

THE RICCI V. DESTEFANO CONUNDRUM: DID THE SUPREME COURT GET IT RIGHT?

Bonnie Roach*

INTRODUCTION

A recent court decision has threatened to change the progress of civil rights and turn the clock back on addressing tests with disparate impact. The City of New Haven faced an unusual problem which involved the test they had just given their firefighters.¹ The tests were going to be used to promote qualified firefighters to the positions of captain or lieutenant. The process that was used to create the tests seemed to be designed to ensure that the tests would be job-related and pose no risk with regards to discrimination. Much time, effort and expense went into the creation of these exams. A consulting firm was engaged to create the exams that had to be in conformity with New Haven statutes and the collective bargaining agreement with the Firefighters union which required that 60 percent of the exam was to be written; 40 percent was oral.² The consultants, ISO, went to great pains to make sure that the items in both exams were the result of a thorough job analysis.³ To ensure that the job data was representative of all firefighters, the ISO team consciously chose to oversample minority job incumbents.⁴ In fact, the whole process that was used to create the tests seemed to follow ideal human resource principles. So, the results of the tests should not reflect any discrimination, right?

Wrong. When the results were tabulated, it was clear that the test failed the four-fifths rule (or the 80 percent rule)⁵ that the courts use to determine a prima facie case of disparate impact. The results were alarming. For the position of Captain, the pass rate was as follows: White: 64%; Hispanic and African American: 37.5%. For the position of Lieutenant: White: 58.1%; African-American: 31.6%; and Hispanic: 20%. Clearly, the test fails the 4/5's rule.⁶ The City was faced with a very strong possibility of facing disparate impact claims if the results of the test were certified. If that was the case, then the City should employ the simple solution of just not certifying the test results. This approach carried its own consequences with it and so the City faced a conundrum - refuse to certify the test results on the basis of the disparate impact evidence and risk the white firefighters bringing a Title VII claim against the City for disparate treatment; or to go ahead and use the test results which have disparate impact and risk subsequent claims from minority firefighters on the basis of disparate impact. When the test results were published, there was an outcry from the public and political factions that subsequently caused "a rancorous public debate" on whether the test results should be thrown out or kept.⁷

* Ohio University

¹ Ricci v. DeStefano, 129 S.Ct. 2658 (2009).

² *Id.* p. 2664.

³ *Id.* p. 2665.

⁴ *Id.* p. 2666.

⁵ Uniform Guidelines on Employee Selection Procedures (UGESP), 41 C.F.R. §1607.4(D), 2010.

⁶ *Id.*

⁷ *Ricci*, at 2667.

The City's Civil Service Board (CSB) had the task of determining whether to certify the results and to best decide whether to do this; the Board held hearings over a period of several days where all interested parties from the firefighters and the test developers were able to testify about the testing process. After hearing much testimony on the issue, the CSB decided not to certify the results. Sure enough, the white firefighters and one of the Hispanic candidates who passed the test brought a Title VII lawsuit against the City of New Haven for disparate treatment under Title VII and for a violation of the Equal Protection Clause of the Constitution.⁸

The District Court granted the City's motion for summary judgment saying that just because the City decided not to use the test results was not evidence of an intentional discrimination since all firefighters were equally affected by this decision and there was no racially based animus on the part of the City. The Appellate Court agreed.⁹

The Supreme Court granted certiorari with the intent of guiding employers on managing the kind of dilemma that was experienced by the City of New Haven.¹⁰ In doing so, the Court developed a new standard known as the "strong basis in evidence" test.¹¹ Unfortunately, the standard was vague in its genesis and its application. The Majority gave no indication as to what factors were necessary to consider, the weighting of those factors, and whether any factor was dispositive in tipping the scales one way or another. The essence of the test is that the only way an employer can disregard the test results is if it could be shown that "the employer can demonstrate a strong basis in evidence that, had it not taken the action, it would have been liable under the disparate-impact statute".¹² In the Court's eyes, the City failed to satisfy the standard. Testimony by the City's consultant and the testimony by other testing experts all convinced the Court that the test was valid and thus the City improperly decided not to certify the test results.

The court's decision is alarming for several reasons: 1) The Court equates corrective actions for discriminatory tests with race-based discrimination; 2) The strong-basis-in evidence test requires an employer to basically mount a disparate impact case against itself; 3) The treatment of the disparate impact evidence by the Court appears to be a return to the *Ward's Cove*¹³ standard which only required a showing that the test was "business-related" and served a legitimate business purpose to defend a claim of disparate impact; 4) The Court appears to ignore the required proof of a test's "job-relatedness" by ignoring the Uniform Guidelines on Employee Selection Procedures of 1978;¹⁴ and 5) By requiring an employer to mount a detailed and expensive disparate impact claim against itself, any incentive in voluntary compliance – the goal of Title VII – is minimized. By applying the "strong basis in evidence" test, the Court seems to have entered a slippery slope in taking a giant step backwards in maintaining equal protection for all protected classes.

The Majority made a conscious decision to ignore statutory, administrative and case law with respect to what constitutes disparate impact and whether voluntary attempts to correct a

⁸ *Id.* at 2667-2669.

⁹ *Id.* at 2671-2674. This case gained much notoriety during the Justice Sonia Sotomayor's Supreme Court confirmation hearings. Justice Sotomayor wrote the Appellate opinion which agreed with the City's action.

¹⁰ *Id.*

¹¹ *Id.* at 2675.

¹² *Id.* at 2664.

¹³ *Wards Cove v. Atonio*, 490 U.S. 642 (1989).

¹⁴ The Uniform Guidelines on Employee Selection Procedures, 20 CFR § 1607, 2010.

discriminatory test must be regarded as disparate treatment. The Majority seemed to refuse to use the Uniform Guidelines on Employee Selection Procedures as a benchmark for determining whether the test was invalid. By refusing to use the Uniform Guidelines, the Court seems to have fundamentally changed the standards for proving disparate impact of a test.

THE EVOLUTION OF DISPARATE IMPACT THEORY.

Title VII prohibits discrimination on the basis of race, color, religion, gender and national origin.¹⁵ Two types of discrimination are prohibited under the law: disparate treatment, which is intentionally treating someone differently because of their membership in a protected class;¹⁶ and disparate impact,¹⁷ where use of a policy or practice results in discrimination. Once a plaintiff makes a prima facie case of disparate impact, the employer must assume the burden of proof to demonstrate that the policy or practice is job-related and consistent with business necessity.¹⁸

The concept of disparate impact – where a facially neutral policy or practice creates an adverse effect on a protected class – first was articulated by the Supreme Court in *Griggs v. Duke Power*. The Court determined that the plain language of Title VII was to achieve equality in employment and to remove any barrier to that goal.¹⁹ Furthermore, the Equal Employment Opportunity Commission (EEOC)–the agency created to enforce Title VII – set up guidelines which declared that tests must be job-related.²⁰ Since the Guidelines were considered “the will of Congress,”²¹ they should be accorded “great deference” by the courts.²²

Like other agencies, the EEOC has the authority to create administrative rules to help clarify how employers must comply with Title VII. The Uniform Guidelines on Employee Selection Procedures²³ carefully explain how an employer can establish the job-relatedness of their selection procedures as well as the methods for determining if any selection method has disparate impact. Statistics play a fundamental part in establishing the test’s validity (or job-relatedness) and the Guidelines specify both the methods and the types of data needed to demonstrate job-relatedness.²⁴ They also specify the type of test required for determining the

¹⁵ 42 U.S.C. §§2000e-2(k)(1)(A)-(C), 2006.

¹⁶ 42 U.S.C. § 2000e-2(a)(1).

¹⁷ 42 U.S.C. § 2000e-2(k)(1)(A)(i).

¹⁸ 42 U.S.C. §§ 2000e-2(k)(1)(A)(ii) and (C).

¹⁹ *Griggs v. Duke Power*, 401 U.S. 424, 429-430 (1971).

²⁰ *Id.* at 433-434.

²¹ *Id.* at 433.

²² *Id.* at 434.

²³ 24 C.F.R §§1607.1-16.07.18.

²⁴ *Id.*, at § 1607.5(B) (“Evidence of the validity of a test or other selection procedure by a criterion-related validity study should consist of empirical data demonstrating that the selection procedure is predictive of or significantly correlated with important elements of job performance. Evidence of the validity of a test or other selection procedure by a content validity study should consist of data showing that the content of the selection procedure is representative of important aspects of performance on the job for which the candidates are to be evaluated. Evidence of the validity of a test or other selection procedure through a construct validity study should consist of data showing that the procedure measures the degree to which candidates have identifiable characteristics which have been determined to be important in successful performance in the job for which the candidates are to be evaluated.”) (internal citations omitted.)

test's disparate impact through the use of the Four-Fifths Rule.²⁵ Finally, the Guidelines talk about information to be examined if a test proves to have disparate impact.²⁶ It would appear that there was little confusion over what constituted disparate impact and what standards would be necessary to prove a test is job-related. The Griggs decision, coupled with Uniform Guidelines, required the employer to demonstrate statistical proof that the test was job related and did not have disparate impact.

The case of *Albermarle v. Moody*²⁷ reiterated the concepts in *Griggs* and expanded the requirement for the employer to defend practices resulting in disparate impact. The *Albermarle* Court agreed that the employer has a responsibility to demonstrate a "manifest relationship" to the job, but demonstration of these relationships was not sufficient in and of itself to remove liability from the employer. The burden would then be with the complaining party: "[I]t remains open to the complaining party to show that other tests or selection devices, without a similarly undesirable racial effect, would also serve the employer's legitimate interest in 'efficient and trustworthy workmanship.'" ²⁸

The burden of proof was significantly reduced for the employer after the decision in *Wards Cove Packing Co. v. Atonio*.²⁹ Rather than having to show an employment practice is essential to the job or serves some significantly job-related purpose, the Court in *Ward's Cove* deemed that the employer merely needs to articulate that a discriminatory test served a legitimate business interest. The Court stated, ". . . , it is generally well established that at the justification stage of such a disparate-impact case, the dispositive issue is whether a challenged practice serves, in a significant way, the legitimate employment goals of the employer."³⁰ Even though the Court acknowledged that an employer could not successfully defend an employment practice by asserting "insubstantial justification" for using a practice with disparate impact, employers have no obligation to demonstrate that the challenged practice is essential to the business. A requirement that the practice be essential poses an impossibly high standard for the employer and would result in a "host of other evils".³¹ Needless to say, this made it much easier on employers to justify using tests which had disparate impact.

²⁵ *Id.*, at § 1607.4(D). ("' 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. 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.)

²⁶ *Id.*, at § 1607.14. (The Guidelines stress the need for a job analysis and proper statistical measures to determine a test's validity. Perhaps more importantly to the *Ricci* court is in section 1607.14(B)(8) which specifies what must be done when unfairness is found in the test: "If unfairness is demonstrated through a showing that members of a particular group perform better or poorer on the job than their scores on the selection procedure would indicate through comparison with how members of other groups perform, the user may either revise or replace the selection instrument in accordance with these guidelines, or may continue to use the selection instrument operationally with appropriate revisions in its use to assure compatibility between the probability of successful job performance and the probability of being selected.")

²⁷ *Albermarle Paper Co. v. Moody*, 422 U.S. 405 (1975).

²⁸ *Id.* at 425.

²⁹ *Ward's Cove*, 490 U.S. 642 (1989).

³⁰ *Id.* at 659.

³¹ *Id.*

Congress was so upset by the decision in *Wards Cove*, that it passed the Civil Rights Act of 1991 for the specific purpose of overturning *Ward's Cove*. The statute specifically requires that the employer demonstrate proof that any test showing disparate impact must be job-related and consistent with business necessity.³² The passage of The Civil Rights Act of 1991³³ sent a clear message that Congress was committed to the requirement that employers must adhere to a strict showing that a test with disparate impact could only be defended by a direct showing that the test in question is job-related and is consistent with business necessity. The Uniform Guidelines provided very specific methodology for demonstrating the reliability and validity of the test. Any attempt by the judiciary to allow a lesser standard would be met with statutory checks such as the Civil Rights Act of 1991.

The *Ricci* case has returned the Court back to the *Ward's Cove* period. The Supreme Court has now articulated a new test, one which seems to ignore statutory requirements, administrative rules and the Court's own judicial history by creating a new standard known as the "strong basis in evidence" standard.

RICCI'S "STRONG BASIS IN EVIDENCE TEST".

The *Ricci* Court took a sharp departure from previous jurisprudence with respect to proving and defending disparate impact cases. The City of New Haven had to decide whether to discard test results that clearly had disparate impact and face charges of disparate treatment or accept the test results and face charges of disparate impact.

The Court emphasized that employers can engage in prohibited disparate treatment only if there is a "strong basis in evidence" engaging in intentional discrimination would be the only way to avoid disparate impact liability.³⁴ In other words, does the employer have a well-founded fear that a claim for disparate impact would be successful? If so, that would be the only justification for the employer to discard or take some other form of action against a test that has disparate impact, especially if that corrective action results in disparate treatment of a protected class.³⁵ The Court claimed that the strong basis in evidence test was derived from Equal Protection cases³⁶ where it was determined that the strong basis in evidence test is derived statutorily and would be appropriately applied to settle the tension between Title VII disparate impact and disparate treatment.³⁷

STATISTICAL PROOF THAT THE TEST HAD DISPARATE IMPACT – THE 4/5'S RULE.

There was no doubt that the test had disparate impact. The statistical results clearly demonstrated that the test discriminated against African-Americans and Hispanics as determined

³² *Id.* at 1071, 1074.

³³ Civil Rights Act of 1991 (Pub. L. 102-166), as enacted on November 21, 1991.

³⁴ *Ricci*, at 2661.

³⁵ *Id.* at 2677. ("under Title VII, before an employer can engage in intentional discrimination for the asserted purpose of avoiding or remedying an unintentional disparate impact, the employer must have a strong basis in evidence to believe it will be subject to disparate-impact liability if it fails to take the race-conscious, discriminatory action.")

³⁶ *Id.* See *Watson v. Fort Worth Bank & Trust*, 487 U.S. 977, 993 (1998); *Richmond v. J.A. Croson Co.*, 488 U.S. 469, 500 (1989); See Also *Wygant v. Jackson Bd. of Ed.*, 476 U.S. 267, 277 (1986).

³⁷ *Id.* at 2662.

by the four-fifths rule³⁸ which is still the accepted statistical rule of thumb to determine a prima facie case of disparate impact.³⁹ However, the *Ricci* Court determined that failing the four-fifths was not sufficient in and of itself to satisfy the strong basis in evidence test. To do that, the Court determined that the test must also be proven to be not job-related or consistent with business necessity and that there was not an existing, equally valid test out there that the City refused to adopt.⁴⁰ The Court determined that the City had not provided evidence of either and that the tests appeared to be job-related and there was no evidence that there was any equally valid test that would not have disparate impact.⁴¹

Although, there is no dispute that the test failed the 4/5's rule, what was very surprising was that the Court equated the City's remedial action of discarding the test results as akin to acting in a race-based way against the non-minority candidates.⁴² The District Court, following the Appellate precedent of *Hayden v. County of Nassau*,⁴³ determined that the City of New Haven was not acting in a race-based way when the test results were discarded.⁴⁴ On appeal to the U.S. Supreme Court, the Dissent, as articulated by Justice Ginsberg, observed that the City was no doubt "conscious of race during the decision making process, . . . , but this did not mean they had engaged in racially disparate treatment".⁴⁵ Yet, the Supreme Court interpreted the City's decision not to certify the test results as a "race-based" decision.⁴⁶ Since the Court viewed the City's decision as a race-based decision, it equated the City's decision as one which must be equated with disparate treatment – intentionally treating a protected class differently because of their membership in that protected class.⁴⁷ Yet, according to Title VII, the City should be immune to disparate treatment liability. Justice Ginsberg for the Dissent argued that because the Court refused to remand the case for further proceedings, The City would not be able to invoke 42 U.S.C. § 2000e-12(b) of Title VII which would shield the City from liability for taking remedial action for a test having disparate impact.⁴⁸ The only remaining defense for the City was to demonstrate to the Court that the test was not job-related nor tied to business necessity.

³⁸ *Uniform Guidelines on Employee Selection Procedures*, 29 C.F.R. §1607.4(D), 2010.

³⁹ *Id.*

⁴⁰ *Ricci* at 2662.

⁴¹ *Id.* at 2663.

⁴² *Id.* at 2697.

⁴³ *Hayden v. County of Nassau*, 180 F.3d 42 (C.A.2 1999).

⁴⁴ *Ricci v. DeStefano*, 554 F.Supp.2d 142, 157 (Conn. 2006). (Stating that "the intent to remedy the disparate impact" of a promotional exam "is not equivalent to an intent to discriminate against non-minority applicants (quoting *Hayden v. County of Nassau*, 180 F.3d 42, 51 (C.A.2 1999)).

⁴⁵ *Ricci*, 129 S.Ct. at 2696.

⁴⁶ *Id.* at 2661-2662. ("All the evidence demonstrates that the City rejected the test results because the higher scoring candidates were white. Without some other justification, this express, race-based decision-making is prohibited.")

⁴⁷ Title VII of the Civil Rights Act of 1964, 42 U.S.C. 2000(e) § 703(a)(1).

⁴⁸ *Ricci* at 2703. See 42 U.S.C. § 2000e-12(b), ("In any action or proceeding based on any alleged unlawful employment practice, no person shall be subject to any liability or punishment for or on account of (1) the commission by such person of an unlawful employment practice if he pleads and proves that the act or omission complained of was in good faith, in conformity with, and in reliance on any written interpretation or opinion of the [EEOC] Such a defense, if established, shall be a bar to the action or proceeding, notwithstanding that (A) after

DID THE TEST HAVE DISPARATE IMPACT?

The Court strongly believed that the evidence and testimony presented regarding the test's development overwhelmingly satisfied the first element of the strong basis in evidence test – that the test was job-related. Specifically, the Court found that the consultant's systematic procedures for test development provided a test that was job related to the positions of Lieutenant and Captain. In fact, the Court stated that the City was turning a “blind eye”⁴⁹ to the test's validity.⁵⁰ What is so troubling about the Court's conclusion is the lack of deference that was given to the Uniform Guidelines which carefully spells out the steps required to develop a test with criterion validity (is statistically linked to job performance) and the information that needs to be examined if it is discovered that a test has disparate impact.

DEVELOPMENT OF THE TEST.

The Uniform Guidelines mandate that a selection test should be built upon the results of a job analysis. That task fell to Industrial/Organizational Solutions, Inc. (IOS) to develop and administer the test. They advertised themselves as specializing in producing promotional tests for police and firefighters. The IOS team did conduct a job analysis and many of their activities appeared to conform to good human resource practices. To ensure that the tests were job-related and identified the relevant Knowledge, Skills, and Abilities (KSAs), the consultants interviewed incumbent lieutenants and captains about their jobs and even participated in “ride-alongs” to construct the job analysis questions that most of the incumbent battalion chiefs, captains, and lieutenants in the Department should be able to successfully answer.⁵¹ Since there had been past concerns of discriminatory testing in the past, IOS took great pains to make sure that minority firefighters were included in the sample of those who participated in the test development. The Court was very impressed with the fact that IOS purposely oversampled minority firefighters in order to avoid racially biased questions.⁵²

The dissent pointed to the testimony of Janet Helms, a professor of counseling psychology at Boston College, who noted that the job analysis information that was submitted by the firefighters was from predominately white firefighters (2/3 of the sample were white). Helms testified that different races will do their jobs in different ways because there is a differential experience between white and black workers and so was a source of bias in the tests.⁵³ So, even though IOS may have tried to come up with a representative sample of firefighters by the inclusion of minority firefighters, the resulting outcome would still produce a bias in the job analysis.

The job analysis, if done correctly, should produce information about the tasks and duties regarding the job. According to the Guidelines, the questions that are used in the selection instrument must be based on the job itself. The extent to which the test questions sample the job

such act or omission, such interpretation or opinion is modified or rescinded or is determined by judicial authority to be invalid or of no legal effect”)

⁴⁹ *Id.* at 2679.

⁵⁰ *Id.*

⁵¹ *Id.* at 2665.

⁵² *Id.*

⁵³ *Id.* at 2694-5.

tasks and requirements, determines the content validity of the test – another requirement articulated in the Uniform Guidelines.

CONTENT VALIDITY OF THE EXAMINATION.

The Uniform Guidelines mandate that any selection test must be content-valid.⁵⁴ In other words, test questions must represent an adequate sampling of the job duties and tasks. Questions that are used should be derived from the job analysis results. The IOS consultants took a different approach. On the written portion of the test they

. . . compiled a list of training manuals, Department procedures, and other materials to use as sources for the test questions. IOS presented the proposed sources to the New Haven fire chief and assistant fire chief for their approval. Then, using the approved sources, IOS drafted a multiple-choice test for each position. Each test had 100 questions, as required by CSB rules, and was written below a 10th-grade reading level. After IOS prepared the tests, the City opened a 3-month study period. It gave candidates a list that identified the source material for the questions, including the specific chapters from which the questions were taken.⁵⁵

Rather than relying on the results of the job analysis, the IOS staff used source materials to create the questions used on the written test. The Majority seemed not to be bothered by this, despite the fact that there was no evidence that the source materials provided a fair representation of the job tasks. In fact, some firefighters testified that the results should not be certified for the reason that several exam questions were not relevant to New Haven procedures and policies.⁵⁶ Some firefighters testified that access to the study materials for the test itself fell along racial lines – minority firefighters were not as able to pay or receive the materials as easily as the white candidates who were able to get the books for free from relatives who were firefighters.⁵⁷

Perhaps the most damning testimony, ignored by the Majority, came from the head of IOS team that developed the test, Chad Legel. Legel testified that he was restricted even in what kind of exam he produced because of the 60/40 rule. Not only that, but because the City was so worried about keeping the tests confidential, no official from the City was allowed to review the content of the test to make sure it was job related.⁵⁸ Instead, a senior officer from Georgia was

⁵⁴ UGESP, 29 C.F.R. §1607.14(C)(4). (“To demonstrate the content validity of a selection procedure, a user should show that the behavior(s) demonstrated in the selection procedure are a representative sample of the behavior(s) of the job in question or that the selection procedure provides a representative sample of the work product of the job.”)

⁵⁵ *Ricci*, at 2666.

⁵⁶ *Id.* at 2692-3. (Hornick also “noted in the literature that [Burgett] sent that the test was not customized to the New Haven Fire Department.” *Id.* at 2687. The Court also noted the City’s argument “that the examinations were not job-related, respondents note some candidates’ complaints that certain examination questions were contradictory or did not specifically apply to firefighting practices in New Haven. But Legel told the CSB that IOS had addressed those concerns—that it entertained “a handful” of challenges to the validity of particular examination questions, that it “reviewed those challenges and provided feedback [to the City] as to what we thought the best course of action was,”) *Id.* at 2679.

⁵⁷ *Id.* at 2687.

⁵⁸ *Id.* at 2706 note 16. (Ginsburg, J., dissenting). (“Under the City’s ground rules, IOS was not allowed to show the exams to anyone in the New Haven Fire Department prior to their administration. This “precluded [IOS] from being able to engage in [its] normal subject matter expert review process”—something Legel described as “very critical. As a result, some of the exam questions were confusing or irrelevant, and the exams may have over-tested some

retained by the City to check the test questions and whether they were related to the source materials that were previously approved by the City and other Districts.⁵⁹ So, all that the Georgia “expert” would be able to do is to say whether or not the test was related to training materials and NOT what happened on the job. Legel testified that the exam was facially neutral and would stand by the validity of the tests.⁶⁰ There is a huge difference between facially valid tests versus tests that are criterion valid. Yet, the Court seemed to accept Legel’s testimony as sufficient in proving the test’s validity. Neither the Uniform Guidelines, nor previous case decisions ever found someone’s word sufficient enough to label a test as being related to the job. It was very telling that the Court took issue with the fact that IOS had contracted to provide the City with a technical report which explained the validity of the tests, yet never produced it. Rather than concluding that the City simply never was given any written information about the test’s validity, the Court instead chastised the City for never requesting it. While the City should have requested the information, this still does not negate the fact that the City had no evidence of the test’s validity which bolsters their position that the test should not be used.⁶¹

More troubling was the total absence of any statistical evidence as to the reliability or the validity of the test. From the testimony, it appears that no statistical validity coefficients were ever calculated to demonstrate the job-relatedness of the test or whether the test had differential validity by race. This is a clear departure from the Uniform Guidelines and should have raised questions about the sufficiency of the written test. Yet, the Majority seemed to be satisfied with testimony that the written exam *appeared* to be related to the job – a step backwards from the requirements of the Uniform Guidelines requirements that a test must be more than just facially neutral.

LACK OF VALIDATION EVIDENCE.

Careful examination of the Uniform Guidelines reveals that proof of a test’s validity requires more than just an expert’s word on the matter. A criterion-validation study is the preferred method for determining a test’s validity where the end result is a validity coefficient

subject-matter areas while missing others. Testimony before the CSB also raised questions concerning unequal access to study materials and the potential bias introduced by relying principally on job analyses from nonminority fire officers to develop the exams”)

⁵⁹ *Id.* at 2666. (Michael Blatchley stated that “[e]very one” of the questions on the written examination “came from the [study] material. ... [I]f you read the materials and you studied the material, you would have done well on the test. Frank Ricci stated that the test questions were based on the Department's own rules and procedures and on “nationally recognized” materials that represented the “accepted standard[s]” for firefighting. Yet, other firefighters claimed that the test questions did not represent the job. “They described the test questions as outdated or not relevant to firefighting practices in New Haven. Gary Tinney stated that source materials “came out of New York.... Their makeup of their city and everything is totally different than ours.” The test materials were criticized for being too expensive at \$500 and too long.)

⁶⁰ *Id.* at 2693.

⁶¹ *Id.* at 2679. (The Court noted “The City, moreover, turned a blind eye to evidence that supported the exams’ validity. Although the City’s contract with IOS contemplated that IOS would prepare a technical report consistent with EEOC guidelines for examination-validity studies, the City made no request for its report. After the January 2004 meeting between Legel and some of the city-official respondents, in which Legel defended the examinations, the City sought no further information from IOS, save its appearance at a CSB meeting to explain how it developed and administered the examinations. IOS stood ready to provide respondents with detailed information to establish the validity of the exams, but respondents did not accept that offer.” Contrary to the Court’s opinion, IOS never produced any evidence of validity that conforms with the Uniform Guidelines.)

which statically measures the relationship between the test scores and some measure of job success. One quick method of validation is to use a concurrent validation study, where incumbent firefighters can take the exam and then their scores would be correlated with a measure of job success. The resulting correlation, a validity coefficient, provides the needed statistical indicator as to whether the test sufficiently is linked to job performance. Yet, because IOS was not allowed to have the resulting test viewed by the firefighters for security reasons, no validation study was attempted. No validity statistics were ever presented at the trial.

Even though the written questions were derived from questionable secondary sources, the oral portion of the exam was more directly related to the job analysis. The IOS team:

“... concentrated on job skills and abilities. Using the job-analysis information, IOS wrote hypothetical situations to test incident-command skills, firefighting tactics, interpersonal skills, leadership, and management ability, among other things. Candidates would be presented with these hypotheticals and asked to respond before a panel of three assessors.”⁶²

The oral questions seemed to have a relationship with actual job tasks, although once again, there were no validity statistics to examine the degree to which the oral exam correlated with job performance.

THE ASSESSORS FOR THE ORAL EXAM.

The testers, or assessors, were of one rank above the candidates and were located from outside the state of Connecticut. Local assessors were not used because previous tests had been so controversial in the past it was felt that nonlocal testers would avoid bias. Previously, passing the tests was not so much based on ability, but on political favoritism and there was much criticism that the decisions were racially biased. Therefore, the test developers felt it was critical that the assessors were not part of the political system. IOS put together a team of 30 assessors which consisted of “battalion chiefs, assistant chiefs, and chiefs from departments of similar sizes to New Haven's throughout the country”⁶³ Great care was taken to make sure that the assessors included 66% minority participants to avoid any racially discriminatory decisions. Each oral exam would be assessed by a three-person panel of assessors and IOS made sure that two of the assessors were minorities. The panelists were trained for several hours as to how to administer and score the tests on the day prior to the administration of the test.⁶⁴

Once again, the intention of the IOS team was good, but the execution was not. By using assessors that were not employed by the City of New Haven guarantees that those individuals are uninformed about the policies and procedures for the City of New Haven and thus, their assessment, at least in part, would suffer. Furthermore, the training of the assessors was insufficient to completely train the assessors. Assessors were trained in one session lasting several hours only on the day before the test was administered. There was no evidence that the reliability or validity of their assessment was tested by the IOS team. More disturbing, the accuracy of their assessment was never tested. It would have been an easy matter to have the assessors score several confederate candidates and then conduct reliability and validity studies on

⁶² *Id.*

⁶³ *Id.*

⁶⁴ *Id.*

the resulting ratings. None of this was done and the Majority never expressed concern over the assessor's ratings.

THE EXPERTS' TESTIMONY REGARDING THE TEST.

Three testing experts provided testimony to the CSB as to their opinion regarding the test's validity. Dr. Christopher Hornick, a testing expert with 25 years of experience in I/O psychology testified that the tests had "relatively high" degrees of disparate impact.⁶⁵ Hornick said the issue of facial neutrality was not the primary concern – it was whether the tests, rules and procedures contribute to the adverse impact of the tests.⁶⁶ Hornick expressed concerns about the 60/40 test required by the union and stated that in his experience, there were many other tests that would be effective at identifying qualified supervisors in the fire department and would not have adverse impact.⁶⁷ His suggestion was to use an assessment center which requires the candidates to perform and act in performance – based situations rather than giving a written test response. The results of the assessment center would be superior in predicting how candidates would actually perform on their job rather than getting a written answer which would not necessarily comport with their on-the-job performance.⁶⁸ Hornick added:

I've spoken to at least 10,000, maybe 15,000 firefighters in group settings in my consulting practice and I have never one time ever had anyone in the fire service say to me, 'Well, the person who answers – gets the highest score on a written job knowledge, multiple-guess test makes the best company officer.' We know that it's not as valid as other procedures that exist.⁶⁹

Hornick stated that the written test was "reasonably good" but was very critical of the City for not allowing the New Haven department officials in the department to review and critique the exam since that would be a principle way of checking to see if the test matched up with the procedures of the department. He stated that if the City was worried about the confidentiality of the tests, the City could require all of the officials to sign confidentiality agreements. When directly asked whether the CSB should certify the test results, he again commented on the fact that the test had disparate impact and that any proceeding would have industrial psychologists on both sides of the issue and the results "would not be a pretty or comfortable position for anyone to be in."⁷⁰ He suggested that the City might certify the results but immediately search for a better type of test that would be less likely to cause disparate impact.

Vincent Lewis, a specialist with the Department of Homeland Security and former fire officer in Michigan, testified that the tests tested relevant material in general, but noted that there was an undue emphasis on duties that were associated with an "apparatus driver", a job which

⁶⁵ *Id.*

⁶⁶ *Id.*

⁶⁷ *Id.* at 2694.

⁶⁸ *Id.*

⁶⁹ *Id.*

⁷⁰ *Id.*

not everyone had performed nor had access to. He expressed concern that not everyone would be able to access the test preparation materials.⁷¹

Thomas Ude, the City's legal counsel expressed concerns about the test, so much so that he advised the CSB should not certify the results. Ude stated that based on the definition of disparate impact and the fact that upon close examination of the questions, the questions appear to measure whether the candidates were able to memorize the test materials and not whether the candidate could do the job. Ude and other officials noted that Hornick and others identified less discriminatory methods of testing and therefore the City would be liable on a disparate impact theory.⁷³

Nevertheless, the pass rate was 64% for whites and 37.5% for Hispanic and Black applicants for the position of captain; 58% for whites; 31.6% for Blacks and only 20% for Hispanic Applicants for the lieutenant's exam.⁷⁴ Since the pass rate is roughly one-half of that of the majority (white) group, this fails the 80 percent or 4/5's test as established by the EEOC.⁷⁵ Even though the results of the test produce a prima facie case of disparate impact, the Court next considered whether the City could demonstrate that there was a valid, less discriminatory test that could have been used.

AVAILABILITY OF ALTERNATIVE, NONDISCRIMINATORY TESTS.

The Uniform Guidelines maintain that one factor in deciding whether an employer should continue to use a test with disparate impact is whether there are alternative, nondiscriminatory tests are available.⁷⁶ The City argued that there were less discriminatory options.

A. The Weighting System. The first, rested on the notion that if the weighting of the two components of the test – the oral and written test – could have been modified to a 30/70 weighting rather than the City's 60/40 weighting, then two black candidates would have been considered for the lieutenant's position and one black candidate for the Captain's position. The Court rejected this argument since there was no proof that the altered weighting neither was equally valid nor was there proof that the original 60/40 weighting system was capriciously set.⁷⁷ Since the weighting was the result of negotiations between the City and the New Haven's firefighters' union, the Court felt it must presume that the weighting system was constructed

⁷¹ *Id.*

⁷³ *Id.* at 2695 (“Upon close reading of the exams, the questions themselves would appear to test a candidate's ability to memorize textbooks but not necessarily to identify solutions to real problems on the fire ground.”).

⁷⁴ *Id.* at 2678.

⁷⁵ *Id.* citing 29 CFR § 1607.4(D) (2008) (selection rate that is less than 80 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”); *Watson*, 487 U.S., at 995-996, n. 3, 108 S.Ct. 2777 (plurality opinion) (EEOC's 80-percent standard is “a rule of thumb for the courts”).

⁷⁶ *Ricci* at 2703, note 11, (Ginsburg, J. dissenting). (“Under the Uniform Guidelines on Employee Selection Procedures (Uniform Guidelines), employers must conduct “an investigation of suitable alternative selection procedures.” Quoting 29 CFR § 1607.3(B)). See also *Officers for Justice v. Civil Serv. Comm'n*, 979 F.2d 721, 728 (C.A.9 1992) (“before utilizing a procedure that has an adverse impact on minorities, the City has an obligation pursuant to the Uniform Guidelines to explore alternative procedures and to implement them if they have less adverse impact and are substantially equally valid”).

⁷⁷ *Id.* at 2679.

based on rational reasons.⁷⁸ Finally, changing the weighting system because of the test results would be an impermissible altering of the test results on the basis of race.⁷⁹

Donald Day, a representative of the Northeast Region of the International Association of Black Professional Firefighters, testified that the situation in New Haven had some similarity to that of Bridgeport. In Bridgeport, there used to be a similar weighting of a written exam (70%); and oral exam (25%) and seniority (5%). However, the City of Bridgeport came to the realization that the oral component of the exam was much more effective at measuring real life scenarios that were relevant for the job. Changing the relative weights of the exam allowed Bridgeport to give the oral part of the exam the heaviest weight and consequently, minorities were now “fairly represented” in the various positions.⁸⁰

B. The Rule of Three. The Respondents argued that the City could have adopted a different interpretation of the “rule of three” which would interpret less discriminatory results. The City’s charter required that the only individuals who would be promoted would be those with the three highest scores on the promotional tests.⁸² The alternative would be to “band” the results which meant that the test scores would be rounded up to the next whole number and then all candidates with the same score would be considered as falling within the same rank and so more candidates could be considered for the position which would have allowed more minority candidates to be considered. However, banding was struck down as an impermissible practice by the state court.⁸³ The state court’s decision would not have automatically disqualified banding as a viable alternative,⁸⁴ but it would have been adopted as measure to correct the racial disparity of the City’s test results which is an impermissible change due to race.⁸⁵

C. Assessment Centers. Hornick had been consulted by the CSB as to the existence of alternative testing methods. It was Hornick’s opinion that an assessment center should have been used because it was more job-related and would produce less disparate impact than the current procedure. Other districts had used the assessment centers for their firefighters with much success and no disparate impact.⁸⁶ Yet, the Court would not rely on Hornick’s statement alone as raising a genuine issue of material fact that assessment centers would be more valid.

There were several reasons cited for discounting Hornick’s statements. One reason was the testimony by the Respondents and statements made by the City that the assessment center would not have been available for use.⁸⁷ Some noted it would take a very long time to set up the

⁷⁸ *Id.*

⁷⁹ *Id.* See 42 U.S.C. § 2000e-2(1).

⁸⁰ *Id.*

⁸² *Id.* citing New Haven, Conn., Code of Ordinances, Tit. I, Art. XXX, § 160 (1992).

⁸³ *Id.* at 2680, citing Kelly v. New Haven, No. CV000444614, 2004 WL 114377, (Conn. Super. Ct., Jan.9, 2004).

⁸⁴ *Id.* See 42 U.S.C. § 2000e-7.

⁸⁵ *Id.*

⁸⁶ *Id.* at 2705. (Ginsburg, J. dissenting). (“Although comprehensive statistics are scarce, a 1996 study found that nearly two-thirds of surveyed municipalities used assessment centers (“simulations of the real world of work”) as part of their promotion processes.” Citing to P. Lowry, A Survey of the Assessment Center Process in the Public Sector, 25 PUBLIC PERSONNEL MANAGEMENT 307, 315 (1996).)

⁸⁷ *Id.* at 2705 at note 15. (Ginsburg, J. dissenting) (“...with respect to assessment centers, the Court identifies ‘statements to the CSB indicat [ing] that the Department could not have used [them] for the 2003 examinations.’

assessment center due to the need of selecting the assessment instruments and the training of the assessors.⁸⁸ Interestingly enough, the Respondents had previously testified that Hornick had pressured the City to reject the test results, but the Court said that the bulk of Hornick's comments actually supported the certification of the test results.⁸⁹ The Court concluded that the strong basis in evidence test can't be fulfilled by a few "stray remarks"⁹⁰ and that Hornick's remarks did not rise to the level of the strong basis of evidence test and concluded that the City did not have an acceptable reason for discarding the test results.⁹¹ The decision was also predicated on the fact that the test takers had relied on the test's credibility and spent much time and money in preparation for the test and would be harmed by the City if it discarded the test.

It was also noted that Hornick was a competitor of IOS and seemed to be more intent on marketing his own services to the CSB. The Court felt that he had not the opportunity to examine the test in detail and the Court concluded that the City did not have an acceptable reason for discarding the test results.⁹² The decision was also predicated on the fact that the test takers had relied on the test's credibility and spent much time and money in preparation for the test and would be harmed by the City if it discarded the test. The bottom line of the decision was that the City did not have sufficient reason to discard the test results and that if a disparate impact case had been brought against the City, the City would have had an adequate defense that if it had not certified the test results, it would have had to face a disparate treatment charge.⁹³

The Majority clearly felt that the City had been not only incorrect in refusing to certify the results of the test, but that it had willfully ignored their own evidence that the tests were job-related. The City had capriciously and hastily discarded the tests solely because of the results of the discriminatory results and in the eyes of the Majority had performed impermissible disparate treatment.

THE DISSENT'S VIEW

Justice Ginsburg, in her Dissent, stated that the context of the case matters.⁹⁴ There is no doubt that discrimination and political forces were at work in New Haven and although the applicants deserve some sympathy if the test results were discarded, they did not have a vested

Ante, at 2680 - 2681. The Court comes up with only a single statement on this subject—an offhand remark made by petitioner Ricci, who hardly qualifies as an expert in testing methods." See ante, at 2686.)

⁸⁸ *Id.* at 2670. (Firefighter Frank Ricci, in arguing for certification stated that although "assessment centers in some cases show less adverse impact," they were not available alternatives for the current round of promotions. It would take several years, Ricci explained, for the Department to develop an assessment-center protocol and the accompanying training. Justice Ginsburg criticized the Court for relying on Ricci's statement since he was not a testing expert.)

⁸⁹ *Id.* (Hornick stated that adverse impact in standardized testing "has been in existence since the beginning of testing," and that the disparity in New Haven's test results was "somewhat higher but generally in the range that we've seen professionally." He told the CSB he was "not suggesting" that IOS "somehow created a test that had adverse impacts that it should not have had." And he suggested that the CSB should "certify the list as it exists.")

⁹⁰ *Id.*

⁹¹ *Id.* at 2680-1.

⁹² *Id.* at 2680-1.

⁹³ *Id.* at 2681.

⁹⁴ *Id.* at 2689, (Ginsburg, J., dissenting).

right in the promotions. Nor were any candidates promoted in preference to others so all of the candidates were in a similar situation when the test results were not certified.⁹⁵

Ginsburg has harsh words for the majority by stating that the majority is “pretending” that the City is discarding the test results solely on the racial outcomes and that the Majority pointedly ignored many serious flaws within the test.⁹⁶ The opinion also takes the Majority to task for ignoring the central decision in *Griggs v. Duke Power*⁹⁷ and thus the decision “will not have staying power”.⁹⁸

The Dissent criticized the City for never having considered what type of examination would be appropriate to test the job knowledge of the candidates but instead, just decided to follow a 20-year-old system that had its genesis in a discriminatory environment. The City testified that it would entertain only “proposals that include a written component that will be weighted at 60%, and an oral component that will be weighted at 40%.”⁹⁹ Legel, the representative from IOS testified on his deposition that he was never asked by the City if there were other viable tests to test firefighter’s abilities.¹⁰⁰ The results not only provided prima facie proof of disparate impact, but a closer analysis revealed that for the lieutenant’s position, the highest scoring African-American candidate ranked 13th; the top Hispanic candidate was ranked 26th. As for the seven then-vacant captain positions, two Hispanic candidates would have been eligible, but no African-Americans. The highest scoring African-American candidate ranked 15th.¹⁰¹ No minorities were even close to promotion eligibility. Thus the City clearly had “cause for concern about the prospect of Title VII litigation and liability.”¹⁰²

The Majority created an impression that the City came to its decision to not certify the results in a capricious and cavalier manner. Contrary to the impression that the Majority was trying to create, the issue regarding the validity of the test was carefully examined by the New Haven’s Civil Service Board which was the entity empowered to certify the test results. The resulting testimony painted a very different picture than the Majority presented. The CSB was given the legal perspective by the city’s corporate counsel, Thomas Ude who proceeded to explain the disparate impact standard to the CSB. He did state that if the test is proved prima facie to have disparate impact, than the employer has the burden of demonstrating the test is job related and consistent with business necessity.¹⁰³ So, the primary task of the CSB, according to the Dissent, was to decide whether the exam was reliable, valid, and whether a less discriminatory test was available.¹⁰⁴ All in all, there were five public meetings to determine

⁹⁵ *Id.* at 2690. (Ginsburg, J. dissenting).

⁹⁶ *Id.*

⁹⁷ *Id.* Quoting *Griggs v. Duke Power Co.*, 401 U.S. 424 (1971).

⁹⁸ *Id.*

⁹⁹ *Id.* at 2691.

¹⁰⁰ *Id.* at 2692. (Legel testified “I was under contract and had responsibility only to create the oral interview and the written exam.”)

¹⁰¹ *Id.*

¹⁰² *Id.*

¹⁰³ *Id.*

¹⁰⁴ *Id.*

this.¹⁰⁵ A broad spectrum of individuals testified at the meetings -- from test takers, the test designer, subject-matter experts, City officials, union leaders, and community members.¹⁰⁶ Some of the firefighters stated that they had put a lot of time and money into the preparation for taking the tests and testified that the tests should be certified.¹⁰⁷

The final vote was a tied vote (2-2) and by the rules of the board, the decision was not to certify. Some members stated that they had voted against certification because of the compelling testimony they had heard. Others were not convinced that the tests were not job-related.¹⁰⁸

The Dissent claims that the decision in *Ricci* flies in the face of established case law and statutory requirements.¹⁰⁹ The view that the City's compliance with Title VII and *Griggs* in addressing a test that has been demonstrated to have disparate impact as equal to that of committing an act of disparate treatment "shows little attention to Congress' design or to the *Griggs* line of cases Congress recognized as pathmarking."¹¹⁰ Here, "Title VII's disparate-treatment and disparate-impact proscriptions must be read as complementary."¹¹¹ The Court summarized the issue this way:

In codifying the *Griggs* and *Albemarle* instructions, Congress declared unambiguously that selection criteria operating to the disadvantage of minority group members can be retained only if justified by business necessity. In keeping with Congress' design, employers who reject such criteria due to reasonable doubts about their reliability can hardly be held to have engaged in discrimination "because of" race. A reasonable endeavor to comply with the law and to ensure that qualified candidates of all races have a fair opportunity to compete is simply not what Congress meant to interdict. I would therefore hold that an employer who jettisons a selection device when its disproportionate racial impact becomes apparent does not violate Title VII's disparate-treatment bar automatically or at all, subject to this key condition: The employer must have good cause to believe the device would not withstand examination for business necessity.¹¹²

The Court noted that no one attempted to justify the 60/40 weighting and no mention was made of the EEOC Guidelines. "[B]y the enactment of title VII," the guidelines state, "Congress did not intend to expose those who comply with the Act to charges that they are violating the very statute they are seeking to implement."¹¹³ Justice Ginsberg made note of the fact that the Guidelines had also been accorded due deference,¹¹⁴ yet such was not the case here.

The Dissent found that the strong basis in evidence test was vaguely stated and left unanswered as to what factors would satisfy the standard.¹¹⁵ Furthermore, the test is based on

¹⁰⁵ *Id.*

¹⁰⁶ *Id.*

¹⁰⁷ *Id.*

¹⁰⁸ *Id.*

¹⁰⁹ *Ricci*, at 2699.

¹¹⁰ *Id.*

¹¹¹ *Id.*

¹¹² *Id.* relying on *Faragher v. Boca Raton*, 524 U.S. 775, 806 (1998) (observing that it accords with "clear statutory policy" for employers "to prevent violations" and "make reasonable efforts to discharge their duty" under Title VII).

¹¹³ 29 CFR § 1608.1(a) (2008).

¹¹⁴ *Ricci*, at 2700. citing *Albemarle*, 422 U.S., at 431; *Griggs*, 401 U.S., at 434.

¹¹⁵ *Id.*

the Equal Protection standard – a standard which does not cover disparate impact.¹¹⁶ Justice Ginsburg also criticizes the cases that were cited by the Majority since they are concerned with absolute racial preferences¹¹⁷ and are not comparable to the current case which tries to avoid Title VII.¹¹⁸ Title VII's mandate is that it “should not be read to thwart efforts at voluntary compliance.”¹¹⁹ Despite this, the Majority's adoption of its vague standard “makes voluntary compliance a hazardous venture.”¹²⁰ The Dissent continues:

As a result of today's decision, an employer who discards a dubious selection process can anticipate costly disparate-treatment litigation in which its chances for success—even for surviving a summary-judgment motion—are highly problematic. Concern about exposure to disparate-impact liability, however well grounded, is insufficient to insulate an employer from attack. Instead, the employer must make a “strong” showing that (1) its selection method was “not job related and consistent with business necessity,” or (2) that it refused to adopt “an equally valid, less-discriminatory alternative.”¹²¹

The Majority's opinion is at odds with the voluntary compliance ideal. In addition, the Court imposed a stricter standard than had been used in the past. In *Croson*, the Court found no strong basis in evidence because the City had offered “nothing approaching a prima facie case.”¹²² In other words, the previous view was that the strong basis in evidence had been met when a prima facie case of disparate impact had been proven.

The Majority seemed to indicate that a heightened standard was necessary because to do otherwise would “upset[s] an employee's legitimate expectation.”¹²³ In other words, the employee would have a legitimate expectation that the test results would be honored. In the Dissent's eyes, “The legitimacy of an employee's expectation depends on the legitimacy of the selection method. If an employer reasonably concludes that an exam fails to identify the most qualified individuals and needlessly shuts out a segment of the applicant pool, Title VII surely does not compel the employer to hire or promote based on the test, however unreliable it may be.

¹¹⁶ *Id.* See *Personnel Administrator of Mass. v. Feeney*, 442 U.S. 256, 272 (1979); *Washington v. Davis*, 426 U.S. 229, 239 (1976).

¹¹⁷ *Id.* See *Wygant*, 476 U.S. at 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, (rejecting a set-aside program for minority contractors that operated as “an unyielding racial quota”).

¹¹⁸ *Id.* at 2700.

¹¹⁹ *Id.* Quoting *Johnson v. Transportation Agency, Santa Clara Cty.*, 480 U.S. 616, 63 (1987). Such compliance, the Court notes, is “the preferred means of achieving [Title VII's] objectives.” *Firefighters v. Cleveland*, 478 U.S. 501, 515 (1986). See also *Kolstad v. American Dental Assn.*, 527 U.S. 526, 545 (1999) (“Dissuading employers from [taking voluntary action] to prevent discrimination in the workplace is directly contrary to the purposes underlying Title VII.”); 29 CFR § 1608.1(c).

¹²⁰ *Id.* at 2701.

¹²¹ *Id.*

¹²² *J.A. Croson Co.*, at 500. The Court did not suggest that anything beyond a prima facie case would have been required. In the context of race-based electoral districting, the Court has indicated that a “strong basis” exists when the “threshold conditions” for liability are present. *Bush v. Vera*, 517 U.S. 952, 978, (1996) (plurality opinion).

¹²³ *Ricci*, at 2677.

Indeed, the statute's prime objective is to prevent exclusionary practices from “operat[ing] to ‘freeze’ the status quo.”¹²⁴

Second, the Court suggests, anything less than a strong-basis-in-evidence standard risks creating “a de facto quota system, in which ... an employer could discard test results ... with the intent of obtaining the employer's preferred racial balance.”¹²⁵ The Dissent points out that the Majority firmly believed that the CSB refused to certify the test results solely because the results demonstrated statistical disparate impact. The Dissent strongly disagreed.

The Dissent criticized the departure of the Majority in how they implemented the strong basis in evidence standard. The typical procedure for applying a new standard is to remand the case to allow the employer an opportunity to conform to the new standard, yet that was not done in *Ricci*.¹²⁶ The only stated reason for the departure was that the City showed a significant statistical disparity. The Dissent regarded the proclamation as “false pretention”¹²⁷ This act, according to the Dissent, precludes the City from invoking 42 U.S.C. § 2000e-12(b) as a shield to liability. Section 2000e-12(b) provides:

In any action or proceeding based on any alleged unlawful employment practice, no person shall be subject to any liability or punishment for or on account of (1) the commission by such person of an unlawful employment practice if he pleads and proves that the act or omission complained of was in good faith, in conformity with, and in reliance on any written interpretation or opinion of the [EEOC] Such a defense, if established, shall be a bar to the action or proceeding, notwithstanding that (A) after such act or omission, such interpretation or opinion is modified or rescinded or is determined by judicial authority to be invalid or of no legal effect¹²⁸

The EEOC guidelines expand on this by stating that employers may “take affirmative action based on an analysis which reveals facts constituting actual or potential adverse impact.”¹²⁹ The Dissent argues that it is far less of a drastic measure to disregard the results of a test than it is to implement an affirmative action program.¹³⁰

It was clear to all parties that the tests had disparate impact.¹³¹ Since the disparity was so great, the City took steps to hear testimony from experts and laypersons to see whether the exams were fair and consistent with business necessity.¹³² Ginsburg noted that the testimony that was presented gave rise to “grave concerns” about the tests.¹³³ The City had failed to

¹²⁴ *Id.* at 2702. quoting *Griggs*, 401 U.S., at 430.

¹²⁵ *Id.* at 2702 and 2675.

¹²⁶ *Id.* See, e.g., *Johnson v. California*, 543 U.S. 499, 515 (2005); *Pullman-Standard v. Swint*, 456 U.S. 273, 291 (1982).

¹²⁷ *Id.*

¹²⁸ 42 U.S.C. § 2000e-12(b) (2010).

¹²⁹ 29 CFR §§ 1608.3 and 1608.4 (2008) and § 1608.3(a).

¹³⁰ *Ricci*, at 2703.

¹³¹ *Id.*

¹³² *Id.*

¹³³ *Id.*

consider alternative methods and adhered to the 60/40 weighting system without ever questioning whether this weighting or methodology was the best way to determine the candidates' aptitude for the job.¹³⁴ Under the Uniform Guidelines on Employee Selection Procedures (Uniform Guidelines), employers must conduct "an investigation of suitable alternative selection procedures."¹³⁵ The Dissent strongly disagreed with the presumption of the Majority that the 60/40 weighting procedure had been adopted for rational reasons just because it was the result of a collective bargaining agreement.¹³⁶ Collective bargaining agreements can easily run afoul of Title VII.¹³⁷

Not only was the weighting a source of concern, but the written test caused some concern for the Dissent. Ginsburg noted that written tests are simply not appropriate for testing firefighters since written tests tend to measure the ability of the test-taker to recall the material from study materials rather than measuring what the candidate would do in the situation.¹³⁸ A fire officer's job, courts have observed, "involves complex behaviors, good interpersonal skills, the ability to make decisions under tremendous pressure, and a host of other abilities-none of which is easily measured by a written, multiple choice test."¹³⁹ The EEOC's Uniform Guidelines on Employee Selection Procedures have stated, measures of "interpersonal relations" or "ability to function under danger (e.g., firefighters)," "[p]encil-and-paper tests ... generally are not close enough approximations of work behaviors to show content validity."¹⁴⁰

The concern of the Majority was that the test results were being discarded because of the disparate impact of the test and in order to successfully defend its action, the City would have to prove a disparate impact case against itself.¹⁴¹ It was the Majority's argument that the test scores could not have been changed without violating Title VII. The Dissent said this wasn't the issue – the issue was whether the City had the opportunity to use other, less discriminatory tests such as the assessment center. The Majority stated that the City was not able to use the assessment

¹³⁴ *Id.*

¹³⁵ 29 CFR § 1607.3(B). *See also* *Officers for Justice v. Civil Serv. Comm'n*, 979 F.2d 721, 728 (C.A.9 1992) ("before utilizing a procedure that has an adverse impact on minorities, the City has an obligation pursuant to the Uniform Guidelines to explore alternative procedures and to implement them if they have less adverse impact and are substantially equally valid").

¹³⁶ *Ricci*, at 2703.

¹³⁷ *Id.* *See, e.g.*, *Peters v. Missouri-Pacific R. Co.*, 483 F.2d 490, 497 (C.A.5 1973) (an employment practice "is not shielded [from the requirements of Title VII] by the facts that it is the product of collective bargaining and meets the standards of fair representation").

¹³⁸ *Id.* at 2704 .” *Vulcan Pioneers, Inc. v. New Jersey Dept. of Civil Serv.*, 625 F.Supp. 527, 539 (NJ 1985).

¹³⁹ *Firefighters Inst. for Racial Equality v. St. Louis*, 616 F.2d 350, 359 (C.A.8 1980).

¹⁴⁰ 29 CFR § 1607.15(C)(4). *See also* *Nash v. Jacksonville*, 837 F.2d 1534,1538 (C.A.11 1988), vacated, 490 U.S. 1103, 109 S.Ct. 3151, 104 L.Ed.2d 1015 (1989), opinion reinstated, 905 F.2d 355 (C.A.11 1990) ("the examination did not test the one aspect of job performance that differentiated the job of firefighter engineer from fire lieutenant (combat): supervisory skills"); *Firefighters Inst. for Racial Equality v. St. Louis*, 549 F.2d 506, 512 (C.A.8 1977) ("there is no good pen and paper test for evaluating supervisory skills"); *Boston Chapter, NAACP*, 504 F.2d, at 1023 ("[T]here is a difference between memorizing ... firefighting terminology and being a good fire fighter. If the Boston Red Sox recruited players on the basis of their knowledge of baseball history and vocabulary, the team might acquire [players] who could not bat, pitch or catch.").

¹⁴¹ *Id.*

center, but their opinion was based on a statement by Ricci who was not an expert on testing and was one of the petitioners in the case.¹⁴²

The central foundation to the Majority's acceptance of the City's testing system was the assertion that the exams as "painstaking[ly]" developed to test "relevant" material and on that basis finds no substantial risk of disparate-impact liability¹⁴³ as constructed by IOS.¹⁴⁴ The showing of the construction of the tests might have sufficed under a *Ward's Cove* proof, but *Ward's Cove* was effectively overturned by the Civil Rights Act of 1991.¹⁴⁵ The City was operating under the notion that it could only use the 60/40 weighting of the exam and no alternatives were even considered. Nor were some of the essential job-related KSAs even measured. For example, "command presence" was never measured. So, while it seemed like the appropriate steps were taken in constructing the exam, there were serious questions as to whether the exam actually was sufficient in measuring job-related KSAs.¹⁴⁶

THE EEOC UNIFORM GUIDELINES AND TESTING STATISTICS

Another disturbing fact was the restriction that the final exam could not be shown to anyone in the department. Legel, the head of IOS said that expert review of the final product was very critical. Since the experts could not review the test questions or the job-relatedness of the questions, some of the questions were confusing and particular areas of the job were over-tested.¹⁴⁷ Something that neither the Majority nor the Dissent mentioned in any great detail is the requirements that is in the EEOC Guidelines on Employee Selection Procedure. The Guidelines mandate that any selection test must be reliable, valid and non-discriminatory. The acceptable proof of these factors is statistical correlations which will demonstrate with subjectivity whether the exam is reliable and valid.

There is no evidence that Legel and his associates ever presented statistics for these tests. In fact, given the prohibition on showing the exam to anyone in the department makes it impossible for IOS to run the kinds of statistical tests to establish the reliability and validity of the test. For example, one of the methods to establish the reliability, or consistency of the test is known as test-retest, where the exam in question is given to a sample of test-takers and one time. The scores are collected and then the test is given to that same sample again at a later time. The later scores are collected and correlated with the test scores taken the first time the test is given. If the reliability coefficient exceeds .85, the test is considered reliable. Consistency of measurement is absolutely critical for any test, yet nothing was mentioned about this important concept.

Finally, the only mention made of the validity of the test was that of face or content validity. Content validity is the least preferred way of assessing the validity of a test and should never be used alone. Content validity is the process that Legel used – where someone examines

¹⁴² *Id.*

¹⁴³ *Id.* at 2778.

¹⁴⁴ *Id.* at 2705.

¹⁴⁵ *Id.*

¹⁴⁶ *Id.* Cf. *Boston Chapter, NAACP v. Beecher*, 504 F.2d 1017, 1021-1022 (C.A.1 1974) ("A test fashioned from materials pertaining to the job ... superficially may seem job-related. But what is at issue is whether it demonstrably selects people who will perform better the required on-the-job behaviors.")

¹⁴⁷ *Id.*

the content of the test and examines the job description of the job and makes a judgment as to whether or not the two are related. It is a highly subjective method and is prone to all sorts of error. The recommended process contained in the Uniform Guidelines is to use predictive validation or its less preferable, but acceptable counterpart – concurrent validation.¹⁴⁸ The City could have at least done a concurrent validation study where the test would have been given to incumbent firefighters in both positions. The test scores would have been collected and then correlated with some measure of job success – most likely, the performance appraisal. At least that would have provided a validity coefficient to allow the City to know if the test was within the range of acceptable validity. The City was so worried about any security breach that it prevented IOS from even having job incumbents review the test items let alone allowing an empirical validation study. Since there was no evidence of any empirical validation studies, the City was perfectly logical in concluding that there would be a significant likelihood that they would lose a disparate impact claim.

It is troubling that the Majority did not even rely on the EEOC Guidelines but instead, were satisfied that the test was criterion-valid (job-related) based on testimony concerning the test's content validity. It is ironic that the Majority criticizes the City for failing to request from Legel and the IOS the technical report which purportedly would have established the test's validity, but as Justice Ginsburg pointed out, Legel had testified that the technical report only contained an account of the methodology of the test development and would not provide the so-called detailed information regarding the validity of the test.¹⁴⁹

One other criticism that the Dissent had was that the test scores were not precise enough to allow meaningful interpretation of the ranking process. A difference of one point would decide whether a candidate was qualified or not qualified for the job. Common sense would tell anyone that one point should not be the deciding factor on whether or not someone is qualified for the job and the Dissent notes that other previous exams have been invalidated on that basis.¹⁵⁰

The Dissent concluded that the City of New Haven had good reason to fear a disparate impact action. The decision to not certify the exam was not solely based on the racial disparity of the test results, but on many deficiencies in the test – deficiencies that raised serious issues as to whether or not the test was really job-related and consistent with business necessity.¹⁵¹

In her dissent, Justice Ginsburg stated that the ruling “will not have staying power.”¹⁵² The ruling of the *Ricci* Majority seems to be a return to the *Ward's Cove* analysis where an employer would only have to show a legitimate business reason for using the test in order to avoid disparate impact liability. The City of New Haven decided not to certify the test results for firefighters in New Haven in order to avoid the certainty of a disparate impact suit. In doing so, the City was then facing the possibility of a disparate treatment lawsuit filed by the firefighters who would stand to gain from the test results, i.e., promoted to lieutenant or captain.

¹⁴⁸ 29 CFR §1607.5

¹⁴⁹ *Id.* at 2706-7.

¹⁵⁰ *Id.* at 2706. See, e.g., *Guardians Assn. of N.Y. City Police Dept. v. Civil Serv. Comm'n*, 630 F.2d 79, 105 (C.A.2 1980) (“When a cutoff score unrelated to job performance produces disparate racial results, Title VII is violated.”); *Vulcan Pioneers, Inc. v. New Jersey Dept. of Civil Serv.*, 625 F.Supp. 527, 538 (NJ 1985) (“[T]he tests here at issue are not appropriate for ranking candidates.”).

¹⁵¹ *Id.* at 2707.

¹⁵² *Id.* at 2690 (Ginsburg, J., dissenting).

The Majority seemed to have ignored all of the case and statutory law as pointed out by the Dissent and decided that the sole reason for the decision by the City was because of the racially skewed test results. In doing so, the Majority seemed to ignore all of the expert testimony that was heard by the CSB during several days of public hearings. The testimony seemed to provide ample evidence that the test was not only flawed, but that other better alternatives existed which were more job-related and less discriminatory. Yet, the Majority claimed that the City only acted solely because of the racial disparity and because the decision was made not to certify the test results, that the City had committed impermissible disparate treatment.

THE FALLOUT FROM RICCI.

It is no surprise that the *Ricci* decision has caused concern and consternation amongst commentators on employment law. Recent articles have expressed concern that employers will now be more reluctant to take voluntary steps to address testing practices having disparate impact.¹⁵⁴ In fact, since the *Ricci* Court seemed to revert back to the *Ward's Cove* approach, the employer only needs to demonstrate a test's furtherance of a legitimate business interest to defend against a claim of disparate impact. If the *Ricci* standard holds, then it would be virtually impossible for an employer to voluntarily remove a test that has disparate impact. In fact, the *Ricci* decision encourages employers to disregard evidence of a test's disparate impact rather than try to adopt a voluntary solution and "relieves employers of the burden of undertaking costly validation processes to determine whether the employment examination is job-related and a business necessity".¹⁵⁵ Furthermore, the standard "frustrates the purpose of Title VII by discouraging employers from voluntarily taking certain measures to eradicate discrimination, a key goal of Title VII".¹⁵⁶

Cheryl Harris and Kimberly West-Faulcon go so far as to say that the *Ricci* Court does not "test discrimination on a level playing field" but rather " 'whitens' discrimination and 'races' test fairness."¹⁵⁷ Before *Ricci*, the Court never would find that an employer would be found guilty of disparate treatment for addressing tests with disparate impact.¹⁵⁸ *Ricci*'s innovation was "to protect discriminatory testing that favors whites by doctrinally treating the City's efforts to avoid disparate impact liability as conclusive evidence of an impermissible racial motive that violates the statute's proscription against disparate treatment. Ameliorating disparate impact against minorities is thus secondary to the core concern of protecting white interests."¹⁵⁹ The authors seem to suggest that the strong-basis-in-evidence standard will not only subvert the original goal of Title VII, but will protect and enhance the status quo of white workers.¹⁶⁰

Furthermore, the authors claim that the *Ricci* decision has had the effect of "racing" test fairness. They explain:

¹⁵⁴ *Id.*

¹⁵⁵ *Id.*, at 482.

¹⁵⁶ *Id.*, at 483.

¹⁵⁷ Harris, Cheryl I. and West-Faulcon, Kimberly. *Reading Ricci: Whitening Discrimination, Racing Fairness*, 58 UCLA LAW REVIEW 73, 73 (2010).

¹⁵⁸ *Id.* at 82.

¹⁵⁹ *Id.*

¹⁶⁰ *Id.*

We use the term “race” as a verb here to refer to the attribution either in political discourse or legal doctrine of an impermissible racial motive to racially attentive but explicitly antiracist and formally race-neutral actions and practices. Calls for employers to adhere to race-neutral and long-standing psychometric best practices are viewed suspiciously as racial preference policies in disguise.¹⁶¹

The authors conclude that while the *Ricci* Court may find it acceptable to be concerned about racial bias in the pre-testing phase, but post-testing assessment of racial imbalances using professional standards will be seen as espousing racial preferences.¹⁶² The *Ricci* court treated the tests in question as “presumptively valid” with virtually no evidence of the veracity of that assumption.¹⁶³ The question remains as to how subsequent court decisions will apply the *Ricci* standard.

POST-RICCI CASES

Subsequent cases seem to apply the *Ricci* standard only in very specific circumstances. For example, in *Carroll v. the City of Mount Vernon*,¹⁶⁴ the District Court refused to apply the *Ricci* strong basis in evidence test for the simple reason that the standard only applies in situations where a defendant must engage in disparate treatment to avoid disparate impact discrimination. Carroll was a Caucasian male who was a firefighter for the City of Mount Vernon. Carroll attempted to use the *Ricci* precedent to bring a claim against the City of Mount Vernon of discrimination for failing to promote him to the position of Fire Lieutenant. Carroll took the Civil Service exam and was ranked third on the list.¹⁶⁵ He became eligible for an open position, but the President of the Vulcan Society objected to any promotion from the list of candidates stating that the list violated a Consent Decree between the Vulcan Society and the City of Mount Vernon.¹⁶⁶ The Vulcan Society was a fraternal organization that represented the interests of African American Firefighters and their consent decree with the City to promote firefighters in a manner which “reflects the proportion of Blacks among the firefighters of [the] Defendant Fire Department.”¹⁶⁷ The primary issue of contention was that Carroll claimed that although the Mayor's Chief of Staff stated that although everything “looked good” for Carroll's promotion, he would not be promoted because the Vulcan Society had challenged and opposed the promotion of a Caucasian candidate.¹⁶⁸ Carroll filed a claim of Title VII disparate treatment discrimination against the City. Because the promotion was denied to the Plaintiff for race-based reasons, the Plaintiff relied on the *Ricci* decision, claiming that the City's refusal to promote him was analogous to the City of New Haven's refusal to certify the test results.¹⁶⁹ The Court was

¹⁶¹ *Id.* at 83.

¹⁶² *Id.* While *Ricci* appears to support the legality of race-conscious employer actions during the test-design stage and before administering a test, it effectively weakens expert fairness standards: Post-testing evaluation of racially skewed tests according to scientific professional testing standards is, under *Ricci*, a racial preference.

¹⁶³ *Id.*

¹⁶⁴ *Carroll v. the City of Mount Vernon*, 707 F.Supp.2d 449, 455 (S.D.N.Y. Apr 26, 2010).

¹⁶⁵ *Id.* at 451.

¹⁶⁶ *Id.*

¹⁶⁷ *Id.* at 452, note 5.

¹⁶⁸ *Id.* at 452.

¹⁶⁹ *Id.* at 454-455.

quick to point out that the *Ricci* case was quite different since the court only had to determine whether allowing Carroll's promotion was a violation of the City's Consent decree, not whether the procedure for promotion had disparate impact. The court determined the Consent Decree not to be discriminatory on its face and thus was very different than the *Ricci* promotional tests. If the Court were to "find that defendants cannot take time to consider potentially valid objections merely because they relate to the representation of racial minorities, it would not only discourage compliance with court-sanctioned efforts at achieving diversity and equal opportunity, such as the Consent Decree, but it would also discourage the voluntary compliance that is 'the preferred means of achieving the objectives of Title VII.'"¹⁷⁰

A recent case seems to both support and ignore the *Ricci* analysis. In the *NAACP v North Hudson*,¹⁷¹ the Court was asked to determine whether the City of North Hudson had engaged in disparate treatment when it changed the residency requirement for North Hudson Regional Fire and Rescue's ("NHRFR"). The residency requirement was part of the hiring requirements established by the Department of Personnel (DOP). Candidates for jobs within the department were required to pass a test developed by the DOP and to reside in the "Member Municipalities" at the time the test was taken. A class action was brought against the City because the residency requirements had disparate impact against African Americans.

In deciding the case, the Court relied on *Ricci* and said "Essentially, a prima facie case of disparate impact discrimination involves a threshold showing that some employment practice causes a significant statistical disparity, or has the effect of denying members of one race equal employment opportunities."¹⁷² It is rather ironic that the Court used this particular quote from *Ricci* since the *Ricci* court had specifically remarked that statistics alone do not make for a case of disparate impact. Nonetheless, the court in *North Hudson* gave great weight to the expert testimony of the plaintiffs' expert, Dr. Richard Wright, "a professor of Geography whose scholarly work focuses on the operation of labor and housing markets in the United States."¹⁷³ Dr. Wright demonstrated statistics which showed that African Americans were statistically more likely to work in state or local government than in the private sector. Statistics were also used to see if the number of African Americans actually employed at the NHRFR significantly differed from expected probabilities derived from African Americans that were available for work in the relevant labor market.¹⁷⁴ It was Dr. Wright's conclusion that the results "clearly shows that African Americans can and do work in substantial numbers and proportions for municipal governments within [NHRFR's] labor market area."¹⁷⁵

¹⁷⁰ *Id* at 457-458 quoting *Local No. 93, Int'l Ass'n of Firefighters AFL-CIO C.L.C. v. City of Cleveland*, 478 U.S. 501, 515, (1986); *See Alexander v. Gardner-Denver Co.*, 415 U.S. 36, 44 (1974) ("Cooperation and voluntary compliance" are preferred means for achieving goals of Title VII).FN12 Employers should be permitted to "self-evaluate," *Abelmarle Paper*, 422 U.S. at 417-18, 95 S.Ct. 2362, and take that "hard look," *Ricci*, 129 S.Ct. at 2678, without fear that doing so would itself violate Title VII.

¹⁷¹ *National Association for the Advancement of Colored People v. North Hudson Regional Fire & Rescue*, 2010 WL 3810632, *1 (D.N.J. Sep 21, 2010).

¹⁷² *City of North Hudson*, at 5 quoting *Ricci* at 2673.

¹⁷³ *Id.* at 5.

¹⁷⁴ *Id.* at 6-7.

¹⁷⁵ *Id.* at 8.

The City's expert, Dr. Bernard Siskin, specializes in the application of statistics to the analysis of employment practices.¹⁷⁶ He analyzed what would happen should the residency requirement be lifted and what would happen should the residency requirement include some of the areas that the City wanted to add. In the first instance, the statistics showed that removing the residency requirement would have only resulted in a small increase in the representation of African American candidates while decreasing the number of Hispanic candidates.¹⁷⁷ However, when the statistics were used to predict the eligibility of candidates when the residency requirement expanded to a ten-mile radius, the change was dramatic. There was a much higher inclusion of African-American candidates.¹⁷⁸

The residency requirements did not prove to be job-related and consistent with business necessity. Here, on the other hand, the Court has determined that the residency requirement does subject the NHRFR to disparate impact liability despite NHRFR's assertions to the contrary¹⁷⁹. Therefore, the NHRFR would not be expanding the residency requirement because of race-bases statistics alone; it would be expanding the list because the residency requirement causes a disparate impact that is not justified by business necessity.¹⁸⁰ The Court was looking for solid evidence that the reasons for the residency actually translated into successful job performance.¹⁸¹

The Court relied on the Court of Appeals by stating, “. . .in determining whether a plaintiff has presented a prima facie case of disparate impact through presentation of statistics, the plaintiff must prove not only statistical disparities in the employer's workforce, but also that the challenged practice causes that disparity.”¹⁸² The use of statistics in North Hudson is notable in that it is a distinct departure from the *Ricci* Court's avoidance of an examination of the statistical proof of disparate impact. The *Ricci* court, unlike the Court in *North Hudson*, minimizes the importance of statistical proof both to prove disparate impact and to also demonstrate the viability of alternative methods.

Although the court in *North Hudson* maintains that they are following the *Ricci* Court, the court is in actuality taking a more traditional approach in examining the evidence. There is heavy reliance on the appellate precedent which takes a much more traditional approach to examining issues of disparate impact.

CONCLUSION

Based on *North Hudson*, the appellate courts might be viewed as supporting the *Ricci* court in form, but not in substance. It seems to be the judiciary view that *Ricci* should be strictly

¹⁷⁶ *Id.*

¹⁷⁷ *Id.* at 9.

¹⁷⁸ *Id.* When the residency area was expanded to only a three mile radius, the increase representation of African Americans was negligible and the decrease in representation for Hispanic candidates was significant.

¹⁷⁹ *Id.* at 16. The City asserted that their residency requirements: (1) increases the likelihood that firefighters in the NHRFR will live in the area, which means they will be more familiar with the buildings and the streets, and also that they will be able to respond more quickly in the case of an emergency; (2) avoids liability to Intervenor and comports with the goals of the Rodriguez settlement agreement; (3) increases the number of Spanish-speaking NHRFR employees, which is helpful since the communities the NHRFR serves are 69.6 percent Hispanic; and (4) fosters pride in the community. The Court will address each of the NHRFR's asserted justifications in turn.

¹⁸⁰ *Id.* at 22.

¹⁸¹ *Id.* at 17-19.

¹⁸² *Id.* at 10. See *Newark Branch, NAACP v. City of Bayonne*, 134 F.3d 113, 121 (3d Cir.1998).

followed in very particular circumstances, but when it comes right down to it, the determination of whether a selection test has disparate impact still requires the kind of proof that is outlined in the Uniform Guidelines on Employee Selection Procedures. Until such time Congress elects to reaffirm the need to follow the Guidelines, it would seem that the courts will continue to rely on tried and true standards of proof with respect to disparate impact. Nonetheless, Congress should take quick action to make sure that the Ricci decision does not give rise to an abandonment of the Uniform Guidelines and the intent of Title VII – to voluntarily remedy tests that have disparate impact.

If there is a lesson to be learned, it is that employers need to be much more aware of what constitutes a good test under the Uniform Guidelines. The critical factor that was missing from all of the careful preparation of the IOS, the essential information was never able to be calculated – reliability and validity statistics. By calculating the statistics, the City would have had objective and solid evidence as to whether or not the tests were consistently measuring what they were intended to measure – the job relatedness of the test. Failing to do that, the Ricci court has essentially stopped any employer from taking remedial actions for tests that are discovered to have disparate impact. Surely the Court could not have meant to thwart all actions on the part of employers to remedy a discriminatory test if that action was even handed and in itself non-discriminatory. If so, the Court's decision is a major step backwards for those who wish to voluntarily comply in the elimination of tests with disparate impact.¹⁸³

¹⁸³ Arakawa, L.L and Sonen, M.P., *Caught In The Backdraft: The Implications of Ricci v. Destefano on Voluntary Compliance and Title VII*, 32 U. HAWAII L. REV., 463, 483 (2010).