Interview Questions – Testing and Selection Procedures in Light of Ricci

1. **Ricci v DeStefano** involved a promotional exam given to city firefighters – government employees – why should private employers be concerned about the Supreme Court’s holding in this case?

   Let’s start by giving some background on prior Supreme Court decisions that initially impacted the public sector and how that has affected private employers. Because state and local governments are public sector employers, their hiring, promotional, and contracting practices are subject to the strict scrutiny theory of law when it comes to affirmative action programs that use race-conscious remedies. The strict scrutiny criterion was codified in Richmond v. Croson (488 U.S. 469, 1989), a case that involved a city program that set aside 30% of city construction funds for black-owned firms that was challenged under the Equal Protection Clause of the Fourteenth Amendment. This set-aside quota was judged as a “highly suspect tool” by the Supreme Court. The Court asserted that such affirmative action steps must be subject to “strict scrutiny” and is unconstitutional unless racial discrimination can be proven to be “widespread throughout a particular industry.” The Court stated that “the purpose of strict scrutiny is to ‘smoke out’ illegitimate uses of race by assuring that the legislative body is pursuing a goal important enough to warrant use of a highly suspect tool. The [strict scrutiny] test also ensures that the [race-conscious] means chosen ‘fit’ this compelling goal so closely that there is little or no possibility that the motive for the classification was illegitimate racial prejudice or stereotype.”

   Both public and private cases that have followed in the wake of the Croson decision have proscribed that affirmative action programs that use race-conscious remedies can pass the “strict scrutiny” test only if they are “narrowly tailored” towards eradicating the effects of past discrimination and preventing current/future discrimination. “Narrowly tailoring” the remedies means that they must take into account factors such as the necessity of the program and the plausibility of alternative remedies, the scope and duration of the remedy, the relationship of the numerical goals to minorities within the relevant labor market, and the likely effect on innocent parties.

   Because they are not publically-funded, private employers may be held to a lesser standard than the strict scrutiny standard; however, similar concepts to those mentioned above will apply when they are challenged under the Equal Protection Clause of the Fourteenth Amendment. Further, the ruling in Ricci states that at least part of the Court’s intention was to “provide guidance to employers and courts for situations when these two prohibitions [disparate-impact liability and disparate-treatment discrimination] could be in conflict absent a rule to reconcile them. In providing this guidance our decision must be consistent with the important purpose of Title VII—that the workplace be an environment free of discrimination, where race is not a barrier to opportunity” (p. 30). This admonition appears to speak to both public and private employers alike.
2. What was the factual setting of the Ricci case in terms of the test that was administered and the results it yielded?

The context of the Ricci case involved the City of New Haven’s firefighter promotional practices. In Ricci, the City administered two promotional selection processes with the intention of selecting people for the positions of Fire Department Lieutenants and Captains. For each position, the final ranking of the candidates was based on a weighting of 60% on the written test and 40% on the oral interview.

The passing rates were lower for each minority group and minorities as a group compared to whites on the Lieutenant and Captain lists. An 80% rule of thumb test is often used by the courts, along with other statistical tests, to determine if there is adverse impact. The passing rate for whites on the Lieutenant list was 58.1%; therefore, the 80% limit of the passing rates of white candidates on the Lieutenant’s testing process was 46.5% (58.1% x 80%). The passing rates for blacks, Hispanics and total minorities were all lower than the 80% limit (31.6%, 20% and 26.5% respectively). These differences were statistically significant as well. The passing rates for whites on the Captain’s testing process was 64%, with the 80% limit for that process being 51.2%. The passing rate for blacks, Hispanics and total minorities was below this limit (37.5%, 37.5% and 37.5% respectively), though not quite statistically significant. Finally, when the Lieutenant and Captain lists were combined, the passing rate of white candidates was 60.3%, setting the 80% limit at 48.2%. The passing rate for blacks, Hispanics and total minorities for the combined lists were each lower than this limit (33.3%, 26.1%, and 30% respectively), and these differences were statistically significant.

3. What was problematic for the City about the results of the examination? What legal issues were implicated by the results? What potential exposure to liability was the City concerned about?

After giving the tests and learning that there was substantial adverse impact on minority candidates, the City threw out the results of that testing process. Their decision was based on the adverse impact finding as well as on limited information that the tests might not have survived a “possible” disparate impact challenge.

The City realized that giving a test that had adverse impact and throwing out the results on that basis alone could put place them in a situation where it faced two lawsuits: (1) one from whites who would have likely been appointed had the lists been adopted; and/or (2) one from minorities who would not have been likely to be appointed had the lists been adopted. A potential suit from whites would be a disparate treatment suit; the suit from minorities would be a disparate impact suit. The City attempted to avoid both potential suits by cancelling the lists. The Court ruled this action was impermissible: “Fear of litigation alone cannot justify the City’s reliance on race to the detriment of individuals who passed the examinations and qualified for promotions. Discarding the test results was impermissible under Title VII.” The Court further noted that a strong basis in evidence would be necessary for such drastic action to be allowed: “If, after it certifies the test results, the City faces a disparate-impact suit, then in light of today’s
holding the City can avoid disparate-impact liability based on the strong basis in evidence that, had it not certified the results, it would have been subject to disparate-treatment liability” (pp. 33–34).

4. What steps did the City of New Haven take when it learned of the results of the test? Why did the City decide not to certify the examination results?

   The City’s response was to convene an ad-hoc test review committee to testify before the Civil Service Board in an attempt to discredit the test’s validity (to justify their actions of tossing out the exam lists, which, according to the court testimony, was at least partially politically motivated). They also blocked the actual test validation report from the Court’s evidence record by specifically requesting that the test developer not send them the validation study. This was done to support the City’s efforts to disregard the list’s results and thereby allow the City to use a different selection device to create a new list. A validation study would have exposed the practices of most concern to the City: arbitrary (but negotiated) 70% cutoff score, arbitrary (but negotiated) weighting, use of rank-ordered list (possibly without the necessary validation support under federal testing Guidelines, Section 14C9), and “rule of three.” By not having a validation study “on record,” the City hoped to avoid litigation, but this also left both the City and the Court without the evidence necessary to conduct a professional job-relatedness review which may, in fact, have actually spared the City litigation.

5. Were there other alternatives available to the City when it learned of the test’s disparate impact on nonwhite candidates?

   Yes. If the City really wanted to cancel the lists, they would have had to establish a strong basis in evidence that the tests had not been properly validated. They would also have had to expose the use of the 70% cutoff without job relatedness support (typically cutoff scores are validated, this one was not); the 60% / 40% weighting scheme which was based upon negotiation with the union and was not supportable based upon job analysis data or subject matter expert opinions (as would be established through a validation effort); rank-ordered lists rather than banding substantially equally qualified candidates together based upon the reliability of the tests; and using the “rule of three” rather than a grouping of substantially equally qualified candidates being submitted to an appointing authority. Alternatively, to establish a strong basis in evidence, the City would have needed to provide the evidence of lack of job relatedness for these practices. This could have been done by City staff, the City attorney, or an outside expert. The City probably knew, however, that had they done this, the evidence gathered could have been devastating if discovered during a suit from minorities (even though such studies are typically conducted using attorney-client privilege).
6. What is the tension between Title VII of the Civil Rights Act of 1964 and the Equal Protection Clause of the Fourteenth Amendment, and why does it pose a challenge for employers?

The Court stated in Ricci that, “This suit does not call on us to consider whether the statutory constraints under Title VII must be parallel in all respects to those under the Constitution. That does not mean the constitutional authorities are irrelevant, however. Our cases discussing constitutional principles can provide helpful guidance in this statutory context” (p. 22). Further clarifying the legal framework of the case, they continue, “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” (p. 25). Drawing this fine line, the Court clarifies that Ricci was reviewed under Title VII, not the Fourteenth Amendment.

Both private and public employers are covered by Title VII. Private employers usually do not use written tests with oral interviews weighted with weights negotiated with a union, with 70% cutoff scores, rank-ordered lists and the “rule of three” to send candidates to the appointing authority. Public sector employers have constraints from civil service rules, merit systems, and union negotiations dealing with promotional tests not shared by private sector employers. The basic concepts and conclusions in Ricci, however, hold for both public and private employers that any disparate treatment race-conscious treatment contemplated in the name of avoiding disparate impact must also have a strong basis in evidence to support it.

7. What analytical framework was applied by the district court and the Second Circuit Court of Appeals in rejecting the reverse race bias claims presented by the 17 white, and one Hispanic, candidates who filed suit challenging the city’s refusal to certify the test results?

Ricci was tried initially in the lower courts as a Title VII disparate treatment case, relevant to its close context with disparate impact theory of law. The Second Circuit assigned three judges from its pool of nine active and twelve senior judges to hear the Ricci case. This three-judge panel affirmed the district court’s ruling in a summary order on February 15, 2008. Shortly after the summary order, a judge on the Second Circuit requested that all judges on the Second Circuit hear the case. The three-judge panel then withdrew its summary order and on June 9, 2008 issued a per curiam opinion (order of the three-judge panel rather than of named judges). A vote of the entire Second Circuit panel (some did not vote) was taken on June 12, 2008 regarding hearing the case as an entire panel. Six judges voted to hear the case again and seven voted not to hear it again (a close call, even at the appellate level). The Supreme Court reversed the Second Circuit’s ruling in this case.
8. As the High Court framed it, the question in *Ricci* was whether “the purpose to avoid disparate-impact liability excuses what otherwise would be prohibited disparate-treatment discrimination?” In answering the question, the Court adopted a new standard requiring that 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. How can employers meet this standard?

Employers can meet the “strong-basis-in-evidence” standard by conducting a Croson Study (or similar process by another name) for the at-issue position. A Croson Study includes an evaluation of the utilization of the at-issue group (by comparing their makeup in the workforce to the qualified applicant pool or census data) as well as the adverse impact and job relatedness of past practices, procedures and tests used to select candidates during prior selection processes and consideration of alternate employment practices. The job-relatedness evaluation includes the assistance of subject-matter experts, who also can provide information for alternative selection procedures (e.g., job related work sample tests or assessment centers). Alternate selection procedures can also be identified using literature searches and exploring the tests that have been successfully used by other similar employers. The data and information generated from a Croson Study can provide the foundation for a strong basis in evidence of prior disparate-impact discrimination and pin-point specific problems, identifying practices, procedures and tests with suggested options.

Next, the employer would need to identify the specific situation they are now facing that may require a race-conscious remedy. For example, if a written test has been given that has severe adverse impact against minorities, the employer can submit this test to a Croson Review. The Croson Review would evaluate the adverse impact of the test and the job relatedness of the test. If the Croson Review demonstrates that the written test indeed adversely impacts minorities, and does not have sufficient job relatedness to withstand a Title VII challenge, the employer could be justified in not using the test results and administering another (properly validated) test. It should not go without saying, however, that throwing out exam results is a drastic measure that has a tremendously negative impact on the employer, especially on those individuals whose employment opportunities have been retracted.

9. In the view of the majority of the Justices, why did the City fail to meet this standard?

The City provided arguments that they did not successfully buttress with convincing evidence, data, and validation reports. The City also allowed their testing consultant to testify in support of the written test and oral interview and to hide (not ask for) the formal validation report (which was scheduled to be submitted) which may have highlighted other practices, procedures, and tests that were not the testing consultant’s responsibility and could not be supported by validation evidence, specifically: (a) 70% cutoff score (an arbitrary percentage based in tradition without competency-based
supporting evidence from subject-matter experts), (b) weights negotiated for written test (absolutely no attempt to support with subject-matter experts in any validation effort), (c) rank-ordered use of the lists (not at all supported by the reliability of the tests, in fact just the opposite, possibly no differentiating validation evidence to address the Federal Guidelines presented, not transformed to standard scores and had more weight unnecessarily creating more adverse impact on minorities than the 60% arbitrarily negotiated), and (d) the use of the “rule of three” (possibly no differentiating validation support, and absolutely no reliability support to show that substantially equally qualified candidates would be presented in groups of three to the appointing authority–more adverse impact built into the system, but hidden from the courts as a result of hiding/suppressing the actual validation study).

Based on the very limited evidence record developed during the evolution of the case, the City was not able to meet the Court’s strong basis in evidence criteria: “Especially when it is noted that the strong-basis-in-evidence standard applies to this case, respondents cannot create a genuine issue of fact based on a few stray (and contradictory) statements in the record” (p. 4, referencing pp. 29-33). Because this unique case created a situation where neither party wanted the validity of the test scrutinized in court, the actual validity of the test was left out of the record (while the painstaking process to develop the test was described during a Civil Service Board hearing by the test’s developer, neither the actual validity study nor any independent review that may have been conducted was included in the evidence). This distinction obviously played a role in the Supreme Court’s deliberation, where they relied almost wholly on an evidence record that consisted of ad hoc testimony from three professionals at a Civil Service Board hearing (only one of whom actually reviewed the tests).

Finally, the only testimony on the record regarding the suitability of assessment centers as a low adverse impact alternative came from Ricci (the plaintiff) himself, and he argued that it would take years to develop: “Firefighter Frank Ricci again argued for certification; he stated that although ‘assessment centers in some cases show less adverse impact,’ id., at A1140, 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 materials” (p. 14).

In summary, to win the case, the City would have had to convince the Court that it was actually under a Title VII disparate impact threat, and would have been very likely to lose the case, either by lack of validity or failure to use suitable alternate employment practices that would have resulted in less adverse impact.

10. Was Ricci a typical Title VII adverse impact testing and selection case? Why or why not?

Ricci was a disparate treatment case (against whites) that included selection procedures that had disparate impact (against minorities)—an important but subtle distinction that makes it different from a typical Title VII case. Facing potential challenges from both sides, the City relied on legal strategies instead of on proven
business practices and procedures to defend their actions (e.g., not asking for the validation study that was already part of the contract the City had with its test developer). Knowing the plaintiffs, who were white, would not want to challenge the very practices that got them to the top of the lists (which included an arbitrary 70% cutoff score and weights, rank-ordered lists, and the “rule of three”), the City suppressed the validation evidence from the Court’s evidence record which could have possibly revealed that their testing practices where not valid (an arbitrary 70% cutoff score and weights; rank-ordered lists; and the “rule of three”). Trying to avoid a challenge by the minority test-takers, the City focused its arguments instead against the selection processes, but did not successfully carry this burden.

In a typical disparate impact case, the defendant (in this case, the City) would need to prove that their test was “job related for the position in question and consistent with business necessity.” For the City to prevail in this case, they had to prove the absence of validity to justify their actions of tossing out the exam results. This is a lesser standard than having to prove validity. In this case, neither party wanted the validity of the test fully vetted so the existing validation study was suppressed/hidden/not admitted as evidence.

11. What was the evidence of record that the Supreme Court concluded was insufficient to meet the “strong basis in evidence” standard?

The Court essentially ruled that the City of New Haven’s race-based rejection of the test results only because of adverse impact against minorities did not satisfy the strong-basis-in-evidence standard (Ricci, pp. 26–34). While there was no dispute that the City was faced with a prima facie case of disparate-impact liability if they had used the lists that had been created, this information alone was not enough to cancel the lists. The problem for the City is that such a prima facie case—essentially, a threshold showing of a significant statistical disparity of test scores and nothing more—is far from a strong basis in evidence that the City would have been held liable under Title VII had it certified the test results.

The court ruled that the City could be liable for disparate impact discrimination only if the exams at issue were definitively not job related and consistent with business necessity, or if there existed an equally valid, less discriminatory alternative that served the City’s needs but that the City refused to adopt. Based on the very limited record the parties developed through discovery (due to the City suppressing the validity study and by the clever legal strategy that allowed the validity study to evade a professional independent review), there was “… no substantial basis in evidence that the test was deficient in either respect” (Ricci, pp. 26–28).

Even so, the City made claims that the exams were not job related and consistent with business necessity, while failing to produce any validation study supporting their claim. A validation study would have looked at the entire selection process and could have shown the parts that were valid and the parts that were not, if any. Of particular interest is the fact that the City attorney asked the test developers to not provide the
validation study already slated for delivery in the contract the test developer had with the City (Ricci, p. 6). This, despite the fact that several people had asked the City’s Civil Service Board to have a validation study conducted so that it could be determined if the test process was job related and consistent with business necessity. As a result of a validation study not being presented, the only evidence available to the Court demonstrated that detailed steps had been taken to fairly develop and administer the tests and that there were painstaking analyses of the questions asked to make certain those questions were relevant to the Captain and Lieutenant positions. Specifically, the testimony showed that complaints by some of the test takers about some examination questions in regards to firefighting as it is performed in the City were fully addressed. Testimony also revealed that the City turned a blind eye to evidence supporting the exams’ validity (Ricci, pp. 28–29). In other words, the City did not provide a strong basis in evidence that the exams were not job related to support its decision to toss out the test results.

During the litigation process, the City made a few vain attempts to show available alternative selection procedures that might have resulted in less adverse impact against test takers. In order for the City to succeed, however, a strong basis in evidence was needed to support each of the City’s proposed alternate employment practices and this was not offered. For instance, the City asked for an opinion from a testing expert who was not involved in the selection process, but who also provided advice to the City without reviewing the selection procedures at issue. Additionally, the City referenced testimony that a different composite-score calculation would have allowed the City to consider black candidates for the open positions while not proving that the weighting actually used was indeed arbitrary, or that different weighting would have been an substantially equally valid way to determine whether candidates were qualified for promotions.

Furthermore, the City argued that it could have adopted a different interpretation of its charter provision limiting promotions to the highest scoring applicants, and that the interpretation would have produced less discriminatory results. The Court said this approach would have violated Title VII’s prohibition of race-based adjustment of test results had it been done after the exams were administered for the exclusive purpose of reducing adverse impact. It should be noted that score banding has been widely supported in litigation in both federal district and appellate level courts and is a commonly applied in personnel testing because it groups applicants in the tightest groups possible given the reliability (consistency) of the tests involved. Banding is an allowable process because it is a psychometrically-driven (and not race-based) process, which is why it is an acceptable process (especially if decided before the test is administered).

Finally, the City asserted that the use of an assessment center to evaluate candidates’ behavior in simulations of typical job tasks would have resulted in less adverse impact than written exams. This assertion, however, was contradicted by other statements the City made on record indicating that it could not have used assessment centers for the exams at issue, because such testing would have taken too long to develop. In summary, the City provided no strong basis in evidence that any alternate employment
practice was available (Ricci, pp. 29–33) and, in fact, contradicted itself on at least one occasion.

12. What are the implications for employers of the High Court’s analysis of the test-related evidence that was available in the record?

This was a unique case with unusual circumstances that are not typically encountered in EEO settings (i.e., tossing out a list solely on the basis that it had adverse impact). The classical Title VII burdens established under the USSC’s unanimous ruling in the Griggs v. Duke Power Company (401 U.S. 424, 1971) case and codified by the 1991 Civil Rights Act remain intact. In the aftermath of Ricci, some recent lower-court cases clarified this point. For example, in the post-Ricci case, U.S. v. City of New York (07-cv-2067, NGG, RLM), the court clarified that the Ricci case does not change the law as it relates to the burden-shifting requirements outlined in Title VII. While the major tenants of Title VII remain, Ricci does provide specific guidance to employers regarding making race-based decisions after an employment test has been given—i.e., let the results stand unless they can prove (using the strong basis in evidence standard) that the test was definitely not valid or that other substantially equally valid (lower adverse impact) options were available but were overlooked.

13. What type of evidence could the City have offered in order to meet the standard applied by the Court?

To win the case, the City would have had to prove that (1) they were definitely going to be sued under Title VII’s disparate impact theory of law, and (2) they would most likely have lost the case either because the tests were definitely not valid or there were substantially equally valid alternatives available.

If they could have proved they were going to be sued, demonstrating a lack of job relatedness (validity) could have been shown by demonstrating one or more fatal flaws, which could have included any one of the following:

1) Lack of validation support for the cutoff score of 70% with adverse impact against minorities (set by tradition and not as a result of a job relatedness competency review by subject-matter experts);

2) Lack of validation support for the 60% / 40% weights used for the written and oral interview tests. Weights that are set by union negotiation are not necessarily “job related” (i.e., valid). Weights should be set by either a job analysis process (whereby the competencies necessary for the job are comparatively evaluated) or using the input from a panel of qualified subject-matter experts who evaluate and rate the tests given the competencies to which they are linked.

In addition, when tests are weighted and combined, they should be standard scored so they can correctly be combined into a composite score. This was not done in the Ricci case, which resulted in the actual weights being different than
the intended weights. Specifically, the actual weights for the Lieutenant’s Oral were 37.2% and 62.8% for the written, and 38.2% for the Captains Oral and 61.8% for the Written. Though slight, these differences resulted in the creation of a final score list where 56% of the 77 Lieutenants were out of their correct rank order and 37% of the 41 Captains.

3) Lack of validation support for the rank-ordered use of the selection process which increased the adverse impact against minorities artificially (set by civil service and not supported by the reliability of the tests involved in the selection process and possibly not supported by differentiating validity evidence required by the Guidelines, Section 14C9); and

4) Lack of validation support for the “rule of three,” set by civil service rules rather than by the reliability of the tests involved in the selection process.

The City could also have prevailed if they could have carried the burden under the Alternate Employment Practice theory of disparate impact law. To do this, the City would have had to successfully argue (by presenting evidence, not just argumentation) regarding any one of the following:

1) That an alternative weighting system would have less adverse impact than the 60% / 40% weighting system it used, but was substantially equally valid. However, the City failed to present any evidence that their alternative weighting system was substantially equally valid to the 60/40 (a basic requirement of alternate employment practices under Title VII). The City would have had to use job analysis data or convene a subject-matter expert panel to evaluate the appropriate job-related weights for each test. The City could have also criticized the weighting system it used (applying weights to the raw scores rather than the standard scores) to demonstrate that the adverse impact was needlessly exacerbated.

2) That it could have adopted a different interpretation of the “rule of three” that would have produced less adverse impact and focused their argument on the race-based issue of lowering adverse impact, rather than on the psychometric fact that the “rule of three” artificially groups candidates in a way that the tests used to generate the list cannot support. The written test and oral interview test are not perfect tests. They each have error in their measurements. The error of measurement is quantified with a reliability coefficient. Using the reliability and related statistics (the conditional standard error of measurement and standard error of difference), the scores could be banded in a way that would group substantially equally qualified candidates together for consideration by the appointing authority rather than the “rule of three” which is not supported by the psychometric properties of the written and oral tests. The City did not present any evidence on the psychometric properties of the written and oral tests.
3) That an assessment center (or other testing practice) could have been used with less adverse impact than the written tests. The City provided no evidence that assessment centers measured the same factors measured by the written test, have had lesser adverse impact elsewhere, or were substantially equally valid.

14. Specifically, what steps could the City have taken, when it learned that its test resulted in a disparate impact on nonwhites, in order to justify under the “strong basis in evidence standard” its refusal to certify the examination?

They could have conducted (or sourced) an independent validation review and fully substantiated the validation support (or lack thereof) for each of the four critical validity elements discussed above (the 70% cutoff score, weights used, rank ordering, and the rule of three). Evaluating the validity of each of these factors requires addressing (at a minimum) Sections 14C and 15C of the federal Uniform Guidelines, as well as professional standards.

To support the alternate employment practices presented, the City should have presented the argument that an alternative weighting system for the written test and oral interview would have less adverse impact than the 60% / 40% weighting system it used but support the argument with evidence that the new weighting system was substantially equally valid to the weights used. The City would have had to use job analysis data or convene a subject-matter expert panel to obtain such weighting opinions. The City could have also criticized the weighting system it used (applying weights to the raw scores rather than the standard scores) to demonstrate that the adverse impact was needlessly exacerbated using raw scores and not standard scores.

Regarding the “rule of three,” they would have had to successfully present the argument that it could have adopted a different interpretation of the “rule of three” that would have produced less adverse impact and focused their argument not on the race-based issue of lowering adverse impact, rather on the psychometric fact that the “rule of three” artificially groups candidates in a way that the tests used to generate the list cannot support (based on the reliability of each test and their weighted composite).

With respect to their argumentation regarding the assessment center as a substantially equally valid alternative that could have been used with less adverse impact than the written tests, they would have had to demonstrate that the assessment center was in fact a substantially equally valid alternative, along with data to show assessment centers have had lesser adverse impact elsewhere.

In hindsight, the City probably should have entered into evidence the full and complete validation study (assuming they were able to uncover fatal flaws in the test development process). However, as mentioned previously, because neither party wanted the validation study fully vetted for strategic reasons, the City wasn’t able to show there was a strong basis in evidence that the test would not have survived legal and professional scrutiny.
15. How could the City have ensured in the first place that its test would meet the standard applied by the court?

By asking its testing consultant to review all practices, procedures, and tests used in the selection process and not just limit itself to the written test and oral interview. The testing consultant did not address the City’s 70% cutoff, the City’s 60% weight on the written test, the rank-order use of the list and the City’s use of the rule of three. If the City wanted to substantiate the validity of their tests, they could have addressed the federal Uniform Guidelines as well as professional standards. Several of these standards were likely addressed by the test development process, however, the validity study was not admitted in the evidence record, deliberated, or professionally reviewed by an independent party. Alternative employment practices with less adverse impact could have been explored and evaluated.

a. How can an employer ensure that a test or selection procedure is a valid and reliable instrument for the purposes intended?

At a minimum, the Uniform Guidelines should be addressed. Specifically, Sections 14C and 15C were relevant to this case because a content validity strategy was employed. Professional standards such as the APA Standards and the SIOP Principles also provide guidance to test development professionals for developing effective and defensible tests. If the employer uses a consultant, as the City did in this situation, the consultant could be asked to evaluate all practices, procedures and tests used—not limit itself to just a written test and oral.

b. Once an employer determines that a test is a valid and reliable selection instrument for a particular job, can the employer presume that the test is also valid and reliable for other purposes, and why or why not?

The 1991 Civil Rights Act specifies that a test is defensible (even if it has adverse impact) if it is “job related for the position in question and consistent with business necessity.” Moving test validity between different positions is permissible if the knowledges, skills, and abilities measured by the test are critically important on the first day of employment (under content validity) for each of the target positions. Of course, a validity study would be required to determine whether this was actually true. Under criterion-related validity (where statistical evidence is the primary basis for validity evidence), tests can be used across positions if the validity can be transported by showing that the positions require employees to perform the same or very similar important work behaviors (as well as the other requirements outlined in Section 7B of the Uniform Guidelines).
c. To what extent should employers use “pre-validated” tests? What, if any, are the pitfalls to using “pre-validated” tests?

Because test validity deals with the interpretation of a certain score for a certain position, there is no such thing as an “automatically valid” or a “pre-validated test” (i.e., one that might be broadly valid across most/all positions), but rather only tests that are valid for a given position, and at a given level for that position. This does not, of course, prohibit the same test being used across multiple positions provided that the competencies (knowledges, skills, and abilities) necessary on the first day of employment are sufficiently similar. In order for a testing process that was found to possess criterion-related validity at another location, the employer would have to satisfy the applicable portions of Section 7B of the federal Uniform Guidelines if the tests exhibited adverse impact. This section requires, among other things, a study to determine the similarity between the job on which the test was originally validated and the job for which the test will be used. Failing to conduct such a study would be equivalent to using an invalid test in the eyes of the courts. If the testing process was found to be valid using a content-related validity approach, a validation study of the test as it would be used for the new position or location would need to be conducted to address Sections 14C and 15C of the Uniform Guidelines. Such studies would help to determine the appropriateness of the use of the test for a new position and also identify if the minimum level of proficiency (e.g., cutoff score) at the new job or location would be the same/similar as that which had been identified for the original job. In this way, the use of the test would be validated for a specific position in a specific setting. For example, a keyboarding speed test may be valid across an entire administrative job group including dozens of job titles, but the actual words-per-minute requirement may differ between job titles.

d. Has the Ricci decision changed what constitutes a valid and reliable test or selection procedure in the view of the courts?

In short, no. Had the Court remanded the case back to the Second Circuit so that specific issues pertaining to the validity of the test could have been deliberated after going through a judicial and professional review process, then certain validity issues could have been changed by the Court. Because the Ricci case was a disparate treatment case that involved a test, the specific elements that are required for making the demonstration of validity as required under the 1991 Civil Rights Act were not changed.

16. Does the Ricci decision impact the relationship between disparate impact and disparate treatment discrimination under Title VII, and if so, how?

No. Because this case was a disparate treatment case that involved a test, no changes were made in this area of the Title VII framework. As the Court stated, “Our holding today clarifies how Title VII applies to resolve competing expectations under the disparate treatment and disparate-impact provisions. If, after it certifies the test results, the City faces a disparate-impact suit, then in light of our holding today it should be clear that the City would avoid disparate-impact liability based on the strong basis in evidence
that, had it not certified the results, it would have been subject to disparate treatment liability” (Ricci, p. 34).

17. What are the practical implications of the Ricci decision for employers and how broadly do they apply?

   a. Does Ricci apply to a private employer that performs a disparate impact analysis prior to implementing a reduction in force, and if so, how?

   b. Does the new standard adopted in Ricci apply to all testing and selection procedures from hiring through firing in both private and public employment? If not, what in your view are the outer limits of the application of this new standard?

The primary practical implication from Ricci is that if employers are using practices, procedures, and/or tests that are fair and valid they are on safe ground for using those procedures. Furthermore, employers should let the results of an employment-related test stand unless they have a strong basis in evidence for disregarding the results from the test (based on the factors discussed above). In addition, any steps taken to reduce adverse impact (that are outside the bounds permitted within Title VII’s “substantially equally valid” doctrine—should be made prior to the test being administered, unless there is a strong basis in evidence that race-conscious remedies are necessary (as would be defined through a Croson Study process).

When employers go through a Reduction in Force (RIF) process, they should be sure that the criteria used for making RIF decisions are based upon a bona fide seniority system and/or are job related and uniformly applied. If the RIF process results in adverse impact based on race or gender and a bona fide seniority system is not being used, the same burdens that are relevant to a hiring case will apply—i.e., the employer will need to justify their RIF criteria based on the 1991 Civil Rights Act validation standard anchored to the Griggs case (“job related for the position in question and consistent with business necessity”), which is essentially “making a demonstration of validity.” Under this same framework, a plaintiff under such a case could prevail under the “alternate employment practice” theory (e.g., by successfully arguing RIF practices or criteria that were “substantially equally valid,” but had less adverse impact).

If the RIF process is not based on a bona fide seniority system and has adverse impact based on age (which is common in RIF cases), the USSC has ruled in Smith v. City of Jackson (03-1160, 544 U.S. 228, 2005, 351 F.3d 183, affirmed) that there are important textual differences between the ADEA and Title VII. The ADEA permits any “otherwise not prohibited” action “where the differentiation is based on reasonable factors other than age.” The result is that the ADEA’s requirement for justifying adverse impact based on age is much less than the Griggs requirement. An employer does not need to “demonstrate validity” but rather only needs to show that reasonable factors other than age accounted for the adverse impact.
So, under Ricci, can an employer calculate adverse impact before making RIF decisions, and even base individual RIF decisions on whether such decisions put them on the adverse impact radar? If their RIF process is defensible (with either a bona fide seniority system or job related validity and uniformly applied) the adverse impact is justified and the employer is safe if challenged under either a race/gender or an age-related claim of adverse impact. However, if their RIF process has adverse impact and is not based on a bona fide seniority system or un-validated criteria, they leave themselves wide open for a Title VII lawsuit.

Another limitation when trying to apply the Ricci concepts to RIFs is that there are major differences between these two practices when it comes to adverse impact theory. For example, Ricci dealt with a promotional process where all of the candidates knew the selection process and criteria in advance. It was also a public process, with a public sector employer. Further, in Ricci, they ruled that the evidence necessary to retract promotional opportunities from candidates based on their higher earned ranking on an (apparently) valid selection process required a strong basis in evidence that the tools used for such ranking were clearly not valid. RIFs are obviously different in most of these respects.

18. What steps should an employer take when it discovers that a test or selection procedure has resulted in a disparate impact against a protected group? What about when the disparate impact is against a non-protected group?

Adverse impact that occurs on a test that has not been “validated in accordance with the Guidelines” typically equates to “disparate impact discrimination” under the 1991 Civil Rights Act. Therefore, to avoid liability and comply with the Civil Rights Act, employers that identify practices, procedures, or tests that have adverse impact are obligated to take one of three choices: (1) conduct a validation study to determine the fairness and job-relatedness of the selection procedure, (2) use the selection procedure in a way that does not generate adverse impact (not a desirable alternative if the tests used in the process were, in fact, valid as it was originally intended to be used), or (3) discontinue the practices in the future (also not recommended if the tests were, in fact, valid since this would be ignoring a job-related tool that is likely to help the employer choose qualified people for the job).

Contrary to the myth of there being “protected groups” and “unprotected groups,” Title VII applies to each and every subgroup—i.e., members of both majority and minority races and colors, as well as applying to majority and minority national origin, sex, and religion subgroups. For example, if adverse impact occurs against whites, the same adverse impact/validity burdens apply as when adverse impact occurs against minority groups. In more recent years, we have noticed that federal enforcement agencies (e.g., the OFCCP) have increased enforcement activities in such cases. Adverse impact is defined by using the comparison group with the highest rate, whatever their race, and the validation requirement follows to justify the test causing the impact.

Copyright 2009, Biddle Consulting Group, Inc.
All Rights Reserved
19. What best practices and policies related to tests and selection procedures do you recommend as a result of the Ricci decision?

My primary recommendation is that employers should evaluate practices, procedures, and tests to identify which have adverse impact and, for those with adverse impact, conduct a validity review and evaluate alternatives that may be substantially equally valid with less adverse impact.

Sometimes employers are apprehensive about conducting validation studies. Objections such as cost or time delays are sometimes used to avoid taking validation steps. This is unfortunate because conducting validation studies provides two major benefits. First, validation helps insure that the selection process is selecting qualified applicants for the target position. The second benefit is that it generates validity evidence that can be useful in defending against Title VII lawsuits. Validation was used as a process years before Title VII required it when the practice, procedure, or test caused adverse impact. Its purpose was to find better ways of selecting people.

Another recommendation is that employers should consider conducting Croson Studies to evaluate whether adverse impact can be reduced in neutral ways, or if race-conscious remedies may be justified. If a properly validated testing practice is exhibiting adverse impact it may be entirely justified. Even so, a Croson Study investigation may uncover important (and measureable) traits necessary for success that are currently not included in the selection process, or other lesser discriminatory alternatives that may exist.

As for policy recommendations, the results of an employment test should never be scratched unless there is a strong basis in evidence that the test was not valid, or that other lesser discriminatory selection procedures were available.

20. Are there any other factors that employers should keep in mind as a result of the Ricci decision?

None to add beyond those discussed above.