3 Legal Issues in Hiring and Promotion of Police Officers

Arthur Gutman

INTRODUCTION

Allegations of discrimination against police departments span a variety of issues, including refusal to hire minorities and females (i.e., facial discrimination), adverse impact of facially neutral selection criteria on minorities (e.g., cognitive tests) and females (e.g., physical requirements), racial and sexual harassment, diversity as a compelling government interest under the 14th Amendment of the Constitution, age limits for hiring and forced retirement of police officers, illegal medical and psychological inquiries prior to a conditional job offer, and more. Comprehensive coverage of these and related issues is provided by Gutman, Koppes, and Vodanovich (2010). The main focus of this chapter is on what is arguably the most important issue facing police forces today: adverse impact in hiring and promotion related to tests and other selection criteria on minorities and, to some extent, females.

As a starting point, adverse impact based on race, sex, religion, or national origin is uniquely associated with Title VII of the Civil Rights Act of 1964 (or simply Title VII).* There are three phases in the Title VII adverse impact judicial scenario. As depicted in Table 3.1, the plaintiff must identify a selection practice that disproportionately excludes one group (e.g., African Americans) relative to another group (e.g., Caucasians) in Phase 1. If the plaintiff succeeds, the defendant must prove that the challenged practice is job related and consistent with business necessity in Phase 2. Then, if the defendant succeeds, the plaintiff must prove that there is an equally valid alternative practice that produces less or no adverse impact in Phase 3.

In general, the most frequent method for assessing knowledge, skills, and abilities (or KSAs) related to police work are written multiple-choice tests because they can be administered to large groups and are easy to score. However, there are other methods of assessing these KSAs, including oral interviews, situational judgment tests, and assessment center methodology, to name a few. These other methods are often used in conjunction with written tests. Indeed, a common claim in adverse impact cases is that nonwritten tests, or different weightings among written versus nonwritten components, are as valid as the actual method(s) used. The discussion below treats all methods of assessing KSAs as a “test” subject to adverse impact analyses.

An additional point to note is that there are common issues relating to hiring and promotion of police officers as there are with firefighters. Although the focus of this chapter is on police hiring and promotion, firefighter cases will be discussed when relevant to police issues. For example, as we will witness below, the U.S. Supreme Court’s 2009 ruling in Ricci v. DeStefano relates to promotion of firefighters. Nevertheless, this ruling has equally important implications for police departments and other entities as well.

There are six sections below. The first section reviews major Supreme Court rulings on adverse impact between 1971 and 1989, and the second section reviews early lower court rulings that were

* Adverse impact is also a valid claim in the Age Discrimination in Employment Act of 1967 (ADEA) based on Supreme Court rulings in Smith v. City of Jackson (2005) and Meacham v. KAPL (2008). However, the rules for deciding ADEA cases are dramatically different as compared to those for Title VII cases.
based on early Supreme Court rulings. The third section features more recent rulings on cutoff scores and related issues, the fourth section features methods of reducing or eliminating adverse impact, the fifth section focuses on the Supreme Court’s recent ruling in *Ricci v. DeStefano* (2009), and the sixth section offers conclusions and recommendations.

### OVERVIEW OF SUPREME COURT ADVERSE IMPACT RULINGS

Table 3.2 depicts nine landmark Supreme Court rulings between 1971 (*Griggs v. Duke Power*) and 2009 (*Ricci v. DeStefano*). Five of the nine cases feature objectively scored written tests, and four of them feature other causes of adverse impact. These topics will be discussed in turn immediately below. For purposes of exposition, *Ricci v. DeStefano* (2009) will be reserved for discussion in the fifth section of this chapter.

### ADVERSE IMPACT AND WRITTEN TESTS

Early precedents for written tests were established in *Griggs v. Duke Power* (1971) and *Albemarle v. Moody* (1975), and the principal rulings in these two cases served as the basis for the *Uniform Guidelines on Employee Selection Procedures* (UGESP) in 1978. *Washington v. Davis* (1976),

### TABLE 3.1

The Title VII Adverse Impact Judicial Scenario

<table>
<thead>
<tr>
<th>Phase</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Prima facie phase requiring statistical evidence that an identified employment practice disproportionately excludes protected group members</td>
</tr>
<tr>
<td>2</td>
<td>Defense phase requiring proof that the challenged practice is job related and consistent with business necessity</td>
</tr>
<tr>
<td>3</td>
<td>Pretext phase requiring proof there is an alternative equally valid practice (or practices) that are equally job related, but that produce less or no adverse impact</td>
</tr>
</tbody>
</table>

### TABLE 3.2

Landmark Supreme Court Adverse Impact Rulings

<table>
<thead>
<tr>
<th>Challenges to Written Exams</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Albemarle v. Moody</em> (1975)</td>
</tr>
<tr>
<td><em>Ricci v. DeStefano</em> (2009)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Challenges to Other Selection Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Dothard v. Rawlinson</em> (1977)</td>
</tr>
<tr>
<td><em>NYC v. Beazer</em> (1979)</td>
</tr>
<tr>
<td><em>Wards Cove v. Atonio</em> (1989)</td>
</tr>
</tbody>
</table>
the first Supreme Court adverse impact ruling involving police officers, addressed the question of whether adverse impact is a valid claim under the 5th or 14th Amendments of the Constitution. *Connecticut v. Teal* (1982), the first Supreme Court Title VII claim against a government agency, addressed the question of whether a Phase 2 defense to adverse impact is required for individual steps in a multiple-hurdle selection procedure when there is no bottom-line adverse impact for the total selection process.

**The Griggs and Albemarle Rulings**

*Griggs* and *Albemarle* featured private sector employers that used cognitive tests and a high school diploma requirement in their selection process. The facts in *Griggs* were that prior to Title VII, Duke Power facially excluded African Americans from upper-level jobs. The company then initiated the testing and diploma requirements for these upper-level jobs on July 2, 1965, the very day that Title VII took effect. The prima facie case featured statistical evidence that the cognitive tests excluded 94% of African American applicants as compared to only 42% of Caucasian applicants. Additionally, the high school diploma requirement was deemed to disproportionately “chill” prospective African American applicants because the graduation rate in North Carolina at that time was 34% for Caucasians as compared to only 12% for African Americans.

Speaking for a unanimous Supreme Court, Justice Burger ruled that Title VII covers the “consequences of employment practices, not simply the motivation” of employers. He then wrote what arguably are the two most important phrases in adverse impact case law: that if the plaintiff proves adverse impact, the defendant must prove there is a *manifest relationship* between the challenged practice and the *employment in question*. There was no evidence to prove the cognitive tests and high school diploma were job related, and Duke Power lost the case.

Aware of the *Griggs* ruling, the Albemarle Paper Company tried to prove that their cognitive tests were significantly and positively correlated with job performance (i.e., using a criterion validity study). However, it was a poor study conducted a month prior to the trial. The Supreme Court struck down the study on four grounds: (1) a lack of quality, or “odd patchwork”; (2) unknown job-performance criteria and subjective supervisory rankings; (3) a focus on high-level jobs rather than the “entering low-level jobs” at issue; and (4) a validation sample that included only “job-experienced white workers.” Relying on the then applicable 1970 *Guidelines on Employee Selection Procedures* by the Equal Employment Opportunity Commission (EEOC), the Supreme Court defined how a manifest relationship should be proven. Accordingly:

> The message of these Guidelines is the same as that of the Griggs case—that 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 evaluated. [emphasis by author]

The *Albemarle* Court then created the third phase of the adverse impact scenario, permitting proof of equally valid alternatives with less or no adverse impact.

**The Immediate Aftermath of Griggs and Albemarle**

The *Griggs* and *Albemarle* rulings were then codified into regulatory law in the UGESP in 1978. The UGESP are still the most important regulatory authority used by courts in adverse impact cases. However, as noted by several authors,* the UGESP were based primarily on the first edition of the *Standards for Educational and Psychological Testing* (or Standards) (1974), and therefore are now outdated. Critically, this was anticipated, and is reflected in Section 1607.5(A) of the UGESP, which states, “New strategies for showing the validity of selection procedures will be evaluated as they become accepted by the psychological profession.”

*See, for example, Landy (1986) and Binning and Barrett (1989).*
As expected, new strategies in validity research have been developed since 1978, as documented in the fifth edition of the Standards (1999) and the fourth edition of the Principles for the Validation and Use of Personnel Selection Procedures (or Principles) published by the Society for Industrial and Organization Psychology (SIOP) (2003). Courts have cited both of these authorities with increasing frequency in recent years. Therefore, test developers and consumers should be equally as familiar with the Standards and Principles as they are with the UGESP.*

The Davis and Teal Rulings

Prior to 1972, Title VII covered only private entities. Coverage of public entities was added in the Equal Employment Act of 1972 (or EEO-72). Consequently, for the plaintiffs in Washington v. Davis (1976), the 5th Amendment was the only basis for challenging Civil Service Test 21, a verbal skills test first used in 1970 by the Washington, D.C., Police Department for entry-level police jobs. The Supreme Court ruled that adverse impact is not a valid claim under the 5th Amendment, or other Constitutional amendments.† Nevertheless, the Supreme Court evaluated Test 21 and ruled it was valid because it (1) measured minimum skills necessary for completing police training school, and (2) criterion validity was established with proof that scores on Test 21 were significantly and positively correlated with scores on written exams given during police training. Also of interest, a high school diploma requirement, deemed invalid for coal-mining and paper mill jobs in the Griggs and Albemarle cases, was deemed valid for police officer jobs in the Washington v. Davis case.

In Connecticut v. Teal (1982), a written test was the first of several hurdles for promotion. After all hurdles were completed, the promotion rate was higher for African Americans (22.9%) than Caucasians (13.5%). However, there was adverse impact for African Americans on the first hurdle. In its defense, the State of Connecticut relied on a provision in Section 1607.4(C) of the UGESP relating to adverse impact of the “total selection process”:

If...the total selection process does not have an adverse impact, the [EEOC]...will not expect a user to evaluate the individual components for adverse impact, or to validate such individual components, and will not take enforcement action based upon adverse impact in any component of that process. [emphasis by author]

Rejecting this guidance in a closely divided 5–4 ruling, the Supreme Court favored the plaintiffs on grounds that “Congress never intended to give an employer license to discriminate against some employees on the basis of race or sex merely because he favorably treats other members of the employee’s group.” This remains a critical ruling for police departments because multiple selection criteria are often used in both entry-level and promotion processes.

Adverse Impact and Other Selection Practices

Among the four cases in Table 3.2 featuring causes of adverse impact other than written tests, Dothard v. Rawlinson (1977) addressed height and weight requirements for prison guards, and New York City v. Beazer (1979) addressed exclusion of transit authority police officers based on past methadone use. The other two cases, Watson v. Fort Worth Bank (1988) and Wards Cove Packing Co. v. Atonio (1989), addressed subjective cause of adverse impact. As important, Watson and Wards Cove were closely connected to each other and led to a temporary alteration of the Phase 2 defense to adverse impact that was subsequently resolved in the Civil Rights Act of 1991 (or CRA-91).

* Readers interested in test validation developments are referred to a comprehensive discussion of the UGESP, Standards, and Principles by Jeanneret (2005).

† The Supreme Court subsequently ruled that adverse impact is not a valid claim under the Equal Protection Clause of the 14th Amendment in Arlington Heights v. Metropolitan Housing Corp. (1977) and Personnel Administrator v. Feeney (1979).
The **Dothard and Beazer Rulings**

In *Dothard v. Rawlinson* (1977), minimum height and weight criteria for selection of prison guards adversely impacted female applicants. The State of Alabama argued that these criteria were indicators of strength, an important job-related KSA. However, the Supreme Court struck down the requirement, ruling, “If the job-related quality that the appellants identify is bona fide, their purpose could be achieved by adopting and validating a test for applicants that measures strength directly.”

The *Dothard* ruling has directly affected police departments. For example, in *Horace v. Pontiac* (1980), featuring a height requirement, the city of Pontiac, Michigan, asserted, among other things, that taller police officers meet “less resistance” and gain “greater respect … from the general public.” However, the 6th Circuit ruled that “there were acceptable alternative policies which would have accomplished the desired purposes other than the obviously arbitrary 5’ 8″ height standard.” Based on *Dothard*, municipalities, including the city of Pontiac, subsequently turned to other selection criteria to assess strength and agility.

In *New York City v. Beazer* (1979), the Supreme Court supported exclusion of methadone users from employment as transit authority officers because it is “obvious” that drug addiction threatens the “legitimate employment goals of safety and efficiency.” Coupled with earlier lower court rulings involving public safety concerns (e.g., *Spurlock v. United Airlines*, 1972), the *Beazer* ruling made it easier for police departments to defend minimum requirements relating to educational requirements, driving records, and prior drug convictions.

For example, in *Davis v. Dallas* (1985), the 5th Circuit found it was necessary for the Dallas police department to exclude recent drug users from police work on grounds that recent drug use shows a disregard for the law. Additionally, exclusion based on poor driving records was supported based on research indicating that past driving habits predict future driving habits, and a requirement of 45 college credits with C or better grades was supported based on federal commission reports in the 1960s that “a high school education is a bare minimum requirement for successful performance of the policeman’s responsibilities.”

The **Watson and Wards Cove Rulings**

In *Watson v. Fort Worth Bank* (1988), an African American woman was passed over for promotion four times, each time in favor of a Caucasian applicant, and each time based on subjective ratings by Caucasian supervisors of (1) job performance, (2) interview performance, and (3) past experience. It was not clear how these ratings were scored or combined, but there was clearly “bottom-line” adverse impact for the “total selection process.” Only eight justices heard this case, and they unanimously agreed that adverse impact based on subjective selection criteria is a valid claim.† However, speaking for herself and three other justices, Justice O’Connor proposed changes in the adverse impact scenario. The most important proposal was to weaken the defense burden in Phase 2 so that the defendant need only articulate (or explain) a legitimate business reason for the challenged practices, instead having to prove that they are job related.‡

In *Wards Cove v. Atonio* (1989), two Pacific Northwest salmon packing companies had a hiring-hall arrangement for selection of unskilled salmon packers, but used word-of-mouth to hire workers

---

*Interestingly, although the State of Alabama could not defend on height and weight criteria in the adverse impact challenge, it succeeded with a BFOQ (Bona Fide Occupational Qualification) defense in excluding all females on grounds that 20% of the population were sex offenders, and it was reasonably necessary to exclude all females because of potential danger to the prison environment.† At the time, Anthony Kennedy’s nomination to the Supreme Court was affirmed, but he was not yet seated for the *Watson* case.‡ The burden to articulate (or explain) without having to prove is termed a burden of “production” and is characteristic of disparate treatment cases such as *McDonnell Douglas v. Green* (1973). It is generally easier to explain a legitimate business reason than to prove job relatedness. These are complex issues, and the reader is referred to a more comprehensive discussion of these issues by Gutman et al. (2010).*
for more skilled positions. As a result, Eskimos and Filipinos were overrepresented in the unskilled jobs and underrepresented in the skilled jobs. It is arguable that *Wards Cove* was more of a pattern or practice case than it was an adverse impact case. More important for present purposes, with the addition of Anthony Kennedy to the Supreme Court, there was now a majority of five justices that turned O’Connor’s proposal in the *Watson* case into case law.

**The Aftermath of *Wards Cove***

Congress attempted to overturn *Wards Cove* (and five other 1989 Supreme Court rulings) in the Civil Rights Restoration Act of 1990 (C CRA-90). However, President George H. W. Bush vetoed CRRA-90, and his veto was nearly overridden. The main reason for the veto was failure by the Democrats and Republicans to agree on the Phase 2 defense burden in *Wards Cove*. Given the close call in CRRA-90, the two parties compromised in CRA-91, restoring the defense burden originally established in the *Griggs* and *Albemarle* cases so that if an identified practice is proven to cause adverse impact in Phase 1, the defendant must prove that the practice is *job related and consistent with business necessity* in Phase 2. Although seemingly a recovery of the original *Griggs* and *Albemarle* precedents, as we will witness later in this chapter, at least two circuit courts have parsed the phrases *job related* and *consistent with business necessity* so that more is needed than proof of test validity to establish cutoff scores.

**EARLY LOWER COURT RULINGS ON ADVERSE IMPACT**

The *Griggs* ruling was rendered prior to the extension of Title VII coverage to public entities in EEO-72. Having no other avenue for lawsuits prior to 1972, plaintiffs in municipal jobs challenged adverse impact via the Equal Protection Clause of the 14th Amendment. There were several lower court rulings in police and firefighter cases in which *Griggs* was affirmed under the 14th Amendment. Although the Supreme Court ultimately ruled that adverse impact is *not* a valid constitutional claim in *Washington v. Davis*, these pre-*Davis* rulings established important precedents for adverse impact challenges that influenced post-*Davis* Title VII rulings. Table 3.3 samples five pre-*Davis* 14th Amendment rulings and three post-*Davis* Title VII rulings.

**PRE-*DAVIS* 14TH AMENDMENT RULINGS**

Among the pre-*Davis* rulings sampled in Table 3.3, two featured fire departments, two featured police departments, and one featured a correctional institution. All but one of these cases (*Kirkland*) originated prior to EEO-72. The circuit courts supported the plaintiff’s 14th Amendment claims based on *Griggs* in each case.

**Carter v. Gallagher (1971)**

*Carter v. Gallagher* was the first post-*Griggs* 14th Amendment ruling. The plaintiffs argued that historical discriminatory practices in Minneapolis, Minnesota, created a nearly all-white fire department. Nested within this claim was an adverse impact charge focused on an entrance exam administered to 2,404 applicants over a 20-year period. Historically, Caucasians passed at rates varying from 40% to 64.7%. Although there were only 22 identifiable minority applicants over that period (six of whom passed), the district court trial judge ruled that minority applicants did “substantially less well” than Caucasians, and there were many minority applicants who took the exam

---

* For example, in *International Teamsters v. United States* (1977), a landmark pattern or practice case, African Americans and Hispanics were congregated in lower paying (short distance) bus-driving jobs and Caucasians were congregated in higher paying (longer distance) bus-driving jobs. Gutman et al. (2010) suggest that pattern or practice rules were also applicable to *Wards Cove*, because minorities were congregated in the salmon-packing jobs and Caucasians were congregated in the more professional jobs.
but could not be identified. The judge struck down the exam, saying, “No effort was ever made prior to the current examination period to analyze the fire fighter examinations to determine whether they were culturally biased or whether they were valid predictive instruments for use in selecting fire fighters.”

The 8th Circuit affirmed this ruling, citing a passage from Griggs that states “what is required by Congress is the removal of artificial, arbitrary, and unnecessary barriers to employment when the barriers operate invidiously to discriminate on the basis of racial or other impermissible classification.”

**Castro v. Beecher (1972)**

Castro v. Beecher was the first post-Griggs police case. The case featured a 100-item multiple-choice hiring test measuring word knowledge, numerical sequence, reading comprehension, reasoning, arithmetic, and analogy. The test had a passing score of 70, and the passing rate was 65% for Caucasians as compared to 25% for African Americans and 10% for Hispanics. The trial judge ruled that the passing score was arbitrary and the test’s “emphasis on academic and verbal skills [had] little relation to a policeman’s job.” The judge also ruled:

> There is almost no effort to test the particular types of observation, memory, statement, and judgment which are useful in connection with the exercise of authority of all kinds, and which are necessary for policemen as guardians, administrators, witnesses, community counselors, social welfare aids, and representatives of a regime of law and order.

Consistent with the 8th Circuit ruling in Gallagher, the 1st Circuit supported the trial judge based on Griggs, ruling that a test that causes adverse impact must be “substantially related to job performance.”

**Bridgeport Guardians v. Bridgeport CSC (1973)**

Bridgeport Guardians v. Bridgeport CSC featured challenges to written exams for hiring and promotion of police officers. The test was developed years earlier by a private company. The trial judge ruled that there was adverse impact for the hiring exam but not the promotion exam. The trial judge
struck down the hiring exam based on Griggs and the 8th Circuit’s ruling in Beecher, and the 2nd Circuit affirmed, ruling that the biggest problem with the hiring exam was that “many of the vocabulary and arithmetic questions [were] only superficially or peripherally related to police activity.”

**Vulcan Society v. CSC (1973)**

*Vulcan Society v. CSC* featured a 100-item entry-level test for firefighters that was deemed flawed by the 2nd Circuit based on the EEOC’s then applicable *Guidelines on Employee Selection Procedures*. Although there were other issues in this case,* the primary focus was on failure to establish content validity in accordance with the EEOC Guidelines. The 2nd Circuit ruled that the test was developed in an “unprofessional manner.” Among the criticisms noted was that the CSC failed to do a job analysis, the test contained 20 items on civics-related issues that were deemed unrelated to firefighting, and the CSC eliminated a physical exam that was used in prior selection procedures. As we will witness shortly, the *Vulcan* ruling was a precursor to *Guardians v. Civil Service* (1980), in which the 2nd Circuit established important precedents for content validity.

**Kirkland v. NY State Corrections (1975)**

*Kirkland v. NY State Corrections* featured Test 34-944, used by the New York State Department of Civil Service for promotion of correctional officers to sergeant. There were 1,389 applicants across the state, and the passing rate was 30.8% for Caucasians as compared to 7.7% for African Americans and 12.5% for Hispanics. As in its prior rulings in the *Bridgeport Guardians* and *Vulcan* cases, the 2nd Circuit, citing *Griggs*, ruled:

> Proof in employment discrimination cases proceeds from effect to cause. Plaintiffs establish the racially disparate consequences of defendants’ employment practices, and defendants must then justify such consequences on constitutionally acceptable grounds. [emphasis added]

Thus, the 2nd Circuit treated the Phase 2 defense outlined in *Griggs* as the “constitutionally acceptable grounds” for defending adverse impact.

**Post-Davis Title VII Rulings**


**NAACP v. Seibels (1980)**

In *NAACP v. Seibels*, written exams were used for entry-level police and firefighter jobs. The exams were correlated with three criteria, most notably training academy performance, for which there were positive and statistically significant correlation coefficients as in *Washington v. Davis*. However, the test in *Davis* was designed to measure minimum skills for succeeding in academy school, whereas the test in *Seibels* was used for strict rank for actual hiring decisions. The 5th Circuit struck down the *Davis* defense, ruling:

> We do not believe the Davis rationale can be extended, as the Board urges, to the general proposition that any test can be validated by showing a relationship to training. More specifically, we reject the Board’s suggested extension of the Davis holding to this case, where the tests were not used to ascertain the minimum skills necessary to complete job-relevant training, but rather were used to rank job applicants according to their test scores and to select only the highest test scorers for job placement. [emphasis added]

* The CSC argued that content validity was unnecessary because their test predicted training school performance. However, there was no evidence of statistical correlations to prove the relationship, as there was subsequently in *Washington v. Davis* (1976).
As important for later cases, the correlations for the other two criteria (efficiency ratings and experimental ratings) were either mixed (i.e., both positive and negative correlations) or not statistically significant. Critically, one of the tests (10-C for firefighters) was rejected even though the correlation was positive (+.21) and statistically significant because it lacked practical significance. Accordingly:

With respect to the 10-C test, the court found that there is a statistically significant correlation between test scores and experimental ratings, but that the correlation is of very low magnitude and lacks practical significance. [emphasis added]

Other courts have since affirmed the need for practical significance in criterion validity studies, requiring, in many cases, correlation coefficients of +.30 or higher.*


Guardians v. CSC was a landmark case that addressed content validity as a basis for strict rank ordering and cutoff scores. In-house personnel in the New York City CSC developed an entry-level police test, and based on test scores, applicants were rank ordered, and a passing score sufficient to generate the required number of potential trainees was established. The plaintiffs challenged content validity as a basis for rank ordering, as well as the cutoff score chosen. The fact that the test was developed in-house ultimately worked to the disadvantage of the CSC.

The challenge to content validity and rank ordering was based on Section 1607C(1) and Section 14(C)(9) of the UGESP. Section 1607C(1) contains the following warning with respect to relying “solely or primarily” on content validity to support “mental processes”:

A selection procedure based on inferences about mental processes cannot be supported solely or primarily on the basis of content validity. Thus, a content strategy is not appropriate for demonstrating the validity of selection procedures which purport to measure traits or constructs such as intelligence, aptitude, personality, common sense, judgment, leadership and spatial ability. [emphasis added]

Additionally, Section 14(C)(9) of the UGESP contains a passage stating that “rank-ordering should be used only if it can be shown that a higher score is likely to result in better job performance.”

The 2nd Circuit ruled that content validity is appropriate for validating tests of “mental processes,” and outlined the five requirements for proving job relatedness and using rank ordering based on content validity:

1. suitable job analysis
2. reasonable competence in test construction
3. test content related to job content
4. test content representative of job content
5. scoring systems selecting applicants that are better job performers

On the issue of cutoff scores, the 2nd Circuit addressed a passage from Section 1607.5H of the UGESP, stating:

Where cutoff scores are used, they should normally be set so as to be reasonable and consistent with normal expectations of acceptable proficiency within the workforce. Where applicants are ranked on the basis of properly validated selection procedures and those applicants scoring below a higher cutoff score than appropriate in light of such expectations have little or no chance of being selected for

---

* The r = +.30 standard for practical significance was actually established in NAACP v. Beecher (1974), a pre-Davis case (see, for example, Landy, Outtz, & Gutman, 2010). However, cases supporting this standard emerged in larger numbers after the Seibels ruling, including Clady v. Los Angeles (1985), Zamien v. City of Cleveland, (1988), and Hamer v. Atlanta (1989) in the 1980s, and more recent rulings such as Williams v. Ford Motor Company (1999) and U.S. v. Delaware (2004).
employment, the higher cutoff score may be appropriate, but the degree of adverse impact should be considered. [emphasis added]

In interpreting this passage, the 2nd Circuit emphasized the importance of professional expertise and logical breakpoints to establish cutoff scores. Accordingly:

As with rank-ordering, a criterion-related study is not necessarily required; the employer might establish a valid cutoff score by using a professional estimate of the requisite ability levels, or, at the very least, by analyzing the test results to locate a logical “break-point” in the distribution of scores. [emphasis added]

Ultimately, the CSC failed on all counts because the in-house test constructors were deemed to lack competence in test construction. However, other defendants profited from the Guardians ruling, most notably the State of Wisconsin.

**Gillespie v. Wisconsin (1985)**

Gillespie v. Wisconsin featured state personnel jobs, but it is as relevant to police selection as any other early lower court ruling. Relying on the 1974 *Standards for Educational and Psychological Tests* and Anastasi’s (1982) *Psychological Testing* text, the 7th Circuit opined that criterion validity is often difficult to accomplish because of technical factors.* Accordingly, criterion-related validation is often impracticable because of the limited numbers of employees available for test development and several measurement errors (APA Standards at 27; Anastasi at 433). Thus, neither the Uniform Guidelines nor the psychological literature expresses a blanket preference for criterion-related validity.

Following the lead of the 2nd Circuit, the 7th Circuit supported the cutoff score used by the state of Wisconsin based on professional expertise, as two qualified experts in test construction testified that the cutoff score was chosen so as to “interview as many minority candidates as possible” while “assuring that the candidates possessed the minimal skills” needed to perform the job in question.

Subsequently, the Guardians and Gillespie rulings were supported in other cases (e.g., *Police Officers v. City of Columbus*, 1990; *Brunet v. City of Columbus*, 1995; and *Williams v. Ford Motors*, 1999). However, as we will witness shortly, agreement among circuit courts on cutoff scores was disrupted by the 3rd Circuit in *Lanning v. Southeastern Pennsylvania Transportation Authority (SEPTA)* (1999).

**RECENT RULINGS ON CUTOFF SCORES AND RELATED ISSUES**

Guardians v. CSC and Gillespie v. Wisconsin established two important rules: (1) if a test is content and/or criterion valid, employers may hire or promote in strict rank order until positions are exhausted; and (2) if strict rank ordering is justified, professional expertise may be used to determine if cutoff scores are “reasonable and consistent with normal expectations of acceptable proficiency within the workforce” in accordance with Section 1607.5(H) of the UGESP. Rule 1 is still applicable across all circuit courts. However, Rule 2 is currently in question based on the 3rd Circuit’s ruling in *Lanning v. SEPTA* (1999), particularly as it applies to initial steps in multiple-hurdle selection systems.

As depicted in Table 3.4, there were four rulings in the Lanning case. Subsequently, the 3rd Circuit’s opinion in Lanning II (*Lanning v. SEPTA*, 1999) was rejected by the 7th Circuit in *Bew v. City of Chicago* (2001) and accepted by the 6th in *Isabel v. City of Memphis* (2005). Additionally,

---

* Technical difficulties associated with criterion validity studies are also discussed by Landy (1986) and Binning and Barrett (1989).
both the *Bew* and *Isabel* rulings addressed important questions relating to the so-called 80% rule for proving adverse impact in Section 1607.4(D) of the UGESP.

**THE LANNING RULINGS**

The *Lanning* case featured a 1.5-mile run as the first step in a multiple hurdle for entry-level transit authority officers. Applicants were required to make the run (in full gear) in 12 minutes or less. Those who passed completed additional physical fitness tests and those who failed were excluded. There was no question of adverse impact in this case; the pass rate on the 12-minute criterion was 60% for males and only 12% for females. The only major issue in this case was the validity of the 12-minute criterion.

In *Lanning I* (*Lanning v. SEPTA*, 1998), the trial judge favored SEPTPA based on job analysis data relating to the importance of aerobic capacity for both officer safety and public safety, and also on criterion validity data on the relationship between aerobic capacity and both successful criminal arrests and commendations for field work. Consistent with the rulings in *Guardians v. CSC* (1980) and *Gillespie v. Wisconsin* (1985), the trial judge supported the 12-minute cutoff.

Then, in *Lanning II* (*Lanning v. SEPTA*, 1999), the 3rd Circuit opined that CRA-91 established new rules for supporting cutoff scores. As written in statutory language in CRA-91, causes of adverse impact must be “job related for the job in question” and “consistent with business necessity.” The 3rd Circuit ruled that proof of test validity is sufficient for job relatedness, and that the defendant succeeded in this respect. However, the 3rd Circuit also ruled that to be consistent with business necessity, a cutoff score must reflect “minimum qualifications necessary for successful performance of the job in question.” Accordingly:

> With respect to a discriminatory cutoff score, the business necessity prong of the Civil Rights Act of 1991, 105 Stat. 1071 (1992) must be read to demand an inquiry into whether the score reflects the minimum qualifications necessary to perform successfully the job in question. [emphasis added]

However, the 3rd Circuit did not strike down the 12-minute criterion, but rather remanded to the district court with instruction to evaluate the 12-minute criterion based on the newly adopted

<table>
<thead>
<tr>
<th>TABLE 3.4 Important Recent Rulings on Cutoff Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lanning Rulings</strong></td>
</tr>
<tr>
<td><em>Lanning v. SEPTA</em> (1999) (Lanning II)</td>
</tr>
<tr>
<td><em>Lanning v. SEPTA</em> (2000) (Lanning III)</td>
</tr>
<tr>
<td><em>Lanning v. SEPTA</em> (2002) (Lanning IV)</td>
</tr>
<tr>
<td><strong>Post-Lanning Rulings</strong></td>
</tr>
<tr>
<td><em>Bew v. City of Chicago</em> (2001)</td>
</tr>
<tr>
<td><em>Isabel v. City of Memphis</em> (2005)</td>
</tr>
</tbody>
</table>
Lanning II standard. Thereafter, in Lanning III (Lanning v. SEPTA, 2000), the trial judge supported SETPA on the Lanning II standard based on evidence that (a) a lower cutoff would endanger officer and public safety; (b) officers who passed the 12-minute criterion made more successful arrests; and, overall, (c) aerobic capacity, the main reason for using the 1.5-mile run, was not overrepresented for the job of transit authority officer. The 3rd Circuit then affirmed district court ruling from Lanning III in Lanning IV (Lanning v. SEPTA, 2002).

The Bew and Isabel Rulings

Unlike Lanning, where adverse impact was clearly established, there were questions in Bew v. Chicago and Isabel v. Memphis relating to the 80% rule for proving adverse impact. As written in Section 1607.4(D) of the UGESP, the 80% rule states:

A selection rate for any race, sex, or ethnic group which is less than four-fifths (4/5) (or eighty percent) of the rate for the group with the highest rate will generally be regarded by the Federal enforcement agencies as evidence of adverse impact, while a greater than four-fifths rate will generally not be regarded by Federal enforcement agencies as evidence of adverse impact.

For example, a selection rate for Caucasians at 70% (e.g., 70 out of 100), requires a selection rate minorities or women at 80% × 70% = 56% (or 56 out of 100 or higher).

However, in a second part of Section 1607.4(D), the UGESP terms the 80% rule a “rule of thumb,” and warns employers not to use it in lieu of other important factors*. Accordingly:

Smaller differences in selection rate may nevertheless constitute adverse impact, where they are significant in both statistical and practical terms or where a user’s actions have discouraged applicants disproportionately on grounds of race, sex, or ethnic group. Greater differences in selection rate may not constitute adverse impact where the differences are based on small numbers and are not statistically significant.

It was the second part of Section 1607.4(D) that was prominently featured in both the Bew and Isabel cases.

In Bew v. City of Chicago, 5,181 probationary police officers were required to pass a written test with a 66% score or higher in order to graduate to full-time appointments. Applicants were given three chances to pass. Only 33 applicants failed three times (less than 1%), and the pass rates for African Americans (98.24%) and Caucasians (99.96%) were within the 80% boundary. However, 32 of the 33 failures were African American applicants, and a test of independent proportions yielded a Z-score of more than five standard deviations, prompting the 7th Circuit to rule that adverse impact was established.† However, as critical for the present discussion, the 7th Circuit supported the 66% cutoff score, ruling that CRA-91 did “not distinguish business necessity and job relatedness as two separate standards.”

In Isabel v. City of Memphis, a written test was the first hurdle for promotion for 120 applicants for police lieutenant. The original cutoff score of 70% would have violated the 80% rule, so the city reduced the cutoff score to 66%, yielding pass rates for African Americans (46 of 63 = 74.6%) and Caucasians (51 of 57 = 89.5%) that were within the boundaries of the 80% rule. The city argued

---

* Comprehensive discussions of the 80% rule compared to statistical and practical significance for small and large sample sizes are provided by Morris and Lobsenz (2000) and Siskin and Trippi (2005).

† In Castaneda v. Partida (1977), plaintiffs charged there was underrepresentation of Mexican Americans in a jury pool where the percentage of Mexican Americans chosen was 39% as compared to 79% in the county as a whole. The Supreme Court, based on testimony from social scientists, ruled that a Z-score difference of two standard deviations or higher constitutes a “gross disparity.” This deviation rule was then applied in Hazelwood v. United States (1977) and was subsequently transported into adverse impact cases as the criterion for proving adverse impact.
that there was no adverse impact, but the plaintiffs prevailed based on (a) a significant difference on mean scores for African Americans (69.17) versus Caucasians (75.59), (b) an effect size of $D = .9$ (which is very large), and (c) a test of independent proportions that resulted in a Z-test difference of 2.35 standard deviations. The trial judge in this case (Isabel v. City of Memphis, 2003) also rejected the defense of the cutoff score based on Lanning II. Accordingly:

The Third Circuit has held that “taken together, Griggs, Albemarle and Dothard teach that in order to show the business necessity of a discriminatory cutoff score an employer must demonstrate that its cutoff measures the minimum qualifications necessary for successful performance of the job in question.” Lanning v. SEPTA, 181 F.3d 478 (3d Cir. 1999). In order to be valid, therefore, the cutoff score of 66 must appropriately measure the minimum qualifications necessary for successful performance of the job of lieutenant in the Memphis Police Department.

The 6th Circuit then affirmed the district court ruling, including the portion relating to Lanning II in its 2005 ruling.

Interestingly, the Lanning II ruling was not that critical to the Isabel ruling. The expert who designed the test for the City of Memphis admitted in open court that the job knowledge component of the test did not represent the full job domain for police lieutenant. Thus, his content validity study was suspect under requirement 4 of Guardians v. CSC (1980) (that test content must be representative of job content). As critical, the expert admitted he “did not condone the usage of a cutoff score” for the promotion process, that he was pressured to do so by the police union, and that the cutoff score adopted was “totally inappropriate,” “logical absurdity,” and “ludicrous.”*

In short, it is relatively clear that employers should not rely solely on the 80% rule to determine if there is adverse impact. However, unless and until the Supreme Court decides the issue, it is an open question as to whether CRA-91 established a new rule for assessing the validity of cutoff scores.

**METHODS FOR REDUCING OR ELIMINATING ADVERSE IMPACT**

Table 3.5 depicts four methods for reducing and/or eliminating adverse impact. Methods 1 and 2 (subgroup norming and banding) use statistical manipulations, Method 3 (alternative tests or combination of tests) was endorsed in two recent district court rulings, and Method 4 (manipulating test content) was endorsed in a controversial ruling in Hayden v. Nassau County (1999).

**SUBGROUP NORMING**

Subgroup norming eliminates adverse impact, but does so by using lower cutoff scores for minorities than Caucasians. Subgroup norming was used routinely in the early 1980s by the U.S. Employment Service (USES) for job referrals based on General Aptitude Test Battery (GATB) test scores.† At the time, subgroup norming was supported by the National Academy of Sciences, but opposed by the Department of Justice (DOJ). However, the debate was abruptly ended by the race-norming provision in Section 106(1) of CRA-91, which makes it unlawful to “adjust the scores of, use different cutoff scores for, or otherwise alter the results of, employment-related tests on the basis of race, color, religion, sex or national origin.”

---

* The city made the mistake of administering the second component of the multiple hurdle, and when both hurdles were complete, one of the African American applicants who finished below the cutoff on hurdle 1 ultimately recorded the second-best score on both hurdles combined.

† Readers interested in a detailed account of the USES procedures are referred to Hartigan and Wigdor (1989) and Sackett and Wilk (1994).
BANDING

Bands are ranges (or bandwidths) in which test scores are treated as being statistically equal.* There are two types of banding: race neutral and race conscious. In race-neutral banding, selections are made randomly within bands. Race-neutral banding is generally legal, even if it increases the percentage of minorities relative to strict rank ordering. This occurred, for example, in Chicago Firefighters Local 2 v. City of Chicago (2001) for promotion of firefighters, where the 7th Circuit ruled that random selection within bands is not a form of race norming. Accordingly:

If banding were adopted in order to make lower black scores seem higher, it would indeed be a form of race norming, and therefore forbidden. But it is not race norming per se. In fact it’s a universal and normally an unquestioned method of simplifying scoring by eliminating meaningless gradations. [emphasis added]

That said, race-neutral banding rarely results in meaningful increases in minority selection as occurred in the Chicago Firefighters case.

Race-conscious banding was originally proposed by Cascio, Outtz, Zedeck, and Goldstein (1991) (see also Cascio, Goldstein, Outtz, & Zedeck, 1995). Proponents of race-conscious banding contend it significantly reduces adverse impact with minimal loss in test utility.† Opposing viewpoints have been expressed by several authors (e.g., Gottfredson, 1994; Sackett & Roth, 1991, Sackett & Wilk, 1994; Schmidt, 1991). The psychometric issues in this debate are complex. For present purposes, it is critical to note that only limited forms of race-conscious banding have been supported in the courts.

Most notably, in Bridgeport Guardians v. City of Bridgeport (1991), the 2nd Circuit supported minority preference within bands, but only as one of nine equally weighted criteria. Then, based on the Bridgeport Guardians ruling, the 9th Circuit supported minority preference within bands as one of four equally weighted criteria in Officers for Justice v. Civil Service Commission (1992).‡

<table>
<thead>
<tr>
<th><strong>TABLE 3.5</strong></th>
<th><strong>Methods for Reducing Adverse Impact</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Subgroup Norming</td>
<td>Use of different norms for minority and nonminority groups; outlawed in the race-norming proscription in CRA-91.</td>
</tr>
<tr>
<td>2. Banding</td>
<td>Arranging scores in bands as an alternative to strict rank ordering. Race-neutral banding has been supported in the courts, but race-conscious banding has been limited at best.</td>
</tr>
<tr>
<td>3. Alternative Tests or Combinations</td>
<td>Valid alternatives with less adverse impact was supported in two district rulings: Bradley v. City of Lynn (2006) and Johnson v. City of Memphis (2006).</td>
</tr>
</tbody>
</table>

* The most common method of computing bandwidth is to compute the Standard Error of measurement (SEM) and multiply it by the square root of 2 to obtain the standard error of difference (SED). Then the SED is multiplied by 1.96 (Z-score associated with 95% of variance in a two-tailed statistical test on the normal curve). For example, assuming an SEM of 3.0, the SED = 4.24 equates to bandwidth of 8.31.

† There are three major methods for race-conscious banding: (1) top-down selection of minorities followed by top-down selection of nonminorities until the band is exhausted; (2) random selection for minorities until either minority candidates are exhausted or the proportion of minorities equals the proportion in the applicant pool, after which nonminority candidates are randomly selected; and (3) the bandwidth slides down after the top-scoring minority and nonminority candidates are chosen.

‡ The eight factors other than minority preference in the Bridgeport case were departmental needs, job dependability, assignment history and job experience in the department, departmental or special skills training, participation in law enforcement/management seminars and courses, formal university education in law enforcement/management, ability to interact with persons of diverse backgrounds, and overall job performance. The three factors other than minority preference in the Officers case were professional conduct, education and training, and experience.
Critically, neither court supported minority preference as the sole basis for selection, and in the Officers case, race-conscious banding was limited to the last 15 of 115 promotions.

In Boston Police Superior Officers v. City of Boston (1998), another notable case, the 1st Circuit upheld promotion of an African American applicant who scored one point less than three higher-scoring Caucasians because it was consistent with the demands of a consent decree. The Caucasian officers challenged the consent decree under the 14th Amendment and lost.

In short, there are some instances in which limited forms of race-conscious banding have been supported by the courts, particularly when there were consent decrees to resolve past instances of discrimination. However, race as the sole basis for selection within bands has never been supported in any court (see Henle, 2004), and it is highly doubtful that race can serve as the sole basis for large-scale selection in view of the race-norming proscription in CR-91, even under consent decrees.

**ALTERNATIVE SELECTION PROCEDURES**

Unlike subgroup norming and race-conscious banding, valid alternative selection procedures with less or no adverse impact have a strong legal basis. As noted earlier in this chapter, the Supreme Court created this standard as Phase 3 of the adverse impact scenario in Albemarle v. Moody (1975). It was then adopted as regulatory law in the UGESP, and written into statutory law in CRA-91. As depicted in Table 3.4, plaintiffs have succeeded with Phase 3 proof in two recent district court rulings.

In Bradley v. City of Lynn (2006), a written test was used for entry-level firefighters. The trial judge ruled there was adverse impact and insufficient evidence of job-relatedness. The judge also cited two valid alternatives: (1) a combination of cognitive tests and physical abilities, and (2) a combination of cognitive tests, personality tests, and biodata. The judge ruled that “while none of these approaches alone provides the silver bullet, these other non-cognitive tests operate to reduce the disparate impact of the written cognitive examination.”

Arguably, the City of Lynn was a likely loser in Phase 2 absent the Phase 3 argument. Not so in Johnson v. City of Memphis (2006), where the trial judge ruled that a promotion exam (for police sergeant) was valid, but the plaintiffs prevailed in Phase 3. The ruling was based on the prior development of a valid promotion test in 1996 that resulted in less adverse impact relative to the current exam. The judge ruled, “It is of considerable significance that the City had achieved a successful promotional program in 1996 and yet failed to build upon that success.”

It is unclear how much weight should be placed on district court rulings absent circuit court appeals. However, it is clear that there is sufficient case law, regulatory law, and statutory law to instruct employers to consider alternative selection procedures with less adverse impact as they commission and/or develop selection tests.

**MANIPULATING TEST CONTENT**

The events preceding Hayden v. Nassau County (1996) were that in 1977, the DOJ sued Nassau County for adverse impact in an entry-level police exam, and the end result was a consent decree in 1982 to construct an exam that produces no adverse impact, or is valid “in accordance with Title VII and the Uniform Guidelines.” However, exams developed in 1983 and 1987 again resulted in adverse impact (and two new consent decrees). Then, in 1990, the DOJ and Nassau appointed a Technical Design Advisory Committee (TADC) to develop a new exam.

The TADC developed and administered a 25-component test to 25,000 candidates. There was “severe” adverse impact after all 25 components were scored. The TADC attempted to eliminate adverse impact completely, but considered the end result invalid. In the exam ultimately adopted, 16 of the 25 components were eliminated, resulting in less adverse impact. However, the subsequent
nine-component test was challenged by 68 unsuccessful candidates alleging that they would have
been selected if all 25 components were used.

The 2nd Circuit ruled “the intent to remedy the disparate impact of the prior exams is *not* equiva-
 lent to an intent to discriminate against non-minority applicants.” The court acknowledged that the
decision to “redesign the exam” was race conscious. However, the court reasoned that the exam was
“scored in a wholly race-neutral fashion” and ruled that the plaintiffs failed to state a claim under
the Equal Protection Clause of the 14th Amendment, Title VII, or the race-norming proscription in
CRA-91.

The Hayden ruling is difficult to interpret, particularly in view of the Supreme Court’s ruling in
*Ricci v. DeStefano* (2009). At the time, Nassau County had the benefit of a court-sanctioned consent
degree with the DOJ to resolve multiple lawsuits over a 20-year period. Therefore, at least in the
author’s opinion, it is not advisable to use the Hayden procedure absent court approval.

**THE SUPREME COURT’S RULING IN THE RICCI CASE**

In *Ricci v. DeStefano* (2009), the New Haven Civil Service Board (CSB) refused to certify pro-
motion exams for firefighter captain and lieutenant, thereby effectively discarding the exam. The
CSB argued it had a good faith belief that it would lose an adverse impact challenge to minori-
ties. Subsequently, 17 Caucasians and 1 Hispanic sued and ultimately won at the Supreme Court
level on grounds that the CSB decision to not certify constituted illegal disparate treatment under
Title VII because there was no strong basis in evidence for believing the minority firefighters
would prevail.*

There were 41 applicants and 7 vacancies for captain and 77 applicants and 8 vacancies for
lieutenant. The CSB used a “rule of three” in which any of the 3 highest scoring applicants could
be promoted for a given vacancy. This left 9 applicants eligible for captain and 10 applicants
eligible for lieutenant. As depicted in Table 3.6, there were racial differences on passing rates
(a score of 70 or higher). More important, no African Americans and two Hispanics were eligible
for promotion to captain, and no African Americans or Hispanics were eligible for promotion to
lieutenant.

The exams were developed by Industrial-Organizational Solutions (IOS) using a content
validity strategy. The CSB sought input from several sources, but relied primarily on one of them
(Dr. Hornick) for their good faith belief that they would lose an adverse impact challenge. Dr. Hornick,

<table>
<thead>
<tr>
<th>TABLE 3.6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Projected Promotions in the Ricci Case</td>
</tr>
<tr>
<td>-------------</td>
</tr>
<tr>
<td><strong>Captain Exam (7 Vacancies)</strong></td>
</tr>
<tr>
<td><strong>Caucasians</strong></td>
</tr>
<tr>
<td>Applicants</td>
</tr>
<tr>
<td>Passing Score</td>
</tr>
<tr>
<td>Top 9 Scores</td>
</tr>
<tr>
<td><strong>Lieutenant Exam (8 Vacancies)</strong></td>
</tr>
<tr>
<td><strong>Caucasians</strong></td>
</tr>
<tr>
<td>Applicants</td>
</tr>
<tr>
<td>Passing Score</td>
</tr>
<tr>
<td>Top 10 Scores</td>
</tr>
</tbody>
</table>

* The facts in this case are very complex and are only summarized here. Readers interested in a more detailed account of
the Ricci case are referred to Gutman and Dunleavy (2009).
a competitor to IOS, testified by telephone. He never reviewed the actual exams. Furthermore, he
opined that the IOS exams were valid. However, he also opined that he generally found less adverse
impact as compared with the IOS tests using an assessment center approach. The CSB ultimately
discarded the exams without requesting a validity report by IOS. They also hired Dr. Hornick to
create and validate new exams.

In court, the CSB argued that they “cannot be held liable under Title VII’s disparate-treatment
provision for attempting to comply with Title VII’s disparate-impact bar.” The plaintiffs, on the
other hand, argued that a “good-faith belief was not a valid defense to allegations of disparate
treatment and unconstitutional discrimination.” Critically, the Supreme Court never evaluated written
tests versus assessment centers in reaching its ruling.

At the district court level, the trial judge (Janet Bond Arterton) acknowledged that the CSB’s deci-
sion was race conscious. Nevertheless, based on *Hayden v. Nassau County* (1996), Judge Arterton
ruled that the noncertification decision was “race neutral.” She ruled further that the “intent to
remedy disparate impact of the prior exams is not equivalent to an intent to discriminate against
non-minority applicants.” A three-judge panel of the 2nd Circuit then affirmed Judge Arterton’s rul-
ing in a short per curiam ruling. Subsequently, a 13-judge panel of the 2nd Circuit refused to further
review the case in a close 7–6 ruling. The Supreme Court then agreed to review the case based on a
written opinion by the six dissenters.

The Supreme Court’s ruling was 5–4, with Justice Kennedy speaking for Justices Alito, Roberts,
Scalia, and Thomas. Kennedy ruled there was a race-conscious motive for discarding the test, and
strongly suggested that the CSB would have certified the tests if the results were more favorable for
minorities. Accordingly:

> Whatever the City’s ultimate aim—however well intentioned or benevolent it might have seemed—
> the City made its employment decision *because of race*. The City rejected the test results solely
> because the higher scoring candidates were Caucasian. *The question is not whether that conduct was
discriminatory but whether the City had a lawful justification for its race-based action.* [emphasis
added]

Nevertheless, this was not the sole basis for the majority ruling, as Kennedy ruled further that it
is necessary to balance the “tension” between disparate treatment and adverse impact. He rejected
a *certainty* criterion for losing on adverse impact. Accordingly:

> Forbidding employers to act unless they know, with certainty, that a practice violates the disparate-
impact provision would bring compliance efforts to a near standstill. Even in the limited situations
when this restricted standard could be met, employers likely would hesitate before taking voluntary
action for fear of later being proven wrong in the course of litigation and then held to account for dis-
parate treatment. [emphasis added]

Kennedy also rejected a good-faith belief on grounds that it was too *minimal*, ruling:

> Allowing employers to violate the disparate-treatment prohibition based on a mere good-faith fear of
dis disparate-impact liability would encourage race-based action at the slightest hint of disparate impact.
A *minimal* standard could cause employers to discard the results of lawful and beneficial promotional
examinations even where there is little if any evidence of disparate-impact discrimination. That would
amount to a de facto quota system, in which a “focus on statistics … could put undue pressure on
employers to adopt inappropriate prophylactic measures. [emphasis added]

Borrowing from Title VII and 14th Amendment case law from reverse discrimination rulings
(e.g., *Wygant v. Jackson*, 1986; *City of Richmond v. Croson*, 1989) and invoking the race-norming
provision in CRA-91, Kennedy ruled it would be legal to discard an exam if there is a strong basis
in evidence for believing an employer would lose on adverse impact. Accordingly:
If an employer cannot rescore a test based on the candidates’ race, §2000e-2(l), then it follows a fortiori that it may not take the greater step of discarding the test altogether to achieve a more desirable racial distribution of promotion-eligible candidates—absent a strong basis in evidence that the test was deficient and that discarding the results is necessary to avoid violating the disparate-impact provision.... For the foregoing reasons, we adopt the strong-basis-in-evidence standard as a matter of statutory construction to resolve any conflict between the disparate-treatment and disparate-impact provisions of Title VII. [emphasis added]

Ultimately, reasoning that the CSB did not pass the strong-basis-in-evidence test, Kennedy overturned both lower courts and granted summary judgment for the plaintiffs, thus ending the case.

In the dissenting opinion, Justice Ginsburg, speaking for Justices Breyer, Souter, and Stevens, criticized the strong-basis-in-evidence standard on grounds that it was inappropriately incorporated into Title VII adverse impact law based on 14th Amendment reverse discrimination affirmative action cases (Wygant v. Jackson, 1986; Richmond v. Croson, 1988). Accordingly:

The cases from which the Court draws its strong-basis-in-evidence standard are particularly inapt; they concern the constitutionality of absolute racial preferences. See Wygant v. Jackson Bd. of Ed., 476 U. S. 267, 277 (1986) (plurality opinion) (invalidating a school district’s plan to lay off nonminority teachers while retaining minority teachers with less seniority); Croson, 488 U.S., at 499–500 22 (rejecting a set-aside program for minority contractors that operated as “an unyielding racial quota”). An employer’s effort to avoid Title VII liability by repudiating a suspect selection method scarcely resembles those cases. Race was not merely a relevant consideration in Wygant and Croson; it was the decisive factor. Observance of Title VII’s disparate-impact provision, in contrast, calls for no racial preference, absolute or otherwise. [emphasis added]

Ginsburg endorsed a lighter reasonableness standard, under which the CSB’s “good faith” belief would be acceptable for discarding the test under disparate treatment rules.

It is critical to remember that Kennedy’s majority ruling was based entirely on Title VII. No precedents were established for the Equal Protection Clause of the 14th Amendment. Indeed, Kennedy suggested that a favorable Title VII ruling on strong-basis-in-evidence standard could at a later time fail under constitutional law. Accordingly:

Our statutory holding does not address the constitutionality of the measures taken here in purported compliance with Title VII. We also do not hold that meeting the strong-basis-in-evidence standard would satisfy the Equal Protection Clause in a future case. As we explain below, because respondents have not met their burden under Title VII, we need not decide whether a legitimate fear of disparate impact is ever sufficient to justify discriminatory treatment under the Constitution. [emphasis added]

In short, much uncertainty remains related to the strong-basis-in-evidence standard advocated by the Supreme Court.

It should be noted that the Office of Contract Compliance Programs (OFCCP) of the Department of Labor, which monitors compliance with Executive Order 11246 on affirmative action, has released guidance that states, among other things, that Ricci will not alter the affirmative action obligations of contractors, and that the OFCCP will continue to use the UGESP to assess adverse impact cases, particularly as it relates to job analysis and test validation procedures. The OFCCP added, however, that it would examine refusals to use tests in accordance with the strong-basis-in-evidence standard.*

An additional point to note is that as this chapter was going to press, the Chicago Police Department was considering scrapping its entry tests. According to an article written on January 6, 2010, in the Chicago Sun Times, this action would “bolster minority hiring, save millions on test

* The full OFCCP report is available at http://www.dol.gov/ofccp/regs/compliance/faqs/Ricci_FAQ.htm
preparation and avert costly legal battles that have dogged the exam process for decades.”* While it is unclear how much weight should be given to this newspaper report, the notion that all legal problems are averted by simply eliminating testing procedures is, for reasons to be discussed shortly, not necessarily true.

CONCLUSIONS AND RECOMMENDATIONS

The case law surveyed in this chapter reveals that municipalities face a variety of Hobson’s choices with respect to development and defense of cognitive tests. By the author’s count, there are four ways for a municipality to be sued relating to adverse impact:

1. Adverse impact is proven by minority plaintiffs, and the municipality fails to prove that a challenged test or other selection criterion is job related.
2. The municipality proves job relatedness, but minority plaintiffs prove that a rank cutoff score for selection decisions is unjustified by the evidence offered.
3. The municipality can successfully defend its cutoff score, but minority plaintiffs prove there are other equally valid methods that produce less or no adverse impact to accomplish the same goals.
4. The municipality uses what it believes is the most valid methods with the least amount of adverse impact, but nonminority plaintiffs challenge these methods on grounds of “reverse discrimination” based on preference for minority candidates.

Additionally, there is potential liability if a municipality scraps the testing process altogether or reduces adverse impact by other questionable means, and the selection process used results in less-qualified police officers. For example, if, because of negligence or intent, an unqualified officer injures or kills a citizen, the municipality could be liable for remedies associated with wrongful death based on its failure to select only the most qualified applicants for police work.

In short, there are good reasons for relying on objective tests and related procedures for hiring and promoting police officers. They represent the best way to make decisions when there are large numbers of applicants. Additionally, many municipalities require such testing. That said, municipalities must be prepared to prove three things: (1) that the tests or other selection criteria used are job related, (2) that there is sufficient evidence for use of rank ordering and/or a specific cutoff score, and (3) that they have examined potentially valid alternative methods that produce less adverse impact before adopting their test strategy.

As discussed by Gutman and Dunleavy (2009), the correct reaction to Ricci is to take a proactive approach when developing selection tests. The Ricci ruling had little to do with whether, for example, written tests are better than assessment centers. That question would have been relevant had the New Haven CSB stood by its tests and offered a content validity defense (which likely would have succeeded), and minority plaintiffs launched a Phase 3 proof on valid alternatives with less or no adverse impact. But that did not happen. However, that is precisely the scenario that municipalities must prepare for regardless of which selection procedures are used.

What Gutman and Dunleavy (2009) recommend is to establish an advisory committee that includes experts in job analysis and test validation prior to issuing an RFP. The advisory committee should then play a key role in awarding a contract and work with the contractor(s) to ensure that tests and other selection procedures are administered and scored in accordance with the UGESPS, and authorities such as the Standards and the SIOP Principles. That would represent an ideal way to establish a strong basis in evidence for discarding a test. Of course, if there is a strong basis for

discarding a test, there is no reason to administer and score the test to begin with, and wait to see the results to determine if it is acceptable.

REFERENCES


CASES CITED


Bridgeport Guardians v. Bridgeport CSC (CA2 1973) 482 F.3d. 1333.
Bridgeport Guardians, Inc. v. City of Bridgeport (CA2 1991) 933 F.2d 1140.
Brunet v. City of Columbus (CA6 1995) 58 F.2d 251.
Chicago Firefighters Local Union 2 v. City of Chicago (CA 7 2001) 49 F.3d 649.
Clady v. Los Angeles (CA 9 1985) 770 F.2d 1421.
Davis v. Dallas (CA5 1985) 777 F.2d 205.
Hamer v. City of Atlanta (CA11 1989) 872 F.2d 1521.
Hayden v. Nassau County (CA2 1999) 180 F.3d 42.
Isabel v. City of Memphis (CA6 2005) 404 F.3d 404.
Kirkland v. New York State CSC (CA2 1975) 520 F.2d 420.
Police Officers for Equal Rights v. City of Columbus (CA6 1990) 916 F.2d 1092.
Spurlock v. United Airlines, Inc. (CA10 1972) 475 F.2d 216.
Vulcan Society v. CSC of New York City (CA2 1973) 490 F.2d 387.
Washington (Mayor, DC) v. Davis (1976) 426 US 229.
Zamien v. City of Cleveland (CA6 1988) 906 F.2d 209.