged components of the process separately for adverse impact. In that context, a showing that an entrance examination significantly disfavors members of an identifiable minority or sex group should be sufficient to shift the burden to the employer to either rebut the *prima facie* showing or prove the job-relatedness of the examination.

Raw statistics showing comparative selection-rejection (or pass-fail) rates for racial, ethnic, or sex groups are highly probative with respect to adverse impact, but they are not necessarily dispositive. A showing that blacks are rejected for hire (or that they fail an entrance examination) in significantly higher proportions than whites should amount to a *prima facie* case of adverse impact. But it should be open to the employer to show that the different rejection or failure rates can be explained on the basis of peculiarities in the applicant population. If, for example, an employer's special recruiting efforts aimed

[17] Excusing the components of a selection process from a job-relatedness inquiry does not, of course, insulate from challenge under Title VII alleged employment discrimination against a particular individual. Even if no racial, ethnic, or sex group is disfavored as a whole by an employer's selection procedures, an employer may not engage in individual acts of discrimination on the basis of race, color, religion, sex, or national origin.

at blacks results in a group of black applicants whose average education or experience is significantly less than that of a normal group of black applicants, statistics showing a higher rate of rejection or test failure for blacks than for whites may reflect, not an adverse racial impact, but merely an atypical applicant population. See the California Fair Employment Practices Commission's *Guidelines on Employee Selection Procedures,* BNA Fair Employment Practices, 451:145, 451:151.

The record in this case shows that, in the years 1968 through 1971, approximately 43 percent of the black applicants and 87 percent of the other applicants passed Test 21 (A. 34). Approximately 17 percent of the black applicants and 33 percent of the other applicants were actually hired by the Metropolitan Police Department (A. 33–35).

These figures—reflecting a total selection rate for blacks [18] and a Test 21 pass rate for blacks that are only about one-half the corresponding rates for others—make out a *prima facie* showing that the Department's total selection process and the written entrance examination used as part of that process have a significant adverse impact on black applicants.

For the reasons stated above, these *prima facie* showings of adverse impact may be subject to re-

[18] The hiring statistics do not conclusively demonstrate what the selection rate is, because some persons selected for hire (*i.e.*, offered appointment) may decline employment. In the absence of any contrary indication, however, it is reasonable to presume that the proportions of blacks and others who are actually hired approximate the proportions who are offered appointment.

buttal at the threshold, without regard to the job-relatedness of the test, if the differences in the total selection rate or the Test 21 pass-fail rate could be explained on the basis of peculiarities in the applicant population. The record here indicates that, at least since August 1969, the Metropolitan Police Department followed a special recruitment program, "a significant portion of which [was] directed towards increasing the numbers of blacks in the department" (A. 66).

There is nothing in the record to indicate whether or to what extent that recruitment program may have resulted in an atypical pool of black (or white) applicants, nor do petitioners suggest that any special characteristics of the applicant population account for the substantially lower selection and passing rates for blacks. Moreover, it would take an extraordinary showing, which we doubt can be made here, to overcome the implications of raw statistics showing that blacks are selected for hire and pass Test 21 at a rate only one-half that for others.

In these circumstances, we believe that, on the existing record, respondents adequately met their burden of demonstrating adverse impact. If the Court determines, however, that the case should be remanded for other reasons (as we argue in the next section of this brief), petitioners should have an opportunity on remand to introduce evidence tending to show that the statistics can be accounted for on the basis of the applicant population.

III. A DETERMINATION OF THE JOB-RELATEDNESS OF TEST 21 SHOULD NOT BE MADE ON THE BASIS OF THE EXISTING RECORD

The job-relatedness of a testing device that disproportionately excludes black applicants must be demonstrated "by professionally acceptable methods" (*Albemarle Paper Co.* v. *Moody, supra,* 422 U.S. at 431). Because the inquiry is necessarily complex and technical, the courts should defer to experts in the fields of industrial psychology and psychometrics for guidance in evaluating evidence that purports to show that a test reasonably predicts job performance.

There are several accessible sources of professional expertise with respect to employee test validation. The American Psychological Association's *Standards for Educational and Psychological Tests* (1974) are generally regarded as stating in a comprehensive fashion the accepted standards of the psychological profession with respect to the development and use of tests. The *Standards* are different in several material respects from the 1966 edition, entitled *Standards for Educational and Psychological Tests and Manuals,* to which this Court referred in *Albemarle* (422 U.S. at 431, 435).[19]

The APA's Division of Industrial-Organizational Psychology (known as Division 14) recently pub-

[19] It is because of "the continuing evolution of the discipline" that "psychologists have avoided setting rigid rules and regulations with respect to research on and use of tests and other selection procedures" (*Amicus* Br. for Exec. Comm. of APA Division 14, p. 12).

lished its *Principles for the Validation and Use of Personnel Selection Procedures* (1975) (set forth in the appendix to Division 14's *amicus* brief in this case). The *Principles* were intended to be consistent with the 1974 APA *Standards* and "to clarify the applicability of the *Standards* (written for measurement problems in general) to the specific problems of employee selection, placement, and promotion" (p. 1).

Both the *Standards* and the *Principles* are designed to state the ideals of the profession. Neither requires satisfaction in all circumstances of each relevant ideal, and neither is meant to serve as an inflexible source of Title VII law.[20]

Apart from the *Standards* and *Principles,* the federal agencies with responsibilities in the area of equal employment opportunities—including the Equal Employment Opportunity Commission, the Department of Labor's Office of Federal Contract Compliance, and the Civil Service Commission—publish guidelines or regulations dealing in more concise fashion with test

---

[20] The *Standards* state (p. 8; emphasis in original), and the *Principles* reiterate (pp. 1–2):

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

validation.[21] Just as the EEOC *Guidelines* are given substantial deference in connection with a private employer's effort to prove the job-relatedness of a testing device, the Civil Service Commission's regulations are the principal source of administrative guidance in the context of a federal employer's use of a testing device.

The Civil Service Commission argued in the district court and the court of appeals that the documentary evidence in this record adequately demonstrated the job-relatedness of Test 21 as a professionally validated predictor of success in training. We have since taken a fresh look at the evidence in light of this Court's decision in *Albemarle,* the 1974 revision of the APA *Standards,* Division 14's recently published *Principles,* and Division 14's *amicus* brief in this case. We conclude that the existing record does not permit a determination of the job-relatedness of Test 21 and that the appropriate disposition is to remand the case to the district court to allow the parties an opportunity to present additional evidence bearing on the issue.

A. THE JOB-RELATEDNESS OF AN ENTRANCE EXAMINATION MAY BE DEMONSTRATED BY EVIDENCE THAT SCORES ON THE TEST ARE PREDICTIVE OF PROPERLY MEASURED SUCCESS IN JOB-RELEVANT TRAINING

Under current professional standards for test validation, the job-relatedness of an entrance examina-

[21] The EEOC *Guidelines on Employee Selection Procedures* appear at 29 C.F.R. Part 1607. The OFCC order on *Employee Testing and Other Selection Procedures* appears at 41 C.F.R. Part 60–3. The Civil Service Commission's regulations governing *Tests and Other Applicant Appraisal Procedures* appear at 37 Fed. Reg. 21557–21559.

tion may be demonstrated by a criterion-related validity study showing a significant positive correlation between scores on the examination and the criterion of success in training. The Civil Service Commission regulations explicitly so provide (¶ S2–2(a)(2)(c), 37 Fed. Reg. 21557). See also 29 C.F.R. 1607.5(b)(3); Division 14's *Principles,* p. 3, ¶ A(2). As Division 14's brief states, "training performance is recognized by a broad consensus of industrial psychologists as a legitimate criterion for validating tests and other selection procedures" (p. 43). That consensus is reflected in the record of this case (A. 174–175, 179, 181, 188, 192–193, 203, 207), and it is wholly consistent with the legislative history of Title VII.[22]

It is not necessary for the employer to show in addition that scores on the entrance examination (or that success in training) predicts successful performance of the post-training job. As the record indicates (A. 183, 195, 206–208), and as Division 14's brief states (pp. 42–43), training performance and job performance are frequently independent. "[T]he same prediction test cannot reasonably be expected to be

[22] Section 703(h), 42 U.S.C. 2000e–2(h), provides that it shall not be "an unlawful employment practice for an employer to give and act upon the results of any professionally developed ability test provided that such test, its administration or action upon the results is not designed, intended or used to discriminate because of race, color, religion, sex or national origin." The original version of the provision was introduced as a floor amendment by Senator Tower in order "to protect the system whereby employers give general ability and intelligence tests to determine the trainability of prospective employees" (110 Cong. Rec. 13492; see also *id.* at 11251). See generally *Developments in the Law: Employment Discrimination and Title VII of the Civil Rights Act of 1964*, 84 Harv. L. Rev. 1109, 1123–1126 (1971).

equally effective in predicting different performances; on the one hand, success in training, and on the other, subsequent success on the job, where the training program is designed to make a significant change in the skills, knowledges, and attitudes of the applicant" (A. 195). Whereas training performance may reflect the ability to acquire "the necessary knowledge and skills to perform the job, subsequent job performance is likely to reflect other characteristics (e.g., motivation) which are not ordinarily incorporated in the training nor necessarily reflected in training performance" (Division 14 Br. 42).

Applicable standards also make clear that the criterion of training success may properly be measured by scores on subject matter achievement tests administered during the training program. The Civil Service Commission's regulations provide that "[c]riterion measures may include * * * tests" (¶S3–3(b), 37 Fed. Reg. 21558). Indeed, the APA *Standards* state that, in certain circumstances, an achievement test may be "the ideal criterion" (¶E4, comment, p. 34).[23]

Proof that scores on an entrance examination predict scores on training school achievement tests, however, does not, by itself, satisfy the burden of demonstrating the job-relatedness of the extrance examination. There must also be evidence—the nature of which

[23] The opinion of respondents' expert that "to attempt to establish validity of an entry test solely on the basis of an achievement test is a somewhat questionable procedure in personnel research" (A. 55) thus appears out of harmony with the current professional consensus reflected by the *Standards*.

will depend on the particular circumstances of the case—showing that the achievement test scores are an appropriate measure of the trainee's mastery of the material taught in the training program and that the training program imparts to a new employee knowledge, skills, or abilities required for performance of the post-training job.

That this is part of the employer's burden of proving job-relatedness is suggested by this Court's decision in *Albemarle*. The employer there had presented evidence showing that the scores of existing employees on some of the entrance tests correlated with supervisory rankings of those employees in terms of a broad standard of overall job performance. The Court found the evidence insufficient to prove the job-relatedness of the tests, in part because the record failed to reveal what factors were actually taken into account by the supervisors in ranking the employees and the extent to which those factors were actually related to performance of the jobs for which the tests were used. The Court stated (422 U.S. at 433; emphasis in original, footnote omitted):

> There is no way of knowing precisely what criteria of job performance the supervisors were considering, whether each of the supervisors was considering the same criteria or whether, indeed, any of the supervisors actually applied a focused and stable body of criteria of any kind. There is, in short, simply no way to determine whether the criteria *actually* considered were sufficiently related to the Company's

> legitimate interest in job-specific ability to justify a testing system with a racially discriminatory impact.

Although the criterion measure in the present case is composed of achievement test scores in training school rather than supervisory rankings under an "extremely vague" standard (*ibid.*) of overall job performance, *Albemarle* suggests that the criterion measure here must also be justified—that is, that there must be evidence showing that the achievement tests actually measure the recruit's mastery of the training program, and that the training program imparts knowledge, skills, or abilities required for the job. In the absence of such evidence, it would be difficult to conclude that the Recruit School Final Averages are "sufficiently related to the [police department's] legitimate interest in job-specific ability" (422 U.S. at 433).

The published standards of the profession [24] and

[24] There are practical reasons why the judicial burden of proof in this technical area should track the professional standards governing the information that should be included in a validation study. Many employers rely on the judgment of psychologists concerning whether a particular test is valid for its intended use, and their judgment is likely to be based on a validation study conducted in accordance with prevailing professional standards. If those standards do not suggest the need for information bearing on a particular question, the psychologist may understandably fail to include that information as part of his report, and the employer may understandably fail to recognize any need to preserve evidence bearing on the question. It might be unfair to the employer if the courts were later to impose upon him the burden of producing evidence that he failed to preserve because the psychologist, consistent with applicable professional standards, thought the information was superfluous.

the relevant agency regulations seem to support this conclusion. First, with respect to the relationship between the achievement tests and the training program, the APA *Standards* state generally that "[t]he merit of a criterion-related validity study depends on the appropriateness and quality of the criterion measure chosen" (p. 27). And, while the APA designates as "Very Desirable" rather than "Essential"[25] the standard that "[a] criterion measure should itself be studied for evidence of validity" (¶E4, p. 34), it considers it "Essential" for the validation report to "comment on the adequacy of a criterion" (¶E3, p. 33) and to set forth "[t]he basis for judgments of criterion relevance" (¶E4.2, p. 34).

Division 14's *Principles* provide that "a reasonably valid, uncontaminated, and reliable criterion is assumed in criterion-related validation," and that "[i]f such a criterion measure cannot be developed, criterion-related validation is not feasible" (p. 5). Division 14's *amicus* brief accordingly states that "[i]t is essential * * * that measures used to evaluate per-

[25] The APA's standards are designated Essential, Very Desirable, or Desirable, depending on their importance and feasibility (*Standards*, p. 6). An "Essential" standard is "intended to represent the consensus of present-day thinking concerning what is normally required for competent use of a test"; it "indicate[s] what information or practices will be needed for most tests in most applications" (*ibid.*). A "Very Desirable" standard indicates "information or practices that contribute greatly to the user's understanding of the test and to competence in its use" (*id.* at 7). It is not listed as "Essential" if its "usefulness is debatable" (*ibid.*). A "Desirable" standard indicates "information and practices that are helpful but not Essential or Very Desirable" (*ibid.*).

formance in training * * * be objective and reliable" (p. 43).

Second, neither the APA *Standards* nor the Division 14 *Principles* expressly address the need for information concerning the relationship between the training program and the post-training job. Civil Service Commission regulations, however, provide generally that a criterion—such as success in training—should be "legitimately based on the needs of the Federal Government" and should represent "performance or qualifications requirements which are relevant to the job or jobs for which candidates are being evaluated" (¶¶S3–2(a)(2), S3–3(a), 37 Fed. Reg. 21558). Unless "the criterion has been defined rationally through a careful empirical analysis of job duties, job-relatedness of the appraisal procedure * * * is inferred but not assured" (¶S3–1(b)(2), 37 Fed. Reg. 21557). The EEOC *Guidelines* state that "[w]hatever 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)). Division 14's brief suggests that the appropriateness of using training success as a criterion depends upon the extent to which the training program "has been developed from an analysis of job requirements so that it represents skills and knowledge required by individuals to perform the job" (Br. 39–40).

Although we recognize that these professional and administrative standards are not free of ambiguity, we believe they tend to support our conclusion that the job-relatedness of an entrance examination may

be established by proof that scores on the examination predict scores on *proper* measures of performance in a *job-relevant* training program. The inquiry to which we next turn is whether, under that standard, the job-relatedness of Test 21 has been established.

B. THE RECORD DOES NOT CONTAIN ADEQUATE EVIDENCE CONCERNING THE RELATIONSHIPS BETWEEN THE RECRUIT SCHOOL EXAMINATIONS AND THE RECRUIT SCHOOL CURRICULUM AND BETWEEN THE CONTENT OF THE TRAINING PROGRAM AND THE POST-TRAINING JOB OF A POLICE OFFICER

The Futransky study found that the higher an applicant scores on Test 21 the more likely he is to achieve a Recruit School Final Average of 85 percent or above (A. 103). Although the study was necessarily restricted to Test 21 scores of 40 and above, it is proper to infer from Futransky's finding that the positive relationship between Test 21 scores and Recruit School Final Averages would hold true over the full range of scores and not merely the restricted range focused on by the study.[26] That is so because, under

[26] The correlation between Test 21 scores and Recruit School Final Averages is higher when the values are corrected for restriction in range (A. 103). Dr. Owens, a former president of Division 14, thus stated in his affidavit in the district court that, "[a]lthough a statistical problem of 'restriction in range' exists, if the lowest scorers on Test 21 *were* present to obtain a Recruit School average, it is my opinion that they would tend to score at the bottom of the distribution and to obtain, hypothetically, failing grades" (A. 206–207; emphasis in original).

The police department's policy, at least in November 1967, was to give additional help to any recruit who needed it to complete Recruit School successfully (A. 102). That policy does not make training success any the less appropriate a criterion by which to

the generally accepted principle of "linearity," valid judgments about a full range of test scores can properly be made on the basis of an evaluation of a restricted range of scores.[27]

While it is therefore fair to say that, the higher an individual's score on Test 21, the more likely he is to achieve satisfactory scores on Recruit School examinations, neither the Futransky study nor other evidence presented by any of the parties in the district court adequately addresses in light of current standards either the appropriateness of using Recruit School Final Averages as the measure of training performance or the relationship of the Recruit School program to the job of a police officer. The training curriculum as of November 1971 is part of the record (A. 110–171). One expert familiar with both the curriculum and the training materials used in Recruit School expressed the opinion that "successful completion of recruit school training requires a relatively high level of verbal ability" and that "the level of

---

validate an entrance examination. It may well be that the policy was feasible only because those applicants who were selected for appointment after passing the entrance examination had sufficient ability, at least with additional help, to master the materials taught in training and to achieve a satisfactory score on the Recruit School examinations.

[27] See, *e.g.*, Thorndike, *Personnel Selection*, pp. 169–176 (1949); Hawk, *Linearity of Criterion—GATB Aptitude Relationships*, 2 Measurement and Evaluation in Guidance 249 (1970); Brewer and Hills, *Univariate Selection: The Effects of Size of Correlation, Degree of Skew, and Degree of Restriction*, 34 Psychometrics 347 (1969); Cronbach, *Essentials of Psychological Testing*, pp. 128–137 (1970); Anastasi, *Psychological Testing*, pp. 72–78, 128–129 (1968).

verbal ability measured by Test 21 is, at most, no higher than that required for successful completion of recruit school training as determined by the final grade average attained by the recruit" (A. 185). The Recruit School examinations are not in the record, however, and it is not clear from the expert's affidavit whether her judgment was based on a familiarity with those examinations.

Similarly, while Dr. Maslow stated that the training program "is designed to equip [the recruit] with the knowledges, skills, and attitudes required for effective police work" (A. 192), it is not clear from his affidavit whether his statement reflects an independent analysis of the job of a police officer and a personal familiarity with the content of the training program.[28]

Because the foundation of these professional judgments is not made explicit in the affidavits, we do not believe that a determination of job-relatedness can fairly be made on the basis of the existing record.

C. THE CASE SHOULD BE REMANDED TO GIVE THE PARTIES AN OPPORTUNITY TO PRESENT EVIDENCE BEARING ON THESE ISSUES

For the reasons stated above, we think that the defendants in this case have the burden of presenting evidence concerning the appropriateness of the criterion measure and the job-relevance of the training

[28] Although lawyers and judges may have a general knowledge of some of the things a police officer does, that is not an adequate basis on which to rest a determination (see Pet. App. 20a–21a, 50a–51a) that the training program is job-relevant. That judgment should be made only on the strength of evidence showing that the content of the training program is related to important elements of a police officer's job.

program. It would be inappropriate, however, to dispose of this case at the present summary judgment stage by simply holding, as did the court of appeals, that Test 21 is unlawful because the record does not contain sufficient evidence of its job-relatedness.

The preferable course, and the one best calculated to result in a fair and proper resolution of the important job-relatedness issue, is to remand the case to the district court so that the parties will have an opportunity to present any additional evidence that may bear upon the questions that now seem critical to the job-relatedness determination. The generally accepted professional standards for test validation procedures have undergone material changes since November 1967, when the Futransky study was performed. For example, the 1966 APA *Standards,* unlike the 1974 version, did not contain material concerning the validity and appropriateness of the criterion measure (see p. 27, *supra*). Division 14's *Principles*, which bear more directly on problems of test validation in the employment context, were not published until June 1975. What may seem an important part of the job-relatedness determination to a psychologist today may have seemed superfluous in 1967.

The applicable legal principles have also evolved substantially since then. The current Civil Service Commission regulations were issued in October 1972; the current EEOC *Guidelines* were published in August 1970; *Griggs* was decided in 1971 and *Albemarle* in 1975.

It is therefore not difficult to understand why the parties did not focus in the district court on the questions that we now see as central. The dispute at the time was essentially over the abstract question whether a test that predicts training success is job-related if it fails also to predict successful job performance. The plaintiffs did not claim that the Recruit School examinations failed to measure training performance, or that the content of the training program was unrelated to the job of a police officer, and the defendants' affidavits did not address those questions.

In these circumstances, a remand for further proceedings in accordance with the principles we have outlined in this brief would be the most appropriate disposition of the case. Cf. *Albemarle Paper Co.* v. *Moody, supra,* 422 U.S. at 436. It is not clear to us whether evidence bearing on the relevant questions for the time period covered by this action[29] has been preserved or can be reconstructed.[30] But the parties should be afforded an opportunity to present whatever evidence they may have on these additional matters.[31]

---

[29] Respondents say (Br. 12–13, 35–36) that the department's training program has been substantially changed. What effect that may have on this litigation should be considered in the first instance by the district court.

[30] If not, the district court should consider whether the failure of the plaintiffs specifically to raise these questions in a timely fashion contributed to the present unavailability of evidence and, if the defendants were prejudiced by that failure, what the consequences should be with respect to this litigation. Cf. *Albemarle*, *supra*, 422 U.S. at 424.

[31] If a remand is ordered, the question of adverse impact should also remain open (see pp. 18–19, *supra*). On the issue of job-

CONCLUSION

The judgment of the court of appeals should be vacated and the case should be remanded to the district court for further proceedings consistent with the principles stated in this brief.

Respectfully submitted.

ROBERT H. BORK,
Solicitor General.
REX E. LEE,
Assistant Attorney General.
MARK L. EVANS,
Assistant to the Solicitor General.
RONALD R. GLANCZ,
HARRY R. SILVER,
Attorneys.

JANUARY 1976.

---

relatedness, the parties should also be free to present additional evidence on whether every applicant who passes Test 21 and is otherwise qualified is offered appointment to the police department (see note 2, *supra*) and, if so, whether the "cutting score" of 40 correct answers on Test 21 can be justified. See APA *Standards*, ¶ I4 and comment, pp. 66–67; Division 14 *Principles*, ¶ 11, p. 14. The need to justify the passing score—a question that relates to the particular use of the scores rather than to the predictive validity of those scores generally—would not arise if, in practice as well as theory (see pp. 2–3, *supra*), only higher scoring applicants are offered appointment. In that event, the "passing" score of 40 would have no special significance, because an applicant who "failed" the test would not have been offered an appointment even if his score had not been automatically disqualifying.