

Seton Hall Law Review

Vol. 10

1980

No. 4

BASING EMPLOYMENT DECISIONS ON TESTS: COMPLIANCE WITH THE NEW UNIFORM GUIDELINES

*Michael J. Zimmer**

I. INTRODUCTION

The development and implementation of testing devices as an integral part of employment decisions is a product of the twentieth century. Accompanied by the growth of experimental psychology, the use of personnel tests is based on the idea that proper responses to particular questions predict successful job execution. Examinations are attractive, in part, because of the appearance of objectivity: the higher the score, the more qualified the person is for any particular job. Thus, subjective or personal preference is eliminated from the employment decision. In the context of equal opportunity employment, a test may serve to avoid arbitrary selections based on race or gender. However, employers often neglect to undertake the burden of showing that test performance in fact predicts job performance. The failure to validate an examination raises serious problems where the test screens out a disproportionate percentage of women or of minorities.

Title VII of the Civil Rights Act of 1964 was enacted to equalize employment opportunities by prohibiting discriminatory practices.¹ Section 703 (a) (1) makes it unlawful for an employer "to fail or refuse to hire . . . any individual, or . . . to discriminate against any individual . . . because of . . . race, color, religion, sex, or national ori-

* A.B., J.D., Marquette University; LL.M., Columbia University; Professor of Law, Seton Hall University School of Law.

A version of this article will appear in a forthcoming book, *Federal Statutory Law of Employment Discrimination*, written by the author with Professor Richard F. Richards of the University of Arkansas (Fayetteville) and Professor Charles A. Sullivan of Seton Hall.

¹ Civil Rights Act of 1964, 78 Stat. 253 (codified at 42 U.S.C. §§ 2000e-2000e-17 (1964)).

gin.”² However, a saving clause is contained in Section 703 (h) which permits an employer “to give and to act upon the results of any professionally developed ability test provided that such test . . . is not designed, intended or used to discriminate because of race, color, religion, sex, or national origin.”³

In 1966 the Equal Employment Opportunity Commission (EEOC) issued “Guidelines on Employment Testing Procedures” to clarify this statutory provision, and these regulations were superseded and expanded in the 1970 “Guidelines on Employment Selection Procedures.”⁴ After the 1972 creation of an Equal Employment Opportunity Council to resolve conflicts among the several federal agencies charged with equal employment opportunity enforcement duties, the Council attempted to produce uniform federal agency testing guidelines.⁵ Despite a period of open dispute, the civil rights enforcement agencies of the government have reached agreement on “Uniform Guidelines on Employee Selection Procedures” which have now been officially promulgated by the Departments of Justice and Labor, the EEOC and the Civil Service Commission.⁶ The major

² 42 U.S.C. § 2000e-2 (1964).

³ 42 U.S.C. § 2000e-2(h) (1964). This language was included in the Act after extensive debate in the Senate. It was feared that Title VII would prohibit the use of all standardized tests without regard to the needs of business, and compel employers to hire unqualified applicants. See 110 CONG. REC. 9600 (1964) (remarks of Sen. Ellender); *id.* at 9599-9600 (remarks of Sen. Fulbright); *id.* at 9025-34 (remarks of Sen. Talmadge); *id.* at 9024-25 (remarks of Sen. Tower); *id.* at 8444-49 (remarks of Sen. Hill); *id.* at 7012-13 (remarks of Sen. Holland); *id.* at 5999-6000 (remarks of Sen. Smathers); *id.* at 5613-16 (remarks of Sen. Ervin). Congressional discussion focused on *Myart v. Motorola Co.*, a decision by the hearing examiner for the Illinois Fair Employment Commission, reprinted in the legislative history of Title VII at 110 CONG. REC. 9030-33 (1964). In this case, the hearing examiner ordered the employer to cease using a short form, five minute test of verbal and numerical skills because it “[did] not lend itself to equal opportunity to qualify for the hitherto culturally deprived and disadvantaged groups.” *Id.* at 9032. *Myart* thus suggested that standardized tests on which whites scored higher than blacks should not be conducted. *Id.* See *Motorola, Inc. v. Illinois Fair Employment Practices Comm’n*, 51 Lab. Cas. 51323 (Ill. Cir. Ct. 1965).

To avoid a similar application of Title VII, Senator Tower introduced an amendment which authorized the continued administration of “professionally developed tests.” 110 CONG. REC. at 13492. This exemption was criticized as too broad since it would permit an employer to give any test “whether it was a good test or not, so long as it was professionally designed.” *Id.* at 13504 (remarks of Sen. Case). Senator Tower offered a substituted amendment which became part of the present section 703(h) of Title VII. *Id.* at 13724.

⁴ 2 EMPL. PRAC. GUIDE, ¶ 4010 (1978). The EEOC recommended that employers using tests have available “data demonstrating that the test is predictive of or significantly correlated with important elements of work behavior which comprise or are relevant to the job or jobs for which candidates are being evaluated.” Guidelines on Employee Selection Procedures, 29 C.F.R. § 1607.4(c) (1975); 35 Fed. Reg. 12333, 12334 (Aug. 1, 1970).

⁵ The Guidelines provided in part that the “criteria of employee adequacy which the test is intended to predict or identify must be fully described” and “must represent major or critical work behaviors as revealed by careful job analyses.” 29 C.F.R. § 1607.5(b)(3) (1972).

⁶ 43 Fed. Reg. 38290 (Aug. 25, 1978). See note 27 *infra*.

thrust of these Guidelines has been to incorporate the validation standards for tests established by industrial psychologists.⁷

This article will present the technical test issues in the context of their legal significance. Lawyers, judges and, indeed, test experts need a common base of knowledge of the professional validation standards and how those standards are used in Title VII litigation.

II. WHAT IS A TEST?

Section 703 (h) is cast in terms of "any professionally developed ability test."⁸ However, since the section is an exemption from Title VII, types of employment inquiries other than tests may be covered by the general precepts of the Act although they do not lie within the purview of the statutory exception. The Uniform Guidelines, drafted in terms of selection procedures, assume a broad view as to what constitutes a test: "[s]election procedures include the full range of assessment techniques from traditional paper and pencil tests, performance tests, training programs, or probationary periods and physical, educational and work experience requirements through informal or casual interviews and unscored application forms."⁹

The Supreme Court has subjected a wide assortment of selection procedures to critical examination under the validation standards of the Guidelines.¹⁰ In *Griggs v. Duke Power Co.*,¹¹ the defendant lost because it failed to prove that its standardized test and high school diploma requirements each bore a manifest relationship to job performance.¹² The more recent case of *Dothard v. Rawlinson*,¹³ in which minimum height and weight prerequisites were challenged as gender discrimination, held that the defendant failed to produce sufficient evidence that height and weight demands correlated with the

⁷ See Uniform Guidelines on Employee Selection Procedures, 29 C.F.R. § 1607.5 (c) (1978). See generally A. ANASTASI, *PSYCHOLOGICAL TESTING* (4th ed. 1976); American Psychological Association, *Standards for Educational and Psychological Tests* (1974); Division 14, Industrial-Organizational Psychology, American Psychological Association, *Principles for the Validation and Use of Personnel Selection Procedures* (1975).

⁸ 42 U.S.C. § 2000e-2(h) (1964).

⁹ 29 C.F.R. § 1607.16Q (1979).

¹⁰ See, e.g., *Dothard v. Rawlinson*, 433 U.S. 321 (1977) (height and weight requirements); *Washington v. Davis*, 426 U.S. 229 (1976) (training program performance); *Albemarle Paper Co. v. Moody*, 422 U.S. 405 (1975) (examinations measuring verbal and nonverbal intelligence); *Griggs v. Duke Power Co.*, 401 U.S. 424 (1971) (standardized intelligence tests; high school education requirement).

¹¹ 401 U.S. 424 (1971).

¹² *Id.* at 431. *Griggs* is discussed more fully at notes 20-22 *infra* and accompanying text.

¹³ 433 U.S. 321 (1977).

degree of strength thought necessary for satisfactory job performance.¹⁴ *Griggs* and *Dothard* establish that a broad range of pre-employment conditions will be treated as tests and selection procedures and, under certain circumstances, will be scrutinized under the technical standards of the Uniform Guidelines and the accepted professional practices of industrial psychology to determine their validity.

III. DISPARATE IMPACT TRIGGERS THE DUTY TO VALIDATE TESTS

Title VII imposes no duty to validate employment tests unless the tests have a disparate or adverse impact upon groups protected by the statute.¹⁵ The Supreme Court has held that where a Title VII violation is alleged, the claimant has the burden of showing a prima facie case of discrimination.¹⁶ This may generally be established by demonstrating that the test in question "has a racially disproportionate impact."¹⁷ Once this is proved, it is the burden of the employer to prove that its tests are job-related—reasonable relationship exists between the test at issue and the requirements of the position for which it is administered.¹⁸ Even if the test is job-related, the claimant may still succeed if he can show that "other tests or selection devices, without a similarly undesirable racial effect, [which] would also serve the employer's interest in 'efficient and trustworthy workmanship.'" ¹⁹

¹⁴ *Id.* at 331-32.

¹⁵ See 42 U.S.C. § 2000e-2(h) (1964). 401 U.S. at 436.

¹⁶ *Albemarle Paper Co. v. Moody*, 422 U.S. 405, 425 (1975); *McDonnell Douglas Corp. v. Green*, 411 U.S. 792 (1973); see generally Note, *Legal Implications of the Use of Standardized Ability Tests in Employment and Education*, 68 COLUM. L. REV. 691 (1968).

¹⁷ *Jackson v. Nassau County Civil Serv. Comm'n*, 424 F. Supp. 1162, 1167 (E.D.N.Y. 1976). Disparate impact may be evidenced where "the tests in question select applicants for hire or promotion in a racial pattern significantly different from that of the pool of applicants." *Albemarle Paper Co. v. Moody*, 422 U.S. 405, 425 (1975). Cf. *Washington v. Davis*, 426 U.S. 229, 238 (1975); *Louisville Black Police Officers v. City*, 20 F.E.P. 1195, 1198 (W. D. Kentucky 1979) (constitutional challenge entails proof of intent to discriminate as well as proof of differential impact).

¹⁸ *Griggs v. Duke Power Co.*, 401 U.S. at 431. The Court declared that [t]he Act proscribes not only overt discrimination but also practices that are fair in form, but discriminatory in operation. The touchstone is business necessity. If an employment practice which operates to exclude Negroes cannot be shown to be related to job performance, the practice is prohibited.

Id. See *McDonnell Douglas Corp. v. Green*, 411 U.S. 792, 802 (1973); *Jackson v. Nassau County Civil Serv. Comm'n*, 424 F. Supp. 1162, 1167 (E.D.N.Y. 1976).

¹⁹ *Albemarle Paper Co. v. Moody*, 422 U.S. 405, 425 (1975) (quoting *McDonnell Douglas Corp. v. Green*, 411 U.S. 792, 801 (1973)). The implication is that the employer administered the test at issue as a device to perpetuate discrimination. *Albemarle Paper Co. v. Moody*, 422 U.S. 405, 425 (1975).

In *Griggs*, Negro employees alleged that their employer had violated the Civil Rights Act of 1964, Title VII, and sought to prohibit the company from demanding a high school diploma and satisfactory scores on a short form intelligence test as a condition of employment or promotion.²⁰ Disparate impact was shown in that "both requirements operate[d] to disqualify Negroes at a substantially higher rate than white applicants"²¹ Chief Justice Burger held that the employer failed to sustain its burden of proof since the requirements were not manifestly related to successful on-the-job performance. "[G]ood intent or absence of discriminatory intent does not redeem employment procedures or testing mechanisms that operate as 'built-in headwinds' for minority groups and are unrelated to measurability."²²

The need for validation standards as a consequence of discriminatory impact was further examined in *Albemarle Paper Co. v. Moody*.²³ In this case involving a class action brought by Negro employees at a paper mill, applicants for employment in the plant's skilled "lines of progression"²⁴ were required to possess a high school diploma and obtain adequate scores on two standardized tests. An industrial psychologist, hired by Albemarle, on the "eve of litigation," to review the job relatedness of its testing program,²⁵ compared test scores with supervisor rankings of incumbent employees.²⁶ The failure to begin the validation with an analysis of the skills and traits involved in the jobs for which the examinations were used led the Court to strike down the tests. Justice Stewart reasoned that *Griggs* demanded adherence in the instant case to the EEOC guidelines:

²⁰ 401 U.S. at 425-26. Prior to the enactment of the Civil Rights Act of 1964, the company overtly discriminated on the basis of race. *Id.* at 426-27. Subsequent to the passage of the Act, eligibility for employment or promotion was dependent upon completion of high school and adequate scores on the two professionally prepared aptitude tests, the Wonderlic Personnel Test and the Bennett Mechanical Comprehension Test. *Id.* at 427-28. The Court of Appeals upheld the company policy because there was no evidence of a discriminatory purpose. *Id.* at 428.

²¹ *Id.* at 426.

²² *Id.* at 432.

²³ 422 U.S. 405 (1975).

²⁴ *Id.* at 409. A line-of-progression determines job advancement. The employee commences at a lower level position and progresses up the line to a more advanced one. *See id.* at 438-39 app. Prior to 1964, the company had saved the more lucrative and skilled positions for whites. After a reorganization in 1968, all that was accomplished in fact was that the formerly "black" lines became tacked on to the bottom of the "white" lines. Pay disparities continued because advancement was based on seniority which only whites enjoyed, due to the plant's "previous history of overt segregation." *Id.* at 409.

²⁵ *Id.* at 410-11. The company used the Revised Beta Examination to measure nonverbal intelligence, and the Wonderlic Personnel Test to score verbal command. *Id.*

²⁶ *Id.*

"discriminatory tests are impermissible unless shown, by professionally acceptable methods, to be 'predictive of or significantly correlated with important elements of work behavior which comprise or are relevant to the job or jobs for which candidates are being evaluated.' " ²⁷

IV. JOB ANALYSIS AND TEST CONSTRUCTION

When the duty to validate arises, it is necessary to select a test expert to conduct the validation.²⁸ The first and critical step is to undertake a job analysis, which is the "process of collecting, and making certain judgments about all the important information relating to the nature of a specific job . . . to determine what workers actually do, how they do it, [and] why they do it" ²⁹ The expert can then conclude what knowledge and skills are needed for the particular task. Data for the analysis should be derived from as many sources as possible, including job descriptions, training manuals, performance records, interviews with incumbent workers and their supervisors, direct observation of the work being done, or even performance of the job by the test expert.³⁰ "The job analysis must focus on those requirements that differentiate this job from other jobs . . . [and] should also concentrate on those aspects of performance that differentiate most sharply between better and poorer workers." ³¹

The job analysis is important because development and use of the test is dependent on the accuracy and precision of the original study.³² In the context of Title VII litigation, the job analysis is crucial because it is the empirical basis for the judgment of the expert (to

²⁷ *Id.* at 431 (quoting EEOC Guidelines at 29 C.F.R. § 1607.4(c) (1975)). Because the EEOC Guidelines represent an "'administrative interpretation of the Act by the enforcing agency,'" the Supreme Court has deemed them to be "'entitled to great deference.'" 422 U.S. at 431 (quoting *Griggs v. Duke Power Co.*, 401 U.S. at 433-34; citing *Espinoza v. Farah Mfg. Co.*, 414 U.S. 86, 94 (1973)).

²⁸ Employers with any expectation of litigation concerning the validation study are advised to engage a person who qualifies as an expert for purposes of testifying in federal court and defending the results of that study. See FED. R. EVID. 702-06.

²⁹ U.S. Civil Service Commission, Bureau of Inter-Governmental Personnel Programs, *Job Analysis: A Guide for State and Local Governments* 1 (BIPP 152-85 1973). The Uniform Guidelines define job analysis slightly differently, depending on the type of validation envisioned: criterion-related, content, or construct. See 29 C.F.R. §§ 1607.14B(2), 1607.14C(2), 1607.14D(2) (1979). Validation Methods are discussed at notes 38-137 *infra* and accompanying text.

³⁰ See A. ANASTASI, *supra* note 7, at 437-38.

³¹ *Id.* at 437.

³² "Job analysis is a most critical and basic problem area in industrial psychology and is, or should be, the foundation of any industrial testing." M. BLUM & J. NAYLOR, *INDUSTRIAL PSYCHOLOGY* 26 (1968).

be challenged by plaintiff's expert and to be reviewed by the court) as to whether any examination is appropriate; if so, what validation strategy should be used; and, finally, whether the test elements selected do predict actual job behavior. The key fault in employment testing to date has been the lack of a job analysis.³³

When the analysis is completed, the test expert must judge whether skills and traits are required for which testing would be appropriate. If the position calls for traits that most of the population share or skills that most people will acquire easily with brief on-the-job experience and training, as most jobs do,³⁴ the expert should decide that no test is necessary for the selection of employees.³⁵

If the work includes critical elements for which examination is appropriate, test items must be prepared in anticipation of the validation study. Formulation of the examination requires expertise³⁶ because the test must reflect the singular aspects of the job revealed in the analysis and at the same time be amenable to validation techniques. Furthermore, the person selecting test items or batteries should be experienced with the vast array of tests that have been created for a wide variety of employment uses.³⁷

³³ See, e.g., *Albemarle Paper Co. v. Moody*, 422 U.S. at 431-32; *Griggs v. Duke Power Co.*, 401 U.S. at 431-34; *Vulcan Soc'y of N.Y. City Fire Dept., Inc. v. Civil Serv. Comm'n*, 490 F.2d 387, 394 (2d Cir. 1973). See also *Harper v. Mayor of Baltimore*, 359 F. Supp. 1187, 1201 (D. Md.), *aff'd in relevant part sub nom. Harper v. Kloster*, 486 F.2d 1134 (4th Cir. 1973); *Bridgeport Guardians, Inc. v. Bridgeport Civil Serv. Comm'n*, 354 F. Supp. 778, 792 (D. Conn.), *aff'd in part, rev'd in part*, 482 F.2d 1333 (2d Cir. 1973), *cert. denied*, 421 U.S. 991 (1975); *Chance v. Board of Examiners*, 330 F. Supp. 203, 220 (S.D.N.Y. 1971), *aff'd*, 458 F.2d 1167 (2d Cir. 1972).

Employers have tended to use tests out of a generalized and unverified notion that an exam will select better employees. A test is an inexpensive method, when not validated, of choosing employees when there are more applicants than the number of positions to be filled. Quite typically these tests have been used to choose employees for jobs that analysis would show require few skills or traits that could not be easily mastered within a short period of time on the job. See 422 U.S. at 428 n.24; note 35 *infra*.

³⁴ See I. BERG, *EDUCATION & JOBS: THE GREAT TRAINING ROBBERY* (1971).

³⁵ The Uniform Guidelines state as follows: "In general, users should avoid making employment decisions on the basis of measures of knowledges, skills, or abilities which are normally learned in a brief orientation period, and which have an adverse impact." 29 C.F.R. § 1607.5F (1979).

³⁶ According to one commentator, "the analyst must . . . be well grounded in psychology. Understanding of motivation, motor habits, and the organization of abilities are required as is knowledge of the multitude of tests now available." L. CRONBACH, *ESSENTIALS OF PSYCHOLOGY TESTING* 407 (3d ed. 1970).

³⁷ For further discussion of the various methods of occupational testing, see A. ANASTASI, *supra* note 7, at 435-63; Lupatka, *A 1977 Primer On The Federal Regulation of Employment Discrimination*, U. ILL. L. F. 69, 89-94 (1977).

V. THE CHOICE OF VALIDATION STRATEGIES

Once it is determined that testing is appropriate, the type of validation strategy must be chosen.³⁸ Two validation techniques, criterion-related and content, are generally accepted for employment testing³⁹ and will be the focus of this article. Construct validation is recognized in a general sense but is probably the least favored technique for use in the employment field.⁴⁰

A. Criterion-Related Validation

Criterion-related validation is a method of verifying that the results of a test predict job ability by comparing those results to present or subsequent job performance.⁴¹ It is a way of confirming the assumption that those who receive high scores on the test will generally be those who do well on the job.⁴² It is demonstrated, according to the Uniform Guidelines, "by empirical data showing that the selection procedure is predictive of or significantly correlated with important elements of work behavior."⁴³

1. Selecting the Criteria

In order to undertake criterion-related validation, job performance measures must be gleaned from the job analysis; these critical items of work behavior then serve as the criteria against which test scores will be judged.⁴⁴ Normally, criteria are derived from stan-

³⁸ Validation may be expensive and time consuming. Although generally less than \$5,000, a study can cost up to \$20,000. Reiter, *Compensating for Race or National Origin in Employment Testing*, 8 LOY. CHI. L.J. 687, 697 & n.49 (1977).

³⁹ The particular type of strategy used would depend upon the type of skills and traits revealed in the job analysis. See 29 C.F.R. § 1607.5A (1979).

⁴⁰ See *Douglas v. Hampton*, 512 F.2d 976, 985-86 (D.C. Cir. 1975); *Jackson v. Nassau County Civil Serv. Comm'n*, 424 F. Supp. 1162, 1171 n.17 (E.D.N.Y. 1976).

⁴¹ The basic idea of criterion-related validation is to compare performance on a test, used as a predictor, against a criterion, that is "a direct and independent measure of that which the test is designed to predict." A. ANASTASI, *supra* note 7, at 140. Criterion-related validity is indicative of "the effectiveness of a test in predicting an individual's behavior" in a given situation. *Id.*

⁴² See generally A. ANASTASI, *supra* note 7, at 140-51.

⁴³ 29 C.F.R. § 1607.16F (1979). See *Kirkland v. New York State Dept. of Correctional Serv.*, 374 F. Supp. 1361, 1370-71 (S.D.N.Y. 1974), *aff'd in part, rev'd in part*, 520 F.2d 420 (2d Cir. 1975), *cert. denied*, 429 U.S. 823 (1976). A. ANASTASI, *supra* note 7, at 146.

⁴⁴ The Uniform Guidelines state that any type of validity study should be based upon a job analysis and "[a]ny method of job analysis may be used if it provides the information required for the specific validation strategy used." 29 C.F.R. § 1607.14A (1979). On the other hand, "[c]ertain criteria may be used without a full job analysis if the user can show the importance of the criteria to the particular employment context." 29 C.F.R. § 1607.14B(3) (1979).

dards of proficiency in the performance of the duties of a particular job in question.⁴⁵ Other criteria, however, may include "production rate, error rate, tardiness, absenteeism, and length of service."⁴⁶

In *Washington v. Davis*,⁴⁷ the test under attack was designed to record verbal ability, reading and comprehension of applicants to a police officer training school.⁴⁸ The criterion against which scores on the exam were compared for the validation study was performance in the training school. No attempt was made to include criteria bearing on the ultimate job performance of police work.⁴⁹ Although a correlation was shown between a high score on the test and high ranking in recruit school,⁵⁰ the training program graduated everyone, and there was no proof that the test predicted any dimension actually associated with the job. Nevertheless, the Supreme Court accepted the use of training program performance as the sole criterion upon which test validity was measured.⁵¹

Such a decision may hinder the movement toward imposition of meaningful test validation standards. The use of training program performance as the sole criterion for test measurement is suspect since success in a training procedure as revealed by test scores may merely

⁴⁵ 29 C.F.R. § 1607.14B(2) (1979).

⁴⁶ 29 C.F.R. § 1607.14B(3) (1979).

⁴⁷ 426 U.S. 229 (1976).

⁴⁸ *Id.* at 234-35. The examination was one utilized generally in the federal service. *Id.* at 234.

Although the complaint in *Washington v. Davis* alleged a fifth amendment due process deprivation of employment rather than a Title VII violation, the court of appeals had stated that Title VII decisions furnished relevant instruction to be considered in the disposition of the case at bar. "The many decisions disposing of employment discrimination claims on constitutional grounds have made no distinction between the constitutional standard and the statutory standard under Title VII." *Washington v. Davis*, 512 F.2d 956, 958 n.2 (1975), *rev'd*, 426 U.S. 229 (1976). However, the Supreme Court reversed and held that a constitutional challenge mandated proof of intent to discriminate as well as proof of disparate impact. 426 U.S. at 239, 252. See note 17 *supra*.

⁴⁹ 426 U.S. at 235.

⁵⁰ *Id.* at 250.

⁵¹ *Id.* at 250-52. Justice Brennan, dissenting, argued as follows:

Where employers try to validate written qualification tests by proving a correlation with written examinations in a training course, there is a substantial danger that people who have good verbal skills will achieve high scores on both tests due to verbal ability, rather than "job-specific ability." As a result, employers could validate any entrance examination that measures only verbal ability by giving another written test that measures verbal ability at the end of a training course. Any contention that the resulting correlation between examination scores would be evidence that the initial test is "job related" is plainly erroneous.

Id. at 270 (Brennan, J., dissenting).

reflect formal educational experience. Grades on many standardized exams simply reflect years of formal education.⁵²

Another problem in regard to selection of criteria for job performance is present in the situation where applicants will be chosen for an entry level job but the employer wishes to test for skills and traits of higher level positions to which employees may be promoted. The question is whether the skills and traits demanded for the higher positions may be criteria included in the job performance measure used to validate the test. The Uniform Guidelines posit that "[i]f job progression structures are so established that employees will probably, within a reasonable period . . . and in a majority of cases, progress to a higher level, it may be considered that the applicants are being evaluated for a [higher level] job. . . ." ⁵³

The Supreme Court appeared to adopt the Guideline stance in *Albemarle*.⁵⁴ In deciding whether a test could be used for applicants to entry level jobs which validly measured qualifications possessed by workers in higher positions, the Court described the factors to be used as "normal speed of promotion . . . , the efficacy of on-the-job training in the scheme of promotion, and . . . the possible use of testing as a promotion device, rather than as a screen for entry into low-level jobs."⁵⁵ Thus *Albemarle*, coupled with the Uniform Guidelines, mandate scrutiny of employer policies regarding promotion where an employer utilizes a screening test based on considerations related to promotion rather than the likelihood of success at an entry level position.

A third problem concerning criteria selection arises where supervisor ratings are included as factors of job performance. This is called

⁵² See A. ANASTASI, *supra* note 7, at 440. Compare the Uniform Guidelines, which accept the use of success in training programs as a criterion but limit the potential for abuse:

[T]he relevance of the training should be shown either through a comparison of the content of the training program with the critical or important work behavior(s) of the job(s), or through a demonstration of the relationship between measures of performance in training and measures of job performance.

29 C.F.R. § 1607.14B(3) (1979).

⁵³ 29 C.F.R. § 1607.5I (1979).

⁵⁴ *Albemarle* concerned the validity of a concurrent criterion-related validation procedure, wherein the validity of a test is determined by administering it to incumbent employees, and then correlating those scores with job performance ratings. 422 U.S. at 429 n.25. For a discussion of predictive and concurrent criterion-related validation, see notes 69-77 *infra* and accompanying text.

⁵⁵ 422 U.S. at 434. The case was remanded to the district court for a factual finding of the test's "job-relatedness" and the fashioning of a remedy. *Id.* at 436. The Court did not hold that concurrent criterion-related validation was an unacceptable technique, but rather that the employer had failed to prove the job-relatedness of the testing procedure to job success. *Id.* See notes 69-77 *infra* and accompanying text.

criterion contamination: the criterion measure is not independent of, but is influenced by, the test score.⁵⁶ For example, if the job performance criterion is partly measured by a supervisor's rating, there would be criterion contamination if the supervisors knew the test scores of the employees they were grading.⁵⁷

Another criterion contamination concern is the potential bias of the supervisors toward those ranked and how this may influence the supervisors' perceptions of employee job performance.⁵⁸ In *League of United Latin American Citizens v. City of Santa Ana* (LULAC),⁵⁹ test experts, in attempting to validate a firefighters' exam using supervisors' ratings,⁶⁰ compiled some "noncognitive" criteria which related to "compatibility, reasonableness, and team work."⁶¹ These were included to discover supervisor bias. The original plan of excluding from the study those firefighters who received extreme reactions from the supervisors⁶² was changed when it was realized that almost all of the minority firefighters would have to be eliminated from the test sample.⁶³ After a review of the validation study's history, the court rejected a finding of job relatedness and, therefore, of test validity. "Having demonstrated that the techniques designed to root out bias had been unsuccessful, they cannot rely on demonstrably biased data."⁶⁴

In dicta, the court in LULAC noted that there was no affirmative duty to root out unconscious bias when undertaking validation studies. Instead, the validation duty was limited to rectifying the ef-

⁵⁶ See A. ANASTASI, *supra* note 7, at 141-42.

⁵⁷ A. ANASTASI, *supra* note 7, at 141; Bass & Burner, *Ethnic Group Differences in Relationships Among Criteria of Job Performance*, 5 J. OF APPLIED PSYCH. 101 (1973).

⁵⁸ See A. ANASTASI, *supra* note 7, at 194.

⁵⁹ 410 F. Supp. 873 (D.C. Cal. 1976).

⁶⁰ The validation method employed in this case is known as concurrent criterion-related validation. It entails the testing and job performance rating of a random sample of incumbent employees. See notes 69-77 *infra* and accompanying text.

⁶¹ 410 F. Supp. at 905. The experts were cognizant of the dangers of subjective criteria. "Employees who are well liked are apt to be treated kindly by raters, while those who are disliked tend to be treated more harshly . . ." *Id.*

⁶² *Id.* This was one of many ways the validation study experts intended to root out supervisor bias. Nine "cognitive" factors supposedly representative of critical job dimensions were stressed, and the supervisor training program for employee ratings emphasized the disregard of general impressions. Employee rankings were not individually established, but rather groups of supervisors met to discuss each one. Safeguards were also taken to prevent supervisors from being apprised of an individual's test score. *Id.*

⁶³ *Id.* There were only thirty-four Mexican-Americans in the sample, which the court characterized as "small to begin with." *Id.*

⁶⁴ *Id.* "In short, the study examined the ratings for bias, found strong evidence of bias, and then proceeded to use the biased results." *Id.*

fects of bias when discovered in the course of a study.⁶⁵ That dicta is questionable. The Guidelines do contain an affirmative obligation to eliminate bias, whether conscious or unconscious, when supervisors' ratings of job performance are used in validation studies.⁶⁶ While the Guidelines fail to specify a method to preclude bias, test experts may apply techniques that reveal and then exclude the consequences of any bias that may exist among supervisors. If this is not feasible, then the experts should consider alternatives to supervisors' ratings. For example, the study in LULAC could have been restructured to provide for an evaluation of work performance by independent examiners who might be expected to respond with sensitivity to the possibility of unconscious bias.

In *Albemarle*, the eve-of-litigation validation study also utilized supervisory ratings as the criterion for validating the test. Supervisors were requested to determine which employees, "irrespective of the job that they were actually doing, . . . did a better job than the person they were rating against."⁶⁷ The Supreme Court found this standard to be fatally vague because of an inability to identify the specific job performance criteria considered. The Court concluded that "[t]here 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."⁶⁸ From the foregoing, it appears that courts will exercise great caution when confronted with subjective supervisory standards or criterion contamination in a validation study. Bias must be rooted out to prove effectively the validity of the employment test.

2. Concurrent vs. Predictive Validation

When the test items are selected and the job performance criteria are well defined, the test expert must choose a medium for conducting the study. There are two recognized methods of criterion-related validation: predictive and concurrent.⁶⁹ In predictive validation, a sample group of applicants are given the exam and, without regard for their individual scores, are assigned to work on the

⁶⁵ *Id.* Noncognitive controls were described as "desirable" but not "necessary." *Id.*

⁶⁶ 29 C.F.R. § 1607.14B(2) (1979).

⁶⁷ 422 U.S. at 433.

⁶⁸ *Id.* (emphasis in original).

⁶⁹ See A. ANASTASI, *supra* note 7, at 140-41; *Kirkland v. New York State Dept. of Correctional Serv.*, 374 F. Supp. at 1370.

job for which the test is ultimately to be used. After sufficient time has elapsed to allow the workers to become accustomed to their duties, the performance of each member of the sample is measured by the test expert in terms of the criteria defined for that job. The test and job performance scores are then compared to determine if scores on the test generally predict work performance.⁷⁰ Concurrent validation, in contrast, involves administration of the test to a sample group of incumbent employees who are presently engaged in the particular duties and measurement of their performance on the job to determine if test scores predict performance.⁷¹

There are practical drawbacks to the predictive correlation method because of the lapse of time necessary between testing and the measurement of job performance. In addition, employers are frequently unwilling to hire the entire group of applicants in a predictive sample.⁷² Yet there are more serious problems associated with the concurrent technique. First, the use of incumbent employees confuses the flow of causation. The employees may have learned to do well on the test because of their good job performance; as applicants, they may not have scored equally well prior to the actual work experience.⁷³ If the predictive method is used, the order of events establishes causality. By issuing the test to a sample of applicants, hiring them, evaluating their performance on the job and establishing a correlation between test and job performance, evidence exists that a high test score predicts good job performance.⁷⁴

⁷⁰ The term "prediction" can be used in the broader sense, to refer to prediction from the test to any criterion situation, or in the more limited sense of prediction over a time interval. It is in the latter sense that it is used in the expression "predictive validity."

A. ANASTASI, *supra* note 7, at 140.

⁷¹ *Vulcan Soc'y of N.Y. City Fire Dept., Inc. v. Civil Serv. Comm'n*, 360 F. Supp. 1265, 1273 (S.D.N.Y.), *aff'd*, 490 F.2d 387, 394 (2d Cir. 1973).

⁷² The Uniform Guidelines specify that employers need not hire employees in order to perform a validation study. 29 C.F.R. § 1607.14B(1) (1979).

⁷³ The court in *Vulcan* referred to expert testimony criticizing the concurrent method because "some distortion may result from either the experience or lack of motivation of the current employees who participate in the examination for experimental purposes." 360 F. Supp. at 1273.

⁷⁴ This is known as a positive correlation between test and job performance. A test may also be valid where those who receive lower scores perform better on the job, creating a negative correlation. A curvilinear correlation, wherein low scores indicate poor job performance, middle scores peak toward very good performance, and higher scores dip back toward poor job performance, may be utilized as well for predictive purposes. See Cooper & Sobol, *Seniority and Testing Under Fair Employment Laws: A General Approach to Objective Criteria of Hiring and Promotion*, 82 HARV. L. REV. 1598, 1661 n.113 (1969).

A second flaw inherent in the use of incumbent employees is that a concurrent study generates no job performance data about individuals who were not and will not be hired because of their low test scores. Instead, the conclusion of job-relatedness must be generalized from incumbents, who have the benefit of their work experience, to applicants who will be rejected. The predictive method produces data concerning low scorers as well as high ones, which serves as a basis for inferences about the job performance of those who, in the routine application of the exam, would and would not be hired.

Third, concurrent validation is inferior to the predictive technique in terms of the composition of the sample. The predictive method draws its sample from the regular flow of job applicants, which is expected to "be representative of the candidates normally available in the relevant labor market . . . and should insofar as feasible include the races, sexes, and ethnic groups normally available in the relevant job market."⁷⁵ Because a concurrent study uses incumbent employees, it is difficult for such a sample to be representative of the typical job applicant group. The Supreme Court in *Albemarle*, relying on a standard of the American Psychological Association, criticized the concurrent study undertaken in that case because the sample of incumbent employees used in the study did not reflect the applicant population who would take the test. The validation study only included "job-experienced, white workers; but the tests themselves are given to new job applicants, who are younger, largely inexperienced, and in many instances nonwhite."⁷⁶ Concurrent validation may not be available where the incumbent employees differ too much in demographic composition from the normal flow of applicants.⁷⁷

3. Differential Validation

Another sample-related controversy is the possibility that test scores will be valid for applicants of one race but not for another. The Uniform Guidelines treat the problem as one of unfairness, which is defined as "[a] condition in which members of one race, sex, or

⁷⁵ 29 C.F.R. § 1607.14B(4) (1979).

⁷⁶ 422 U.S. at 435.

⁷⁷ Whether the study is predictive or concurrent, additional sample problems may occur. For example, statistical techniques used in validation become more persuasive and meaningful as the number of participants increase. Hence, it is important that the test constructor have a group large enough to yield significant results. See generally A. ANASTASI, *supra* note 7, at 89-90.

ethnic group characteristically obtain lower scores on a selection procedure than members of another group, and the differences are not reflected in differences in measures of job performance.”⁷⁸ Differential validation subsumes two possibilities. One is that a test may be valid for one group but not for another. For example, a sufficient correlation may exist between test scores and job performance for blacks, while for whites the test simply cannot predict how well the group will perform.⁷⁹ The second possibility is that the same test scores might predict different levels of job performance depending upon the race of the test-taker. In this manner, the use of a common test score could underpredict the job performance of one ethnic or racial group and overpredict that of another.⁸⁰ Although the earlier EEOC Guidelines required employers to validate tests by racial groups where “technically feasible,”⁸¹ the Uniform Guidelines are less demanding. Recognizing that the concept of fairness of selection procedures is in a state of evolution, the Guidelines place the duty to conduct fairness studies only upon employers with a large number of applicants of a particular race for particular jobs, and test developers.⁸² Hence, the employer whose work force is not large enough to validate a test by separate racial or ethnic groups may be excused from this obligation.⁸³

A validation study which indicates that the test is job-related for one group but not valid for the other justifies utilizing the examination only for members of the first group, and provides a defense to any claim of disparate treatment discrimination.⁸⁴ A study may re-

⁷⁸ 29 C.F.R. § 1607.16V (1979).

⁷⁹ This is termed slope bias, because the plotting on a graph of the separate regression lines for the two groups would have them cross. A. ANASTASI, *supra* note 7, at 192-95.

⁸⁰ *Id.* at 195-97. This is called intercept bias. The validity coefficients are the same for the two groups and therefore the graph of their respective regression equations would show the same slope. The lines are parallel, however, instead of identical, and cross or intercept the job performance axis (or test score axis) of the graph at different places. *Id.*

The issue of differential validity is acute in the world of psychology due to conflicting findings of recent research. A. ANASTASI, *supra* note 7, at 194-97; see Bartlett, Bobko & Pine, *Single-Group Validity: Fallacy of the Facts*, 62 J. OF APPLIED PSYCH. 155 (1977); Katzell & Dyer, *Differential Validity Reviewed*, 62 J. OF APPLIED PSYCH. 137 (1977).

⁸¹ 29 C.F.R. § 1607.4(a) (1978). “The term ‘technically feasible’ as used in these guidelines means having or maintaining a sufficient number of minority individuals to achieve findings of statistical and practical significance, the opportunity to obtain unbiased job performance criteria, etc.” 29 C.F.R. § 1607.4(b) (1978).

⁸² 29 C.F.R. § 1607.14B(8) (1979).

⁸³ *Id.* Thus, the “technical feasibility” standard remains, but has been modified. *Id.*

⁸⁴ See 29 C.F.R. § 1607.11 (1979); EEOC v. Local 638, Sheet Metal Workers, 532 F.2d 821 (2d Cir. 1976); *Kirkland v. New York State Dept. of Correctional Serv.*, 520 F.2d 420 (2d Cir. 1975), *cert. denied*, 429 U.S. 823 (1976); Reiter, *supra* note 38, at 702 n.76.

veal that the test is valid as to both groups but that the same score predicts different levels of job performance for the members of each group. In that situation, separate cut-off scores for each group could be used to reflect the different meaning the same score has depending upon the racial affinity of the test-taker.⁸⁵

In *Albemarle*, the Supreme Court appeared to accept the notion of differential validation. A concurrent study of a test was run among incumbent employees, mostly whites, in the higher level jobs. Justice Stewart acknowledged that differential validation would not be "technically feasible" in that particular sample because there were very few incumbent minority employees at those levels. Nevertheless, the Court criticized the employer for failure to direct the study at entry level positions. There was "no clear showing that differential validation was not feasible for lower level jobs."⁸⁶

4. Correlation Coefficient

When the sample has been run and all data collected and recorded according to the initial design of the validation study, the test and job performance measures can be compared to determine if a positive correlation exists between the two. The relationship between the two measures can be visualized by graphing the scores of each individual in the sample by their test and job performance scores. Thus, Figure 1 has the horizontal axis representing test scores and the vertical axis representing job performance scores. An example for an individual in the sample, "z", indicates that she received a test score of 6 and a job performance score of 10.

If every score fell on the dotted line O-A and the scales for scoring test and job performance were the same, the correlation between test score and job performance would be perfect: for every increase of one in the test score, the performance measure would also go up one. Thus, this perfect positive correlation is expressed numerically as +1. A perfect negative correlation, where every step up of one in test scores is matched with a decline of one in job performance, would be stated as -1. But the actual scores in validation studies invariably fall somewhat between the extremes. If the scores fell completely at ran-

⁸⁵ See 29 C.F.R. § 1607.5(b)(5) (1978). Where tests are generally geared to white cultural norms and backgrounds, and the subject tested is unfamiliar with these norms, the results may reflect cultural differences rather than ability, so that differential scoring may be appropriate. Cooper & Sobol, *supra* note 74, at 1645-46.

⁸⁶ 422 U.S. at 435.

dom on the graph, there would be no relationship at all between test scores and job performance and that would be expressed as a zero correlation.

The scores in Figure 1 appear to express some positive relationship between the two variables because the scores are grouped along the line O-A. To move from the visual to the numeric or statistical expression requires a method of computation called the Pearson Product-Moment Correlation Coefficient.⁸⁷ The basic theory is to determine mathematically how much each employee is above or below the mean or average scores of each variable. Thus a person above average on both test and job performance measures would have two positive scores; a person below average in both would have two negative standard scores. The next step is to multiply each individual's standard score, the difference from the mean, on each variable.

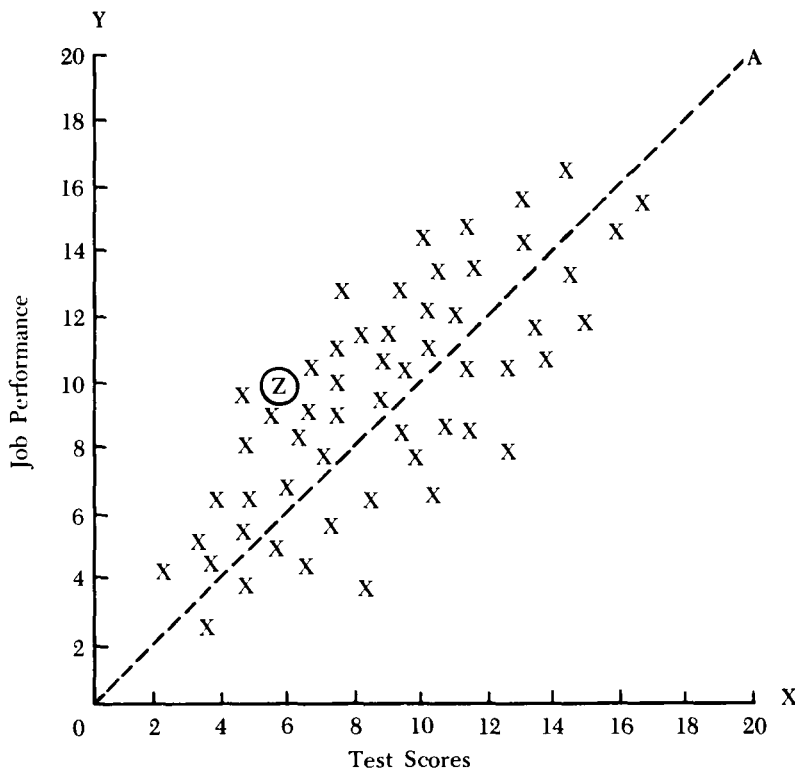


Figure (1)

⁸⁷ A. ANASTASI, *supra* note 7, at 106.

The correlation is the coefficient mean of those products. Suppose that the job performance test scores were tabulated for test validation of a very small sample,⁸⁸ as in Figure 2. One must tabulate the results and then develop the data into means, deviations from the means, standard deviations, product of the deviations, and, finally, compute the correlation coefficient. In short, the correlation coefficient is a mathematical way to show that two characteristics, test score and job performance, are related to each other.⁸⁹

It is also important to demonstrate that the relationship can be used to predict whether all the individuals expected to take the test

1.

N Test Takers	X Test Score	d _y Test Deviation from Mean	d _x ² Squared Test Deviations	y Job Score	d _y Job Deviation from Mean	d _y ² Squared Job Deviations	(d _x)(d _y) Product of Deviations
1	20	+6	36	20	+1	1	+6
2	18	+4	16	25	+6	36	+24
3	15	+1	1	19	0	0	0
4	15	+1	1	20	+1	1	+1
5	14	0	0	19	0	0	0
6	14	0	0	20	+1	1	0
7	14	0	0	18	-1	1	0
8	13	-1	1	19	0	0	0
9	13	-1	1	18	-1	1	+1
10	10	-4	16	16	-3	9	+12
11	8	-6	36	15	-4	16	+24
	154 Total		108 Total Variance	209 Total		66 Total Variance	68 Total

$$2. \quad a. \text{ Test Mean: } \frac{154}{11} = 14$$

$$b. \text{ Job Mean: } \frac{209}{11} = 19$$

$$3. \quad a. \text{ Test Standard Deviation:}$$

$$SD_x = \sqrt{\frac{\text{Variance}}{\text{Number of Cases}}} = \sqrt{\frac{108}{11}}$$

$$SD_x = \sqrt{9.82} = 3.1$$

$$b. \text{ Job Standard Deviation:}$$

$$SD_y = \sqrt{\frac{66}{11}} = \sqrt{6}$$

$$SD_y = 2.4$$

$$4. \quad \text{Product-Moment Correlation } (r) = \frac{\sum(d_x)(d_y)}{N(SD_x)(SD_y)}$$

$$(r) = \frac{68}{(11)(3.1)(2.4)}$$

$$(r) = .83$$

Figure (2)

⁸⁸ See L. TYLER, TESTS AND MEASUREMENTS 16-24 (1963).

⁸⁹ See A. ANASTASI, *supra* note 7, at 104-08.

will perform on the job as foreseen by their test scores. The correlation must be statistically sound rather than a product of chance since, once a test is validated, it can be administered without further study. The Uniform Guidelines establish that a correlation coefficient is statistically significant at the .05 level, which means that the results reported could have been produced by chance not more than five times out of one hundred.⁹⁰

5. Usefulness

Assuming that the correlation coefficient is statistically significant, the next question is whether the test is valid enough to be utilized. That decision is a matter of judgment. Factors to be considered include the impact of the use of the test upon minorities and women, the importance of the job, the need for safety or other reasons to avoid hiring unqualified individuals, and the rate of selection for the job.⁹¹ An example of a rate of selection issue is the use of ranked tests to choose only a few workers from many applicants. Civil service systems in particular rank applicants by their individual test scores and then hire the person with the highest grade. This may be of dubious legitimacy from a testing point of view⁹² in that a standard error of estimate "shows the margin of error to be expected in the individual's predicted criterion score, as a result of the imperfect validity of the test."⁹³ At the other extreme, it may be pointless to administer a test when most applicants will be hired.⁹⁴ Once it is decided to use a test, the final step in the validation study is to establish the appropriate cutoff scores of job performance and test measures. The cost to the employer of false acceptances must be weighed against the cost of false rejections.⁹⁵

⁹⁰ 29 C.F.R. § 1607.14B(5) (1979). The determination of statistical significance is a function of the size of the sample (the number of employees participating in the validation study) and the size of the correlation. The lower the correlation coefficient, the larger the sample size must be to show statistical significance. A. ANASTASI, *supra* note 7, at 108-09. Even where a sample size is as small as 11, as in Figure 2, the high correlation of .83 can be statistically significant. See G. WEINBERG & J. SCHUMAKER, *STATISTICS: AN INTUITIVE APPROACH* 279-81 (2d ed. 1969).

⁹¹ See 29 C.F.R. § 1607.14B(6) (1979).

⁹² Hunt, *Civil Service Testing and Affirmative Action: A Psychologist's Perspective*, 44 U. CINN. L. REV. 690, 693 (1975). Civil Service testing systems mandate employment based on merit. *Id.* at 690.

⁹³ A. ANASTASI, *supra* note 7, at 165.

⁹⁴ *Id.* at 171. If the selection ratio (the proportion of applicants who must be accepted) is 100%, "no test, however valid, could improve the selection process." *Id.*

⁹⁵ A. ANASTASI, *supra* note 7, at 167-69.

In *Jackson v. Nassau County Civil Service Commission*,⁹⁶ plaintiffs had been hired as provisional community service assistants, a position not covered by the Civil Service system. It was subsequently resolved to make the jobs permanent and therefore subject to Civil Service testing requirements. As a result of their test marks, the plaintiffs lost their jobs or were demoted, despite their proven ability to do the work. Those hired in their stead did achieve higher test scores but had not yet shown capability to perform the duties for which they were selected.⁹⁷

The court in *Jackson* failed to deal with several issues. The replacement of well qualified incumbents, black and white, with whites less qualified but for their test scores, should have been analyzed in the first instance in terms of the utility of imposing the test system. The Civil Service test had used a two level scoring procedure. To pass the exam an applicant was required to receive a minimum score of 60%.⁹⁸ This entitled the test-taker to be ranked by test score. Actual employment decisions were made strictly according to that ranking. The court upheld the cut-off mark of 60% and found no disparate impact in the rate of passing.⁹⁹ Although 88% of the whites passed as compared to 73% of the blacks, the difference did not violate the "Rule of Four-Fifths."¹⁰⁰ The Uniform Guidelines state the rule, in pertinent part, as follows: "[a] selection rate for any race, sex, or ethnic group which is less than four-fifths . . . of the rate for the group with the highest rate will generally be regarded . . . as evidence of adverse impact" ¹⁰¹ However, the court did not discuss the appropriateness of interpreting test data on the basis of a passing score which played no role in the actual employment decisions.¹⁰² This raises a possibility that employers who impose a cutoff score pegged at the level where no disparate impact exists may then

⁹⁶ 424 F. Supp. 1162 (E.D.N.Y. 1976).

⁹⁷ *Id.* at 1164-65. Eight of the ten plaintiffs received passing grades on the exam, but their scores were not high enough to qualify for appointment to a permanent position. *Id.*

⁹⁸ *Id.* at 1167.

⁹⁹ *Id.* Seventeen applicants who achieved 90% or better were offered the jobs. *Id.* at 1168.

¹⁰⁰ *Id.* at 1167. See 29 C.F.R. § 1607.4D (1979).

¹⁰¹ 29 C.F.R. § 1607.4D (1979). The court refused to find a violation of the four-fifths rule because blacks passed the exam "at a rate better than 4/5 that of the passing rate for whites." 424 F. Supp. at 1167. Cf. *Vulcan Soc'y of N.Y. City Fire Dept. Inc. v. Civil Serv. Comm'n*, 490 F.2d 387 (2d Cir. 1973) (qualifying exam for firemen found to have discriminatory impact because white candidates passed at rate more than twice that of minority applicants). See generally Shoben, *Differential Pass-Fail Rates in Employment Testing: Statistical Proof Under Title VII*, 91 HARV. L. REV. 793 (1978).

¹⁰² See 424 F. Supp. at 1167-68.

manipulate the test results for actual employment decisions without regard to any impact on groups protected by Title VII.¹⁰³

B. Content Validation

Content validation is authorized where the critical elements of the job can be replicated in a test so that test-takers are in fact subjected to a sample of the work for which they are applying. "An examination has content validity if the content of the examination matches the content of the job."¹⁰⁴ The typical example is a typing test for a typist job.¹⁰⁵ The Uniform Guidelines expand the domain of content validation to defined knowledge, skills or abilities which are part of the actual content of the job.¹⁰⁶

A case that appears to have foreshadowed the new guidelines is *Arnold v. Ballard*,¹⁰⁷ where the court held that a high school education requirement for employment in a city police department was valid under a content validation approach.¹⁰⁸ Under the earlier framework, only criterion validation would have been proper since achievement of a high school diploma is not a sample of police officers' functions. Defendant's expert performed a job analysis which indicated that police department bulletins were geared toward a twelfth-grade reading level and that as much as 15% of the work time involved preparing reports.¹⁰⁹ However, where the job analysis showed reading and writing to be elements of the job, a test expert

¹⁰³ The court also failed to discuss the standard error of estimate issue. See text accompanying note 93. The impact of the decisions to appoint weighed much more heavily upon blacks than whites since only one of the seventeen appointees was black. If two black applicants had each answered one more question correctly on the test, they would have achieved scores of 90%, the appointment score. Instead of questioning whether there was any meaningful difference among these closely ranked candidates because of the standard error of estimate, the court found that no disparate impact would have been made out if those two black applicants had actually scored 90% by each getting one more answer correct. 424 F. Supp. at 1168.

¹⁰⁴ *Vulcan Soc'y of N.Y. City Fire Dept. Inc. v. Civil Serv. Comm'n*, 360 F. Supp. 1265, 1274 (S.D.N.Y.), *aff'd*, 490 F.2d 387, 395 (2d Cir. 1973); see 29 C.F.R. § 1607.14C(1) (1979).

¹⁰⁵ *Kirkland v. New York State Dept. of Correctional Serv.*, 374 F. Supp. 1361, 1371 n.9 (S.D.N.Y. 1974), *aff'd in part, rev'd in part*, 520 F.2d 420 (2d Cir. 1975). "It is essential that the examination test these attributes both in proportion to their relative importance on the job and at the level of difficulty demanded by the job." *Id.* at 1372 (quoting *Vulcan Soc'y of N.Y. City Fire Dept. Inc. v. Civil Serv. Comm'n*, 360 F. Supp. 1361, 1374 (S.D.N.Y. 1974), *aff'd in part, rev'd in part*, 520 F.2d 420 (2d Cir. 1975)).

¹⁰⁶ 29 C.F.R. § 1607.14C(4) (1979). The distinction between job elements subject to criterion validation and those subject to content validation is almost eliminated under the Guidelines. See *id.*

¹⁰⁷ *Arnold v. Ballard*, 390 F. Supp. 723 (N.D. Ohio 1975).

¹⁰⁸ *Id.* at 728.

¹⁰⁹ *Id.* at 727-28.

should have used reading and writing samples from the job itself in order to rely on content validation. Even under the new Uniform Guidelines there is some doubt as to the *Arnold* validation because it was not demonstrated that those with high school diplomas knew how to perform the reading and writing necessary for their employment.¹¹⁰

In *United States v. City of Chicago*,¹¹¹ the district court had upheld the validity of a written promotional examination for Chicago firemen because a job analysis had been conducted.¹¹² The Seventh Circuit reversed because there was no showing that the test's content closely approximated the critical elements of the job.¹¹³ Further, the court concluded that "there must be a correlation between the importance of a job function as determined by the job analysis and the weight given to this function on the examination."¹¹⁴

Where the job analysis indicates that a critical part of the job can be incorporated into a test, the test expert makes an assessment of the similarity between the test and the job. The power of that connection is a matter of rational rather than empirical verification, since content validation typically does not involve a sample of employees who take the test and whose subsequent job performance is compared statistically with their test scores.¹¹⁵ The Uniform Guidelines instruct that "[t]he closer the content and the context of the selection procedure are to work samples or work behaviors, the stronger is the basis for showing content validity."¹¹⁶

¹¹⁰ See 29 C.F.R. § 1607.14C(6) (1979). A prerequisite of certain previous training "based on content validity . . . should be justified on the basis of the relationship between the content of the training . . . and the content of the job" being applied for. *Id.*

¹¹¹ *United States v. City of Chicago*, 573 F.2d 416 (7th Cir. 1978).

¹¹² *Id.* at 419-20.

¹¹³ *Id.* at 425. See *Firefighters Institute for Racial Equality v. City of St. Louis*, 549 F.2d 506 (8th Cir. 1977) (promotional examination which did not measure supervisory skills not content valid).

¹¹⁴ 573 F.2d at 426.

¹¹⁵ For a discussion of criterion-related validation, see notes 69-77 *supra* and accompanying text.

¹¹⁶ 29 C.F.R. § 1607.14C(4) (1979). The Uniform Guidelines further provide as follows: As the content of the selection procedure less resembles a work behavior, or the setting and manner of the administration of the selection procedure less resemble the work situation, or the result less resembles a work product, the less likely the selection procedure is to be content valid, and the greater the need for other evidence of validity.

Id.

Content validation is attractive because it does not involve the complexities or expenses of criterion validation.¹¹⁷ The principle danger is that a test which intuitively appears related to the job in question will be accepted as content valid. This intuitive type of approach is called "face validity,"¹¹⁸ and "refers, not to what a test actually measures, but to what it appears superficially to measure."¹¹⁹

*Coopersmith v. Roudebush*¹²⁰ is an example of a court accepting as content valid a test that had only face validity.¹²¹ In this case, a woman with a law degree was rejected for a job as an attorney with the Veterans Administration, partly because of alleged poor performance on a legal writing test. The exam was based on an actual case history and the applicant was asked to write an opinion with supporting reasons.¹²² Although the test was similar to work entailed in the job, there was no indication that a job analysis had been undertaken. Further, there was no evidence of how closely the test content and its administration actually fit the job.¹²³

Content validation is aimed at situations where the test is actually a trial run at critical parts of the job. Therefore, the exam must not require skills which are not necessary to perform the work. The problem typically occurs where the employment does not require verbal skills but the test does. *Frontera v. Sindell*¹²⁴ illustrates this dilemma. Plaintiff, of Spanish origin, had been a temporary employee performing carpentry work which did not require much knowledge of English. To qualify for a permanent position as a carpenter, plaintiff took a civil service test which asked him to demonstrate his skills, the kind of exam that could be content validated. The test, however, was administered in English. Plaintiff claimed that he failed the test because he was unable to understand and follow the test directions, not because he was unable to perform the carpentry work.¹²⁵ Since the case was not brought under Title VII,¹²⁶ the Sixth Circuit upheld the

¹¹⁷ See notes 50-116 *supra* and accompanying text.

¹¹⁸ A. ANASTASI, *supra* note 7, at 139.

¹¹⁹ *Id.* Face validity, therefore, is not validity in a technical sense. It should never be considered a replacement for actual "objectively determined validity." *Id.* at 140.

¹²⁰ 517 F.2d 818 (D.C. Cir. 1975).

¹²¹ No disparate impact was shown in *Coopersmith*; the discussion of test validity was dicta. See 517 F.2d at 823-24.

¹²² *Id.* at 820.

¹²³ See *id.* at 823-24.

¹²⁴ 522 F.2d 1215 (6th Cir. 1975).

¹²⁵ *Id.* at 1216.

¹²⁶ Plaintiff claimed a violation of his Civil Rights under 42 U.S.C. §§ 1981, 1983 and 1985. *Id.*

test without basing its affirmation upon standards established by the Act for content validation. However, the court recognized that the words and phrases used in the exam, which Frontera could not comprehend, would ordinarily be understood by a person familiar in the field, so that "such terminology would not . . . interfere with the test's objective of identifying competent carpenters and ranking them for Civil Service."¹²⁷ From a testing point of view, if the ordinary English carpentry terms were not used in the job, the test and its administration should not rely upon them.

C. Construct Validation

The construct validity of a test is the extent to which the test measures a theoretical construct or trait. It involves a "more abstract kind of behavioral description"¹²⁸ than content and criterion validity. Examples of such constructs are intelligence, mechanical comprehension, verbal fluency, speed of walking, neuroticism, and anxiety.¹²⁹ In an examination for the position of typist, "[a] construct valid approach would identify certain traits essential to success as a typist, such as ability to concentrate, perseverance and attention to detail, and would examine the applicant for those traits."¹³⁰

In the job analysis for construct validation, as in content or criterion-related validation, the critical work behaviors for the job must be identified. However, the job analysis for construct strategy must further locate the mental traits which underlie successful performance of these critical elements of the job.¹³¹ In view of the fact that construct validation is not often utilized in the employment context¹³² because it is a relatively new procedure, the Uniform Guidelines caution users to proceed carefully: "the effort to obtain sufficient empirical support for construct validity is both an extensive and arduous effort involving a series of research studies, which in-

¹²⁷ *Id.* at 1219. The implication is that the test would have been valid under a Title VII inquiry as well. *See id.*

¹²⁸ A. ANASTASI, *supra* note 7 at 151.

¹²⁹ *Id.*

¹³⁰ *Kirkland v. New York State Dept. of Correctional Serv.*, 374 F. Supp. 1361, 1371 n.9 (S.D.N.Y. 1974), *aff'd in part, rev'd in part*, 520 F.2d 420 (2d Cir. 1975); *see Vulcan Soc'y of the N.Y. City Fire Dept. Inc. v. Civil Serv. Comm'n*, 490 F.2d 387, 395 (2d Cir. 1973).

¹³¹ 29 C.F.R. § 1607.14D(4)(a) (1979).

¹³² *See* 96 L.R.R. 197, 198 (1977). The exception is the Civil Service Commission, which utilizes this strategy to validate almost all examinations. *Id.* *See generally* *Washington v. Davis*,

clude criterion related validity studies and which may include content validity studies.”¹³³

VI. CONCLUSION

The Uniform Guidelines provide that any of the validation strategies is acceptable in the proper context.¹³⁴ Although the earlier EEOC Guidelines required the use of criterion-related validation unless that was technically not feasible, the Supreme Court made clear in *Washington v. Davis* that any appropriate technique was permissible.¹³⁵ While the presumption of the EEOC Guidelines has been eliminated, the same presumption is still reflected in the professional standards of industrial psychology.¹³⁶ The reason professional test standards emphasize criterion validation is because it is based on the comparison of two sets of empirical data—test scores and job performance scores—which result in a number, the correlation coefficient, a mathematical expression of the relationship between test scores and job performance. In contrast, content validation begins with one empirical basis—job analysis data—and requires the test expert to argue rationally that the examination in fact is a sample of critical job behaviors shown in the analysis.¹³⁷

Employers have increasingly relied on content validation since it is easier to validate from a testing point of view and provides less ammunition for plaintiffs to attack the validation study in Title VII litigation. The key element in content as well as criterion validation remains the job analysis. Content validation should be utilized only where the analysis reveals that critical tasks are required which cannot be simply learned on the job and which can be measured by a test that is a sample of those tasks.

426 U.S. at 247 n.13; *United States v. City of Chicago*, 573 F.2d at 426 n.10.

¹³³ 29 C.F.R. § 1607.14D(1) (1979). The Uniform Guidelines warn that “there is at present a lack of substantial literature extending the concept to employment practices.” *Id.*

¹³⁴ 29 C.F.R. § 1607.5A (1979).

¹³⁵ 426 U.S. at 247 and n.13.

¹³⁶ A. ANASTASI, *supra* note 7, at 108.

¹³⁷ See *Vulcan Soc’y of N.Y. City Fire Dept. Inc. v. Civil Serv. Comm’n*, 490 F.2d 387, 395; *Kirkland v. New York State Dept. of Correctional Serv.*, 374 F. Supp. 1361, 1370 (S.D.N.Y. 1974), *aff’d in part, rev’d in part*, 520 F.2d 420 (2d Cir. 1975). See generally *Western Addition Community Organization v. Alioto*, 360 F. Supp. 733, 736 n.5 (N.D. Cal. 1973), *appeal dismissed*, 514 F.2d 542 (9th Cir. 1975); A. ANASTASI, *supra* note 7, at 134-61; L. TYLER, *TESTS AND MEASUREMENTS* 16-24 (1963).