

EEO

INSiGHT®



Volume 2 : Issue 1

*The Leading Resource for the
EEO/AA Community*



EEO INSIGHT®

The Leading Resource for the EEO/AA Community

Volume 2: Issue 1

EEO Insight® is published quarterly by Biddle Consulting Group, Inc., 193 Blue Ravine Road, Suite 270, Folsom, California 95630-4760. Electronic subscriptions (PDF e-mail delivery) are free for a limited time. Sign up online at www.EEOInsight.com/subscribe.html. Annual hardbound editions containing a compilation of published articles will be printed and made available for purchase. For more information on submitting an article for publication in EEO Insight, please contact Christine Anthony, Managing Editor, at canthony@eeoinsight.com or (916) 294-4250 ext. 108.

© 2010 Biddle Consulting Group, Inc. All Rights Reserved.

ISSN: 1947-1548

Dear READERS

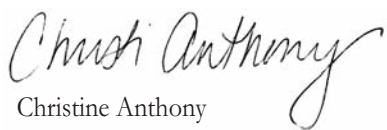
AS WE WORK through a new year and close out the first complete volume of EEO Insight, we would like to thank the readers, authors, reviewers, and editorial board members for all of your support. Together, we have accomplished our mission to equip and inform the equal employment and affirmative action community. We would like to continue to build and expand on our commitment to forge ahead and promote diversity, fairness, and equity in our workplaces and communities. Over the past several months, the Obama administration has confirmed its top officials and they, too, have the drive to improve equal employment in our country.

On February 24, 2009, Hilda L. Solis was confirmed as Secretary of Labor after being recommended by President Obama on January 20, 2009. Prior to confirmation as Secretary of Labor, Secretary Solis represented the 32nd Congressional District in California, a position she held from 2001–2009. During those years, Solis was dedicated to improving the lives of working families, protecting the environment, and expanding access to affordable health care. A recognized leader on clean energy jobs, Solis authored the Green Jobs Act which provided funding for “green” job training for veterans, displaced workers, at risk youth, and individuals in families under 200 percent of the federal poverty line. The Department of Labor has implemented the enforcement of the American Recovery and Reinvestment Act (ARRA) to protect workers who have lost their jobs, to provide new training opportunities for workers looking to upgrade their job skills, and to create new job opportunities in emerging sectors such as clean energy and health information technology. For more information regarding ARRA please visit, <http://www.recovery.gov>.

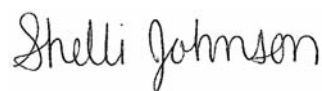
Patricia A. Shiu assumed the position of Director of the Office of Federal Contract Compliance Program (OFCCP) in October of 2009. Shiu brings to her new position a passion for workplaces free of race and gender discrimination as well as equality regarding family and medical leave. In 1983, Shiu joined the Legal Aid Society–Employment Law Center in San Francisco, California as an attorney and has most recently served as their Vice President of Programs. The OFCCP budget for the upcoming year has permitted Shiu to authorize the hiring and training of over 200 new Compliance Officers. Together with the increase in OFCCP staff and her passion for equal employment, we are sure to see an increase in the number and scope of compliance reviews under this administration.

This issue of EEO Insight will provide you with articles covering topics that the Administration has focused on over the past several months. Selection decisions have been and will continue to be heavily investigated during compliance reviews. This issue offers guidance on how to manage selection decisions and avoid adverse impact findings (please see *Managing the Impact Ratio Analysis: Strategies for Framing and Defending the Inference of Adverse Impact in Selection Decisions* and *Strategies to Avoid Adverse Impact When Choosing a Test for Selection or Promotion*). We also investigate compensation matters at a deeper level (please see *Digging Below the Surface: A Closer Look at Multiple Regression Analyses*). Finally, we offer to you a look at the history of the strong-basis-in-evidence standard that became the center of the *Ricci v. DeStefano* case and how the standard impacts your day-to-day employment decisions (please see *How Croson Studies can be used to Effectively Develop Defensible Diversity Initiatives in a Post-Ricci World*).

We invite our readers to reach out to us (editor@eeoinsight.com) with suggestions on how we can strengthen and improve EEO Insight. Authors are welcome to submit their work for consideration of publication in future EEO Insight issues.



Christine Anthony



Shelli Johnson

Editorial Directors

Dan A. Biddle, Ph.D.

Patrick M. Nooren, Ph.D.

Jim Higgins, Ed.D.

Managing Editor

Christine Anthony

Assistant Editor

Shelli Johnson

Production Editor

C. Lori Lee

Contributor

Chris Lindholm

Production Artist and Cover Design

Susan M. Sheldrake

Review Board

Margaret S. Carter, *Raytheon Space and Airborne Systems*

George Desloge, *ERS Group*

John C. Fox, Esq., *Fox, Wang & Morgan P.C.*

John Garza, *JAG Specialties, LLC*

Lisa Harpe, Ph.D., *Peopleclick Research Institute (PRI)*

Valerie Hoffman, *Seyfarth Shaw LLP*

Greg M. Hurtz, Ph.D., *California State University, Sacramento*

Brent Kleiman

James E. Kuthy, Ph.D., *Biddle Consulting Group, Inc.*

Van M. Latham, *PathPoint Consulting Incorporated*

Randy Mizelle, *Progress Energy*

Patrick Nooren, Ph.D., *Biddle Consulting Group, Inc.*

Chester I. Palmer, Ed.D., *ERS Group*

Michael Reagan, *CIGNA Corporation*

Kurt Ronn, *HRworks*

Olivia Sanchez, *Metropolitan Water District of Southern California*

David Scheffler, *Pinnacle Affirmative Action Services, LLC*

Jennifer L. Seda, Esq., *Jackson Lewis LLP*

Cathy Seward, *SONY Electronics, Inc.*

Mickey Silberman, Esq., *Jackson Lewis LLP*

Kay Todd, *Paul, Hastings, Janofsky & Walker LLP*

William H. Truesdell, *The Management Advantage, Inc.*

Kristi Williamson, *McKesson Corporation*

The opinions expressed are those of the authors. They do not necessarily represent the views of the publisher, editorial board or peer reviewers involved in the article review process. The EEO Insight editorial board realizes that practitioners in the HR/EEO field often hold widely divergent views and opinions on the various subjects that are included in this journal. EEO Insight is designed to provide a venue for professional dialogue to occur on these topics. As such, we invite practitioners to submit articles that contain differing viewpoints to those expressed in this journal forum.

TABLE *of* CONTENTS

<i>Introduction</i>iii
<i>Contributors</i>	iv
Managing the Impact Ratio Analysis: Strategies for Framing and Defending the Inference of Adverse Impact in Selection Decisions <i>Kyle C. Goodridge, SPHR, Citigroup, Inc.</i>	1
Digging Below the Surface: A Closer Look at Multiple Regression Analyses <i>Desir�e Throckmorton, Biddle Consulting Group, Inc.</i>	11
Strategies to Avoid Adverse Impact When Choosing a Test for Selection or Promotion <i>Meghan R. Lowery, Southern Illinois University Carbondale and Matthew J. Borneman, University of Minnesota</i>	21
How Croson Studies can be used to Effectively Develop Defensible Diversity Initiatives in a Post-Ricci World <i>Dan A. Biddle, Ph.D., Biddle Consulting Group, Inc. and Richard E. Biddle, Founder, Biddle & Associates, Inc.</i>	27
Additional Resources	61
Calendar of Events	63

Managing the Impact Ratio Analysis:
**STRATEGIES FOR FRAMING AND DEFENDING
THE INFERENCE OF ADVERSE IMPACT
IN SELECTION DECISIONS**

Kyle C. Goodridge, SPHR | *Citigroup, Inc.*

The Office of Federal Contract Compliance Programs' (OFCCP) increasingly successful enforcement of the nondiscrimination provisions of Executive Order 11246 requires increased vigilance in the contractor community as it pertains to the inference of adverse impact in selection decisions. Moreover, OFCCP's re-emphasis on affirmative action means that contractors must possess the tools necessary to defend against the adverse impact indicator and assess the interaction between nondiscrimination and affirmative action. This article discusses some of the methods equal employment and affirmative action professionals can use to build a framework for analyzing adverse impact within the context of affirmative action.

INTRODUCTION

THE PAST several years have seen the Office of Federal Contract Compliance Programs (OFCCP) working to “. . .concentrate Agency resources on identifying and remedying cases of systemic discrimination, thereby enabling the Agency to use its resources in a more effective and efficient manner” (U.S. Dept. of Labor, Employment Standards Administration, 2008). With this approach, OFCCP has refined its strategy for evaluating contractor compliance with Executive Order 11246 (E.O. 11246), Section 503 of the Rehabilitation Act of 1973, and the Vietnam Era Veterans Readjustment Assistance Act (VEVRAA). Using the Active Case Management Directive (U.S. Dept. of Labor, Employment Standards Administration, 2008), OFCCP shifted its attention away from affirmative action compliance (*e.g.* technical requirements of the affirmative action plan, good faith

efforts, goal attainment) towards identifying contractors whose employment practices have a negative impact on job applicants and employees. Along with the Equal Employment Opportunity Commission (EEOC), OFCCP's efforts at identifying systemic discrimination¹ have resulted in the recovery of over \$67.5 million in back pay, salary, and benefits for 24,508 workers in fiscal year (FY) 2008. Remarkably, 99% of the damages recovered were collected in cases of systemic discrimination (U.S. Dept of Labor, 2009). These results represented a 133% increase in the financial remedies recovered by OFCCP since FY 2001 and clearly demonstrate that OFCCP has “stepped up its game” in terms of its enforcement practices. It should be noted that the reporting strategies have significantly changed from FY 2008 to FY 2009 resulting in a dramatically lower FY 2009 financial remedy value. It is expected that the strategies used for FY 2009 will

continue to be the reporting method used under the current Administration.

Although OFCCP does not publish detailed data analyzing its enforcement results, research of FY 2007 enforcement statistics by Cohen and Dunleavy (2009) found that the vast majority (90%) of OFCCP audits resulting in financial remedies were the result of hiring discrimination. While the authors concede that the FY 2007 findings were the culmination of audits that may have begun prior to the advent of OFCCP's systemic discrimination initiative, these results, coupled with OFCCP's Active Case Management and Systemic Discrimination efforts, suggest that hiring discrimination will continue to be an important area of focus in the future.

In light of OFCCP's increasingly successful approach to the investigation of hiring practices, this article will discuss the primary method used by OFCCP to bring a charge of discrimination—adverse impact—and will propose strategies contractors may use to position its impact ratio analyses (IRA) towards minimizing or eliminating the inference of adverse impact that may be presented during the course of a compliance evaluation.

Because OFCCP's focus has been on hiring activity, this article will direct most of its attention on managing the inference of adverse impact in hiring. However, this article will also briefly discuss strategies contractors may use in managing adverse impact in promotions and terminations. In closing, this article will look at where OFCCP is today with its shift back toward affirmative action compliance and what contractors might expect as a result of this change, particularly in light of Active Case Management and the Systemic Discrimination Initiative.

ADVERSE IMPACT: DEFINITIONS AND CONTEXT

Although the reader is undoubtedly aware of what adverse impact is, it is worthwhile to provide a definition here to ground our discussion on what can be done to manage it.

*Adverse impact. A substantially different rate of selection in hiring, promotion, or other employment decision which works to the disadvantage of members of a race, sex, or ethnic group.*²

For the purposes of the Uniform Guidelines on Employee Selection Procedures (UGESP), the “substantially different rate of selection” is triggered at 80%:

*Adverse impact and the “four-fifths rule.” 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. Greater differences in selection rate may not constitute adverse impact where the differences are based on small numbers and are not statistically significant, or where special recruiting or other programs cause the pool of minority or female candidates to be atypical of the normal pool of applicants from that group. Where the user's evidence concerning the impact of a selection procedure indicates adverse impact but is based upon numbers which are too small to be reliable, evidence concerning the impact of the procedure over a longer period of time and/or evidence concerning the impact which the selection procedure had when used in the same manner in similar circumstances elsewhere may be considered in determining adverse impact. Where the user has not maintained data on adverse impact as required by the documentation section of applicable guidelines, the Federal enforcement agencies may draw an inference of adverse impact of the selection process from the failure of the user to maintain such data, if the user has an underutilization of a group in the job category, as compared to the group's representation in the relevant labor market or, in the case of jobs filled from within, the applicable work force.*³

When considering the adverse impact definition and how UGESP characterizes the 80% standard, it is critical for the contractor to not only be mindful of the standard but, more importantly, to understand how the standard is applied during a compliance review. Areas of interest to OFCCP can be identified in the above subsection and contractors should be knowledgeable of the strategies OFCCP will use to assert that selection rates of less than 80% - significant or not - means that there is adverse impact in the selection process.

In terms of OFCCP strategy, Cohen and Dunleavy's (2009) research shows that tests of statistical significance were the standard of measurement used to determine adverse impact in 95% of settlements. Clearly, the use of statistics is a wise and prudent strategy for evaluating an inference of adverse impact and UGESP suggests as much in the Supplement Q&A to the Guidelines at Question 22⁴. However, contractors should keep in mind that sole reliance on statistical tests carries a risk of error (finding a non-significant difference between selection rates and inferring that there is no adverse impact). Zedick (2009) articulates this concern in light of UGESP's statement that "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" (p. 6)(emphasis added). Moreover, non-significant adverse impact may not be sufficient to disprove the inference when ". . .the user had not maintained data on adverse impact as required by the documentation section of applicable guidelines . . . if the user has an underutilization of a group in the job category, as compared to the group's representation in the relevant labor market or, in the case of jobs filled from within, the applicable workforce." (p. 6) Therefore, contractors should be prepared to address the adverse impact indicator, regardless of its statistical significance.

STRATEGIES FOR MANAGING THE IMPACT RATIO ANALYSIS

OFCCP's increased focus on the nondiscrimination provisions of E.O. 11246 requires contractors to give careful consideration to their IRA for it is highly likely that OFCCP will take a particular interest in employment transactions during the compliance review. Additionally, contractors must become more pro-active in the evaluation of these employment practices so that questions of adverse impact are not discovered at the point of the compliance review. With that in mind, the following outlines strategies contractors can use not only to frame IRA results for OFCCP but also enhance the contractors own self-evaluation practices per 41 CFR 60-2 (b) – (d).

TERMINATION ANALYSIS

UGESP requires that each step of the selection process must be analyzed for impact when adverse impact is indicated in the overall selection system.⁵ When analyzing adverse impact in terminations, the same logic should apply whereby the contractor should analyze the organizational and individual behaviors that lead to the termination of employment. By taking this approach, the contractor can be prepared to articulate differences in termination rates between men and women, minorities and non-minorities, and between racial/ethnic subgroups. This strategy has particular significance in light of current economic conditions and the mass downsizings, layoffs, and business reorganizations that have resulted. That said, current laws such as the Worker Adjustment and Retraining Notification Act (WARN)⁶ and court rulings⁷ have provided contractors ample motivation to analyze termination decisions carefully. Regardless of case law and statutory requirements, contractors must be able to support termination decisions with rationales that are unbiased. A good strategy in crafting the rationale should be to first make distinctions in the type of terminations to be analyzed. To that end, the

contractor should separate terminations into four categories: employee-initiated (voluntary), employer-initiated (involuntary), reductions in force (RIF), and overall terminations (voluntary, involuntary, RIF). It is recommended that reductions in force be analyzed separately, even though they are typically involuntary, because they represent business decisions not directly related to the employee (although an employee criterion, such as performance, may be used to categorize employees slated for the reduction). Moreover, the RIF populations sometimes include employees who volunteer for the reduction (in exchange for early retirement, severance). For most readers, this approach will be quite familiar, so what follows are some suggestions on how to look at your data once you've categorized the terminations.

- Do not assume that voluntary terminations absolve you from responsibility. Even though an employee may have voluntarily terminated employment, keep in mind that the context of the termination is still important. If overall adverse impact is indicated, the contractor should ensure that the voluntary terminations were not the result of workplace factors that cause women and minorities to disproportionately exit.

- Involuntary Terminations (without RIFs). These analyses should always be conducted with special care since they are usually the result of an antecedent which led to this outcome. In this sense, analyzing involuntary terminations should be approached much like the analysis of adverse impact in hiring whereby the policies, practices, and procedures leading up to the termination are assessed for fairness.

- Reductions in Force. As discussed, RIFs typically involve the establishment of business-based (*e.g.*, reorganizations, expense management) or employee-based (*e.g.*, performance, tenure) criteria to determine who will be slated for downsizing. Although these criteria may be facially neutral, the contractor should know how these

criteria impact protected groups. As with other involuntary terminations, employee-based criteria should be evaluated carefully and never assume that RIFs are not germane to the discussion of an inference of overall adverse impact.

- Other terminations. To ensure a complete analysis, other termination conditions such as retirees or divestiture often can be removed from the IRA but should also be vetted for impact.

PROMOTION ANALYSIS

OFCCP regulations⁸ define a job group as a category of job titles with similar wage rates, content, and opportunity for advancement. Following this definition, any increase in wages, job duties or other changes in employment status may represent a promotion for purposes of adverse impact analysis or goal attainment. For many contractors, increases in wages, job duties, or responsibilities are often reflected in an increase in pay grade. However, local business practices may not recognize a simple increase in grade as a promotion. On the other hand, the job evaluation process may reassign a position to a higher grade and the incumbent in the re-evaluated position may not be considered promoted in light of the reclassified position. It is within this context that the EEO/AA specialist must determine what a promotion actually is by reconciling job changes—within the parameters of the job group definition—and local business practices. For many contractors, this reconciliation represents a significant challenge in that there is usually a strong disinclination for having compliance requirements driving business practices. Nevertheless, it is vital that the EEO/AA specialist partner with human resources and the business in an effort to best define a promotion that satisfies business objectives and works within the context of the job group definition. Absent a common definition, it is impossible to accurately calculate adverse impact in promotions or progress toward goal attainment.

In addition to the challenge of simply defining a promotion, strategizing the promotional adverse impact must also consider promotions that are the result of internal hiring practices (competitive promotions) and those which are non-competitive. For competitive and non-competitive promotions, promotional criteria must be clearly defined (*e.g.* performance, tenure) valid, and unbiased. However, these promotions differ in that employees may self-nominate in the case of competitive promotions via an internal job posting system and non-competitive promotions are usually the result of an external management decision. Therefore, assessment of the promotional adverse impact should, where possible, segregate these promotion types in an effort to isolate the appropriate promotion pools. From that point a more accurate adverse impact analysis can be performed and communicated to OFCCP.

Promotions tend not to receive much attention during compliance reviews due to the challenge of analyzing the non-transfer movement of employees within and between job groups which is often a function of the contractor's inability to articulate the organizational context from which the promotion is conferred. While it is usually easy to identify who was promoted, the promotion transaction often lacks the procedural foundations that typically support hiring or termination decisions. As a result, the OFCCP often gives up on promotional IRAs as a line of inquiry and the contractor breathes a sigh of relief because it avoids the challenge of preparing a cogent promotion analysis.

Leveraging the aforementioned strategies for analyzing promotional IRAs can be a good first step in organizing the employee data so that the promotion analysis can be attempted. But, regardless of the strategies contractors may use, contractors will do well to craft meaningful approaches to the promotion IRA in advance of OFCCP directing its attention to this area. Keep in mind that with OFCCP's renewed emphasis on affirmative action, goal attainment will likely receive more attention in

the coming years and contractors will need to be able to explain promotional activity clearly. More importantly, however, is the opportunity effective promotional analysis creates in support of the contractor's overall diversity agenda through the development of an analytical framework by which internal applicant pools can be built and analyzed.

APPLICANT ANALYSIS

Applicant and hiring activities are the greatest areas of focus and success by OFCCP and, as such, are the areas of greatest interest to the contractor. Because there can be many factors to consider when analyzing hiring activity, the contractor is obligated to analyze each component individually when overall adverse impact is indicated. To that end, the following outlines some of the most common factors and discusses how they are framed within the context of the Internet Applicant Rule.

The first step in analyzing the hiring adverse impact is ensuring that the applicant pool is accurate within the definition of the Internet Applicant Rule. With that, what follows are a few things to keep in mind when developing the pool.

Expression of Interest.

At this stage of the process, the issue relative to adverse impact is one of determining who has entered the applicant pool. For most contractors, this usually means having an effective intake (*i.e.* requisition management/job posting) system that tracks expressions of interest against specific, open positions. An effective intake system also supports the recordkeeping requirements and the affirmative action requirement to post certain open positions with the State Employment Service. An effective intake system can also aid in structuring the IRA in situations when combining multiple requisitions for the same or similar position is advisable or necessary or when analyzing the applicant pools for positions that are filled on a regular basis, such as positions with high turnover or for job postings that are open for a long period of time.

Considering the Applicant and Basic Qualifications.

Although these parts are separate prongs in the Internet Applicant definition, they typically operate together. Many applicant tracking systems (ATS) feature functionality whereby a position description is affixed to a requisition/job posting and job candidates may express interest in the position based on a perception of their qualifications. Then, the system or recruiter can consider the application as the first step in assessing basic qualifications. This two-part process is critical to how the adverse impact analysis is structured because it determines who is in the applicant pool based on the contractor's control. Therefore, it is of utmost importance that the contractor be able to articulate who was considered and, most importantly, how the candidates' qualifications compared to the stated qualifications in the position description and what criteria was used to advance or disqualify the candidate. Given the criticality of these stages in the selection process, EEO/AA specialists should be working closely with their staffing and recruiting partners to ensure that effective requisition management and screening protocols are in place so as to meet the internet standard.

Candidate Removes Him/Herself from Further Consideration.

This prong of the standard is straightforward but is important to recordkeeping requirements in terms of being able to demonstrate who withdrew from the selection process and for what reason.

Adherence to the Internet Applicant Rule, of course, will not guarantee that you will have no adverse impact. However, following the rule and understanding its components relative to your hiring practices will assist your adverse impact analysis since the rule can provide the contractor with the contextual framework within which hiring decisions were made. With that structure in place, analysis can now be directed toward identifying and describing the data

points that can be used to demonstrate that hiring decisions do not adversely affect protected groups.

CONTRACTOR—KNOW THY DATA

Upon receipt of applicant and hiring data, experienced EEO/AA professionals often recognize adverse impact before any analysis is conducted. At this point the process involves deconstructing applicant flow and hiring decisions to understand what caused the adverse impact. While following the Internet Applicant Rule provides an analytical foundation for assessing selection activity, one of the most important things the EEO/AA specialist can do in strengthening that foundation is to work closely with staffing professionals and other human resources partners to understand selection practices and selection decisions and work with business clients to become knowledgeable of the business' labor requirements. Additionally, the EEO/AA specialist must have a thorough understanding of the contractor's ATS and how it is used to record the creation of a job opportunity, how it announces that opportunity (internally and externally), what recruiting and selection data the ATS collects, and how it tracks candidates through the selection process.

In terms of recording the creation of the opportunity, the EEO/AA specialist should be familiar with the requisition/job posting process: of particular interest are the accuracy of the job description and an understanding of the requisition life cycle. In terms of the job description, the EEO/AA specialist should be satisfied that the stated qualifications accurately reflect job duties. This knowledge provides the foundation of who meets the basic qualifications per the Internet Applicant Rule. Concerning the requisition life cycle, the EEO/AA specialist should be able to answer certain questions on how requisitions are managed. Are requisitions discrete events (one unique job opportunity per requisition)? Can multiple requisitions be opened for the same position

(important for requisition-level analysis of selection activity)? Does the requisition stay open for positions filled on a regular basis, thus requiring analyses to identify the appropriate time parameters of the applicant pool? By answering these questions, the EEO/AA specialist can begin to sort and organize the applicant data in an effort to construct meaningful applicant-to-hire ratios.

In terms of data collection and tracking, the EEO/AA specialist should be familiar with the data points that can be used to defend the adverse impact. For example, data on sourcing can demonstrate that the contractor is taking *affirmative action* to ensure a diverse applicant pool. Additionally, accurate data points on the stages of the selection process (*e.g.* screen for basic qualifications, testing, interview) are invaluable tools for analyzing how job candidates fall out of the applicant pool and aid in isolating stages of the process where additional adverse impact analyses are required.

Effective analysis of the selection process is a task whose purpose cannot be understated. For all that this article has discussed, the ability to effectively analyze the multiple components of employment transactions is the essential tool the EEO/AA specialist must possess to successfully defend the inference of adverse impact and to demonstrate that the contractor takes affirmative action to ensure equal employment opportunity. Absent this tool or the ability to obtain it, any effort to strategize or frame the adverse impact analysis will be compromised from the outset. As a result, OFCCP will be left with no option other than to conclude that any impact ratio less than 80% constitutes adverse impact and potential discrimination.

Not only must the EEO/AA specialist be knowledgeable about the data points relevant to each stage of the selection process, s/he must understand what they mean. That is, the EEO/AA specialist should be aware of not only recruiting results or sources, but also recruiting strategy; the EEO/AA

specialist should also understand how each stage of the selection process was validated to ensure that the stages appropriately predict success on the job.

Lastly, when considering the IRA, a decision must be made regarding the use of internal applicant and hire data. While UGESP is silent on the requirement to use internal applicants and hires in the calculation of adverse impact, in most cases internal and external applicant pools are distinct enough to warrant their separation. Usually, internal applicants are sourced differently, they may have advance knowledge of job opportunities, and they may have the natural advantage of being internal candidates (*viz.* institutional knowledge, may be able to bypass some selection stages). Additionally, keep in mind that internal hires may also be considered promotions in instances where the candidate obtains a position higher than the current level. As a practical matter, it is recommended that internal selections be segregated from external selections, in the first instance, with the understanding that all pools may be combined to support a defense against the inference of adverse impact.

AFFIRMATIVE ACTION AND NONDISCRIMINATION

In the past, OFCCP has treated the nondiscrimination and affirmative action requirements of E.O. 11246 as two fairly distinct constructs, as reflected in its enforcement history. For example, the Bush Administration's OFCCP emphasized nondiscrimination, as seen in Active Case Management and the Systemic Discrimination Initiative. Conversely, the Clinton Administration's OFCCP focused on affirmative action compliance. In turn, contractor compliance efforts also have been directed in a similar fashion. Today, the Obama Administration's OFCCP has indicated that there will be a renewed emphasis on affirmative action (*e.g.* recruitment and outreach; Sec 503 and VEVRAA compliance; goal attainment). This shift, however, should not be construed as OFCCP retreating from

its nondiscrimination enforcement efforts, or, adverse impact. In fact, it is reasonable to assume that, in light of OFCCP's enforcement successes, contractors should be prepared to understand the interaction between nondiscrimination and affirmative action. This may be especially true as it pertains to goal attainment and adverse impact to the extent that disproportionate rates of selection have a deleterious effect on goal attainment. In other words, contractors should consider OFCCP's new emphasis to be another potential tool in its effort to enforce nondiscrimination in employment and should plan their compliance activities with that in mind.

CONCLUSION

OFCCP's increased efforts to enforce nondiscrimination in employment means that government contractors must be pro-active in ensuring that their selection practices are valid and

fair. Furthermore, renewed emphasis on affirmative action compliance suggests that OFCCP will investigate contractor compliance with an integrated approach not before seen in the contractor community. As such, contractors must have strategies in place that take into account the interaction between affirmative action and nondiscrimination. By taking this approach, contractors will be able to position their IRA towards minimizing or eliminating the inference of adverse impact.

This article has offered a few suggestions on what strategies contractors may use to manage the adverse impact indicator. However, close collaboration with human resources and business partners is the necessary ingredient for success. Only through effective partnerships will the contractor be prepared to address questions concerning adverse impact and affirmative action that may present themselves during the compliance review. ☒

ENDNOTES

1. For a good review of the Systemic Discrimination Initiative, see Silberman & Lewis, “The rules have changed” in *EEO Insight*, 1, 7 – 18.
2. Equal Employment Opportunity Commission. 29 CFR 1607.16(b)
3. Equal Employment Opportunity Commission. 29 CFR 1607.4(d)
4. Adoption of Questions and Answers to Clarify and Provide a Common Interpretation of the Uniform Guidelines of Employee Selection Procedures. Equal Employment Opportunity Commission, et al. March 21, 1979.
5. Equal Employment Opportunity Commission. 29 CFR 1607.4(c)
6. Worker Adjustment and Retraining Notification Act. 29 USC 23.
7. *Smith, et al v. City of Jackson*, MS (03-1160) 544 U.S. 228 (2005) 351 F.3d 183, affirmed. March 30, 2005.
8. Office of Federal Contract Compliance Programs. 41 CFR 60-2.12(b)

REFERENCES

- Cohen, D. & Dunleavy, E. M. (2009). A review of OFCCP’s enforcement statistics: A call for transparency on OFCCP reporting. The Center for Corporate Equality, Washington, D.C,
- United States Department of Labor, Employment Standards Administration (2008). OFCCP Directive, transmittal number 285.
- United States Department of Labor, Employment Standards Administration (2009). ESA news release.
- Zedick, S. (2010). Adverse impact: History and evolution. In J.L. Outtz (Ed.), *Adverse impact: Implications for organizational staffing and high stakes selection* (pp. 3-27). New York, NY: Routledge.

Digging Below the Surface: **A CLOSER LOOK AT MULTIPLE REGRESSION ANALYSES**

Desireé Throckmorton | *Biddle Consulting Group, Inc.*

Companies are under a great deal of pressure to explain and defend decisions made regarding employee compensation. With renewed attention from the Lilly Ledbetter Fair Pay Act, the proposed Paycheck Fairness Act, and the Office of Federal Contract Compliance Programs (OFCCPs) continued efforts during audits, human resource professionals are being pushed to conduct statistical analyses of their compensation practices that go well beyond the earlier practices of looking at pay grades and average salaries. Through court precedent and the recent OFCCP guidelines regarding compensation, the use of multiple regression has become one of the accepted tools in the industry.

The goal of this article is to highlight the complexities of multiple regression and explain considerations that should be made prior to conducting the analysis and making pay changes.

OVERVIEW

In today's equal employment opportunity (EEO) compliance environment, human resource (HR) professionals are feeling increased pressure to analyze their company's compensation systems to identify possible discrimination. Traditionally, HR training and education does not include the use of advanced statistical tools such as multiple regression analysis or independent samples t-tests. This lack of statistical expertise is especially problematic when the current statistical tool endorsed by the Office of Federal Contract Compliance Programs (OFCCP) for compensation analysis, as detailed in the Voluntary Guidelines for Self Evaluation of Compensation Practices (p. 35121), is multiple regression (hereinafter referred to as "regression").

Federal contractors are required to analyze their compensation system on, at least, an annual basis to

determine whether gender, race/ethnicity based pay disparities exist. Additionally, the Lilly Ledbetter Fair Pay Act (Ledbetter), which was signed into law by President Obama, will compel employers to carefully scrutinize their entire compensation practices to ensure they are free from discrimination. Furthermore, OFCCP has announced publicly that uncovering unfair compensation discrimination is a priority. With that, and with Ledbetter opening new avenues for pay equity complaints, the HR community is being forced to learn about advanced statistical techniques for conducting compensation analyses, such as regression.

Regression is a complex and powerful statistical tool that can help determine whether observed disparities in pay are due to legitimate job-related factors or, if after controlling for such factors, the disparities appear be related to protected

characteristics such as gender or ethnicity. The purpose of this article is to accomplish three goals. The first is to provide insight into setting up and conducting proper compensation analyses using regression. The second is to illustrate the importance of interpreting the results of a regression analysis correctly. Finally, the article will underscore the reality that it is much easier to conduct (and interpret) regression analyses incorrectly than it is to conduct them correctly. A qualified professional with experience using regression to identify pay disparities should be involved in the development of the statistical model and in the interpretation of the results. Given the potential repercussions of finding problem areas that don't exist, or worse yet, missing problem areas that do, having a second set of eyes assisting in the analysis begins to make a lot of sense.

Included here are a series of key items that a practitioner should consider when utilizing regression as a tool for conducting advanced pay equity analyses. While the items discussed below are not intended to be an exhaustive list of how to prepare for regression analyses, a better understanding of the concepts will help a contractor have confidence in their findings.

GOOD RECORDKEEPING IS CRITICAL

Contractors with inadequate recordkeeping provide an easy target for OFCCP. The Compliance Officer may ask a simple question, "Do you have information X to explain compensation decision Y?" If the employer's response is, "No" or if inaccurate information is provided to the OFCCP then there is a serious recordkeeping issue even before analyses are conducted. The issue of absent or flawed data is pervasive among most Human Resources Information Systems and Applicant Tracking Systems. Typically, the responsibility of recordkeeping lies with the HR professional and their often limited resources. With proper recordkeeping fundamental to compensation analyses, it is crucial that HR professionals have access to the necessary

data. This need for comprehensive and accurate recordkeeping only increases with the passing of Ledbetter. This Act establishes that the 180-day (or 300 days depending on the state jurisdiction) statute of limitations for filing claims of unlawful pay discrimination resets each time an employee receives a paycheck. For companies with more than 150 employees, the regulations require that Federal contractors retain records for two years from the date of the employment decision. However, with Ledbetter this means that companies will need to retain records for a longer period of time and in more detail than ever before. The recordkeeping required by Ledbetter will be a challenge to most companies because it will require retention of detailed records concerning pay decisions (including reasons for starting salaries) as well as information regarding performance appraisals for a potentially unlimited amount of time.

To explain pay disparities within jobs, an employer would likely be required to preserve detailed employment records such as prior salary, tenure within the industry, level of education, defensible performance scores and more for *each* employee. More often than not, employers do not collect and/or retain this type of information, at least not electronically. While most employers will likely have hire date and perhaps performance scores, the more pertinent explanatory data is far less likely to be available in a readily-analyzable format. Regression utilizes these explanatory variables and more to statistically explain pay differences. If the relevant data has not been tracked or retained, then the value of the analysis could be extremely limited.

The *accuracy* of the data is another key point that deserves consideration. If the data is inaccurate or incomplete then regression may not provide meaningful results. One of the assumptions of multiple regression is that valid data is available for all employees for each variable included in the regression model. If data is missing for an employee

for one or more of the explanatory variables being used to model compensation, that employee will be dropped from the analysis. This becomes especially important when dealing with either small sample sizes or large amounts of missing data. For example, if 20 females are employed in a particular job title and education data is missing for 10 of them, half of the female employees would be dropped from the analysis. This situation leads to invalid and indefensible results.

One way of addressing this issue of missing data is to perform what is called “data imputation.” To impute data means estimating missing data based on available data. For example, one technique that could be used in the situation above would be to calculate the average education of females who have valid data and then “replace” or “impute” that value to those with missing data. The problem with imputing data is that there is no way of knowing if the employees were assigned accurate values of the variable. Therefore, the results will be highly questionable and likely will not stand legal scrutiny. The author does not encourage imputing data. Rather, if valid data cannot be gathered for all employees (or nearly all employees) for a specific variable, it is recommended to exclude that variable in the regression analysis. This is the only way to have confidence in the results obtained.

CHOOSING THE CORRECT VARIABLES

In an ideal situation where clean and accurate data are available, the next step is to choose the variables

that a company would expect to predict employee compensation. It is common to begin the analysis with readily-available variables that are universally applicable across all (or most) job titles, such as tenure and performance. From there, the user can run the initial analysis to identify problem areas that exist after accounting for tenure and performance. If problem areas remain, it is recommended that the user identify additional explanatory variables for just those analyses where tenure and performance were not sufficient to explain the differences. To be precise, these additional variables should be specific to the job(s) being analyzed and it should never be assumed that one set of variables will apply to all jobs. For example, it may be appropriate to consider education as an explanatory variable for particular jobs, but certainly not for all jobs. Be wary of approaches that suggests including all variables for all jobs regardless of their statistical value without considering the variety of jobs within the company.

As previously mentioned, it is important to first evaluate the variables used in the analysis. Are they significant predictors of pay? Can their inclusion be supported by company policies? One technique to accomplish this is to correlate salary with each explanatory variable. Examining a scatterplot of compensation and each predictor variable is an appropriate technique for determining which variables affect compensation. See Table 1 for a sample correlation matrix that shows a statistical relationship between Tenure and Salary. The *p*-value

Table 1: Correlation between Tenure and Salary

CORRELATION		Salary
Tenure	Pearson Correlation Coefficient	.918**
	Sig. (2-tailed)	.000
	Sample Size	1591

**The correlation is significant at the 0.01 level (2-tailed).

for the correlation ($p=.000$) is well below the .05 threshold for being considered significant, therefore there is a statistically significant relationship between tenure and salary. The closer the Pearson Correlation value is to either -1.0 or 1.0 the greater the correlation. Notice that the correlation in Table 1 is positive and close to perfect – as tenure increases, so does salary. In other words, the longer an employee is with the company, the more money they tend to make.

Conceptually, an HR manager may assume that performance would be a reliable predictor of pay; however, a statistical analysis may reveal otherwise. Therefore, it is important to determine whether each variable correlates with salary. HR professionals should have the ability to recognize the importance of correlating the variables with pay, or have the ability to interpret the results correctly. It is also important to remember that certain variables may be more subjective than others and are therefore likely to be held to additional scrutiny by OFCCP. For example, if performance scores are used to explain differences in compensation, the employer should expect OFCCP to investigate whether the performance scores themselves are *tainted* (i.e., discriminatory). This is not to say that performance scores fail to predict compensation, but rather that they must be non-discriminatory to be used to explain compensation. And, rather than just evaluating average differences between groups on performance scores, complex multiple regression techniques—known as moderated multiple regression (MMR)—are required to determine whether performance scores do in fact operate differently between groups with respect to predicting pay.

According to OFCCP's Federal Register *Interpreting Nondiscrimination Requirements of EO 11246 With Respect to Systemic Compensation Discrimination*, (p. 35138), it is not appropriate to use tainted variables to defend pay decisions. This means that how the explanatory variables relates to both pay and the protected variables (e.g., gender/race) should be

examined to be sure they are not discriminatory. This is another important layer of regression analysis that needs to be considered. The appropriateness of the selected variables can be determined by using MMR to evaluate the “interaction” between gender or minority status and each predictor variable.

EXPLORE THE DATA FOR POSSIBLE INTERACTIONS

As mentioned above, OFCCP's Federal Register regarding *Respect to Systemic Compensation Discrimination* (p. 35124) declares that any variables used to explain differences in compensation cannot themselves be discriminatory. Such variables are often referred to as being tainted. In regression analyses, users can investigate whether variables are potentially tainted or not by reviewing their interactions. An interaction is said to exist when the relationship between a predictor and salary is different across levels of some other variable such as gender or ethnic group.

In regression, the “main effect” is the relationship between each single explanatory variable with salary. “Standard regression models that include only ‘main effects’ are forced to assume that these [predictor] variables impact all the employees in the group in the same way” (Biddle, 2006). This assumption, however, is not always correct. An example of this is when compensation is predicted by tenure, but where each year of tenure means something different for men versus women (e.g., men receive an additional \$500 per year of tenure while women receive only \$400). Even if a protected variable such as gender or race is not itself a significant contributor to salary differences, employers should be aware of such significant interactions because they too can be considered indicators of potential discrimination.

Graph 1 below exhibits an interaction where males and females are compensated differently based on their tenure. Notice that during the early years of employment females are paid noticeably less than males for the same tenure. More specifically females

who have been with the company for approximately 2.5 years are paid around \$5,000 to \$10,000 less than males with roughly the same tenure. Meanwhile during the higher years of tenure males are paid noticeably less than females. Over time the compensation of females surpasses that of males. It is useful, therefore, to analyze compensation data to identify such interactions.

As a safeguard against interpreting statistics out of context, OFCCP typically gives serious consideration to anecdotal evidence to support any position they may take regarding the results of a regression analysis (Compliance Manual, at Section 7D05(e)). This means they are likely to conduct a cohort analysis to further explain statistical results. A cohort analysis is a non-statistical comparison of the employees who were associated with the compensation disparity. Employers can protect themselves by conducting these proactive internal analyses to determine why these interactions may be occurring.

A forthcoming book by Nooren and Biddle (in press) titled, “Compensation Analyses: A Practitioner’s Guide to Identifying and Addressing Compensation Disparities” explains how to use MMR to investigate tainted variables more in depth.

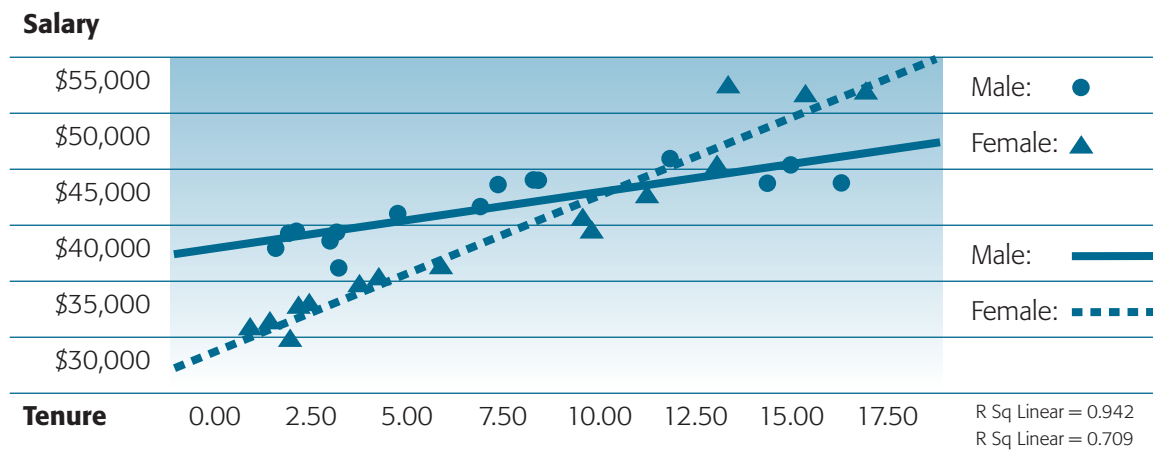
MISUSE OF VARIABLES

There are additional considerations to make after the data have been deemed complete and accurate but before analyses are begun. For example, have the

variables been coded properly? Some variables are not appropriate for use in statistical analyses because the values do not represent a true number (*i.e.*, they are measured on a nominal scale). Ethnicity is an example of a relevant variable (it represents a protected characteristic) that is appropriate to include in a compensation analysis. However, users cannot calculate an “average ethnicity” even if each ethnicity is associated with a number (such as 1=Caucasian, 2=African American, 3 = Hispanic, 4 = Asian, 5 = Native American) because ethnicity is a nominal level variable. A nominal level variable is the least precise level of measurement and it is not appropriate for use in complex mathematical calculations unless such data are coded appropriately, as described below. Nominal level variables simply provide a name or label rather than a value that is considered a number with an absolute zero or other properties with quantitative meaning.

It is possible to utilize ethnicity and/or gender in regression but a technique referred to as “dummy coding” must be used. Dummy coding refers to a process where variables are recoded into a dichotomous variable, such as 1 = non-minorities and 0 = minorities. When dummy coding is employed, it results in a special case variable that is appropriate for regression. The key point is that it is a common mistake to misuse data in regression. Misusing variables can result in meaningless or misleading analyses.

Graph 1: Scatterplot of Salary by Tenure by Gender



Multicollinearity is a technical term for what happens when two or more explanatory variables in a regression model explain the same variance. In other words, two or more of the explanatory variables are either perfectly or nearly perfectly correlated with each other and therefore are explaining the same variation in compensation. While it might seem logical to include as many variables as possible to explain as much of the expected pay as possible, the reality is that it could do more harm than good if those variables are redundant. OFCCP specifically cautions federal contractors regarding multicollinearity in the Federal Register regarding *Respect to Systemic Compensation Discrimination* (p. 35131) stating that when two variables are significantly correlated the accuracy of the regression model’s prediction of compensation becomes less precise.

Statisticians tend to agree that it is ideal to explain the most variance with the least amount of variables. If *tenure*, as defined as the time the employee has been with the company, greatly impacts how a person is paid, then *tenure* would hopefully account for the most variance in a regression model. Additionally, the variable *time-in-job*, as defined as the time the employee has been in their current position, can be added to the regression model to help successfully explain differences in pay. Before calling OFCCP to tell them there are no longer unexplained pay disparities after including *time-in-job*, employers should first determine whether *tenure* and *time-in-job* are

significantly correlated themselves. It is possible that the two variables are overlapping and account for the same variance, thus compromising the analysis. Industry professionals suspect multicollinearity exists if the correlation coefficient between the two variables is .80 or greater (Biddle, 2006). Table 2 is a correlation matrix that illustrates a significant relationship between *tenure* and *time-in-job*. The correlation coefficient is .889, which is above the .80 threshold that Biddle applies. This relationship suggests that the two variables are accounting for the same variance and both should not be included in the model. In this situation it is best to include the variable that has the greatest relationship (*i.e.*, correlation coefficient) with salary and exclude the other variable.

It is important to note that investigating multicollinearity is a very complex process and a thorough investigation requires more than only looking at the correlation between two variables. Advanced statistical diagnostics—such as the Variance Inflation Factor and Tolerance—that are available in statistical packages such as SPSS can be useful tools for this.

Multicollinearity is just the tip of the iceberg of complexity that is multiple regression. While a few key issues have been introduced here, there are other considerations that need to be made before users can adequately interpret the results and have confidence in their findings.

Table 2: Correlation between Tenure and Time-in-Job

CORRELATION		Time-In-Job
Tenure	Pearson Correlation Coefficient	.889**
	Sig. (2-tailed)	.000
	Sample Size	550

**The correlation is significant at the 0.01 level (2-tailed).

ASSUMPTIONS OF MULTIPLE REGRESSION

“Assumptions” are essentially requirements that must be met for regression to be appropriate. One of the key assumptions of regression is called “the assumption of homoscedasticity.” Homoscedasticity suggests that predicted compensation should be equally valid across the entire range of the predictor variables (*e.g.*, tenure). The opposite of homoscedasticity is “heteroscedasticity” which results in predictions of compensation that are more accurate at some points in the explanatory variable and less accurate at others. For example, one possible pattern of heteroscedasticity is one where the regression model has a higher accuracy of predicting compensation for employees with lower tenure than it does for employees with higher tenure. Heteroscedasticity can lead to inaccurate and misleading results. One of the most common approaches to evaluating this issue is through a careful review of scatterplot graphs where each explanatory variable is plotted against compensation.

Regression also assumes a normal distribution, so evaluating the scatterplot graph of each variable allows the user to confirm that the distribution is normal and not skewed. Graphically viewing the data is also helpful in determining whether outliers exist. Some practitioners simply remove the outlier from the group and reassess the data; however, outliers with respect to compensation analysis could be evidence of discrimination. It’s important not to neglect to ask “why is that employee paid so much more, or less, than the others in the group?” Perhaps

they were classified incorrectly or the outlier could be a potential issue worth investigating.

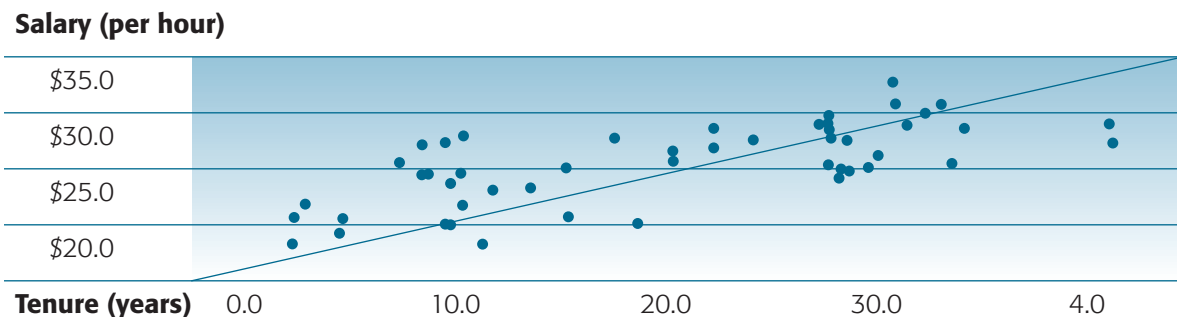
Another assumption is that of “linearity.” Regression assumes that there is a linear relationship between all predictor variables and the dependent variable (*i.e.*, salary). For example, as tenure increases is there a straight line relationship where people earn higher compensation? To determine this, users can create a scatterplot graph; see Graph 2 for an example. Notice that the majority of the scores are concentrated around the regression line, indicating that a linear relationship exists between salary and tenure. As tenure increases so does salary, in most cases.

However, it is rare to find a perfect (or near-perfect) linear relationship between variables such as those found in compensation analyses. The reason is that there are so many factors involved in compensation that it is virtually impossible to isolate each individual factor. Anyone attempting to conduct regression analyses must make a judgment as to whether the relationship is sufficiently linear enough to support a valid analysis. In cases where the relationship is non-linear but the variable is clearly related, there are mathematical techniques called non-linear transformations that can be employed to correct for non-linearity.

CONCLUSION

It is not the intent of this discussion to discourage the use of regression as a tool for determining statistical disparities in compensation; rather, it is the intent of this article to reinforce the understanding

Graph 2: Scatterplot of Salary (per hour) and Tenure



that conducting a valid regression analysis is technical and requires at least a moderate understanding of the statistic. HR professionals who intend to utilize regression should first become educated regarding all of the assumptions, techniques, and considerations that are involved before interpreting the results and potentially making changes to a company's compensation system.

HR professionals should be wary of automated

regression approaches that do not consider the assumptions broached in this article. There is no way to take the “art” out of the process of determining whether assumptions are being violated, data are appropriate, or other factors exist that impact the analysis. Whether completed in-house or outsourced, it is important that an organization's data integrity is analyzed prior to completing and implementing results of regression. ☒

REFERENCES

Biddle, D. (2006) Using multiple regression analysis to examine compensation practices, *Adverse Impact and Test Validation* (pp 145-146). Hampshire, England: Gower Publishing Limited.

Federal Register – Office of Federal Contract Compliance Programs, Department of Labor, Interpreting Nondiscrimination Requirements of Executive Order 11246 With Respect to Systemic Compensation Discrimination, Vol. 71, No. 116, Friday, June 16, 2006, Notices.

Office of Federal Contract Compliance Programs (1993), Federal Contract Compliance Manual. Washington, DC: Department of Labor, Employment Standards Administration, Office of Federal Contract Compliance Programs, Section 7D05(e).

Strategies to Avoid Adverse Impact
**WHEN CHOOSING A TEST FOR SELECTION
OR PROMOTION**

Meghan R. Lowery | *Southern Illinois University Carbondale*

Matthew J. Borneman | *University of Minnesota*

Adverse impact is an important consideration for any organization. According to the Uniform Guidelines for Selection Procedures, organizations are responsible for the effect of what they use as a selection device.

Not only will adverse impact against a protected class result in increased risk for a lawsuit, but there is also the social imperative to increase diversity within the company. As such, it is necessary to examine strategies to decrease adverse impact. The goal of this paper is to provide practical advice for HR professionals when choosing a new test or assessment for selection or promotion purposes. The advice presented here includes both what to look for in the test manual and questions to ask the consultant providing the test. Guidelines on how to interpret the answers to these questions are also provided.

**STRATEGIES TO AVOID ADVERSE
IMPACT WHEN CHOOSING A TEST
FOR SELECTION OR PROMOTION**

ADVERSE IMPACT (AI) is the result of an employment practice that has an adverse effect on the hiring rates of a protected class; that is, the percentage of people selected for one protected class is substantially lower than that of another group. Protected groups can include (but are not limited to) those groups based on age, gender, and ethnic group. The protected class member may be disadvantaged—or adversely impacted—by the particular employment practice; this is often the case in employee selection/promotion. This adverse impact is often shown by a statistical disparity between the selection rates for two groups of people; that is, the selection rate for one group is less than 80% of the rate for another group (*i.e.*, the 80% rule), and/or there is a statistically and practically significant difference in selection rates.

Historically, finding AI within an organization can increase legal risk (Zedeck, 2010) for lawsuits stemming from discrimination claims. Organizations may also feel it is necessary to address AI as a social or moral imperative as well as to increase diversity at multiple levels of the organization. Avoiding AI is a complex adventure in employment practices; this paper seeks to provide practical advice as to how to avoid adverse impact when choosing a new assessment solution to select or promote employees.

ASSESSMENT OF THE TEST MANUAL

Three important criteria for any selection device are having a test that predicts relevant organizational criteria, shows substantial return on investment, and does not result in needless risk of AI. Fulfilling all three criteria is a difficult, but worthy goal. Although this article emphasizes considerations involving AI, the user should keep in mind that absence of AI does

not make a selection procedure useful if it does not meet the other criteria. The following are suggestions as to what to look for in a test manual to help avoid AI. When considering a test or assessment battery to enhance current selection/promotion practices, it is extremely important to examine components of the test manual provided by the test vendor.

Comparison of Sub-Groups

Actual differences between groups is not proof of AI; however, it does provide an indication of the likelihood that AI might result from use of the test. To compare group differences, the standardized mean difference, or d , should be provided. To calculate a d -value, the mean score of the minority group is subtracted from the mean score of the majority group and the difference divided by the combined standard deviation of the two groups. The resulting value is represented in standard deviation units: a positive value represents a greater mean value for the majority group, a negative value represents a greater mean value for the minority group, and a value of zero indicates no difference (Sackett, DeCorte, & Lievens, 2010). Smaller d -values (e.g., around 0.20) are less likely to yield AI, whereas higher d -values (e.g., above 0.80) carry a greater risk. When examining a test, it is helpful to look for multiple d -values to compare each subgroup to its corresponding majority group. For instance, find assessments of d -values in the technical manual between males and females for any expected gender differences. Additionally, the test manual should provide a d -value score between each race and the majority group (Whites) to determine if there are any ethnic group differences. To avoid violations of the Age Discrimination Act, be mindful of any differences between people under forty and over forty years of age. It is also useful if there is information on the performance of the younger members compared to the over forty group (perhaps 40 to 54) and the older members (perhaps 55 and

older). The test manual should provide assessments of each of these possible group differences. Additionally, these d -values should be based on large sample sizes; this helps to ensure that they converge on the true value. When d -values are based on smaller sample sizes, they may substantially over- or underestimate the actual magnitude of the group differences. This, in turn, leads to inaccurate assessments of the risk of AI.

To illustrate this issue, we offer the following example by examining two meta-analyses that document the magnitude of ethnic group differences. The first study, Roth, Bevier, Bobko, Switzer, and Tyler (2001), researched differences between ethnic groups on cognitive abilities tests; the second study, Foldes, Duehr, and Ones (2008), researched differences between ethnic groups on personality variables. Both of these studies have combined multiple individual studies, resulting in large sample sizes and accurate estimates of the group differences. The tests based on cognitive abilities have large group differences between Whites and many other ethnic groups; as such, using these tests is likely to result in AI. However, personality variables tend to have substantially smaller group differences and therefore tend not to have much risk of AI. Comparing the results of these meta-analyses gives some idea of the potential risk for AI depending on the variables; looking for similar data in test manuals will aid the decision-making process when choosing among tests.

Studying information on group differences in the manual is useful and important. But it is also important to keep in mind the extent to which the measured group of applicants or employees in your organization is likely to be similar to or different from the groups on which the differences were measured. For example, if differences were previously measured on a group with markedly different educational characteristics, the measured differences may not be very relevant to what you might find in your own use.

Adverse Impact Analyses from Other Studies

Examination of AI analyses from other studies is recommended. These analyses help give some indication as to how the test/assessment battery (hereafter referred to only as the test) has been used in other selection/promotion situations. If the test is used to control applicant flow or if a strict cutoff value is used, it is preferable to be able to examine AI using multiple cutoff values. These multiple cutoff values can be found in the technical manual. Multiple cutoff values can lend valuable information in decision-making about using a selection device. For example, if a test shows no AI with a cutoff value at the 70th percentile, but yields AI with a cutoff value at the 80th percentile, this might suggest that a lower cutoff would be useful to help avoid AI risk. In this case, even a 10% drop in the cutoff score can dramatically decrease the risk of AI and even increase the representation of the protected group among those passing the test. Additionally, this comparison between cutoff values allows an examination of performance of the test under different situations. A test may yield different results under a variety of testing environments, and it is helpful to establish a range of possible performance scores to indicate the likelihood of creating an AI-inducing situation.

Predictive Validity Evidence

The predictors being considered for selection and/or promotion purposes should show substantial and meaningful correlations with relevant criteria. Not only is this a good idea in order for these selection measures to have a large return on investment, but these relationships will help to mitigate legal risk. It is a well known fact within the testing and legal communities that when a selection/promotion practice shows evidence of AI, that practice must show business necessity; this is typically shown by substantial statistical relationships with meaningful criteria. A wide variety of predictors have typically

been shown to have substantial relationships with job performance (Schmidt & Hunter, 1998); however, the choice of any new specific predictor must still be examined for predictive validity evidence, as previous research does not *necessarily* translate to new measures.

When looking for validity evidence, larger *d*-values tend to need to show larger predictive validity coefficients (*i.e.*, correlations with relevant criteria). The reasons for this are twofold. First, while larger *d*-values increase legal risk from AI, proportionally large predictive validity coefficients tend to mitigate this risk. Secondly, legal guidelines imply that AI can be tolerated if the tests predict relevant criteria and there is not an alternative that has substantially equal predictive validity. Additionally, when examining a test manual, predictive validity coefficients should have larger sample sizes (*N*) and be drawn from multiple samples; this helps ensure the stability of the predictive validity coefficients and makes sure they converge on the true values. Also, results from previous studies should be assessed in a job or industry setting as close to the organization as possible. Specifying and matching settings will aid in generalization of the results to the current setting (Guion, 1998). Once again, the more similar your group is to the groups previously studied, the more relevant the previous results are to what might happen in your organization.

Finally, determine whether or not any statistical corrections have been made to the validity coefficients provided. These corrections serve to provide better estimates of the true predictive validity that can be expected in an operational setting. Preferably, examine both uncorrected and corrected correlation coefficients; this will help ensure that tests can be compared on equal footing when choosing among tests from different vendors. Coefficients can be corrected for criterion unreliability because the outcome does not influence decisions on individual scores/applicants and because these corrected relationships better estimate how well test scores

predict the true nature of the criterion. On the other hand, corrections for unreliability in the predictor variable(s) should never be employed, as selection/promotion decisions are made based on actual scores; predictors with considerable randomness to their scores will necessarily be less useful and show less return on investment than those predictors with little randomness.

Construct Validity Evidence

Another aspect of drawing inferences from test scores that must be considered is construct validity. Construct validity refers to whether the test measures what it purports to measure. In the case of personnel selection tests, the test needs to measure knowledge, skills, and abilities that are directly relevant. As such, to develop construct validity evidence, an effort first must be made to link desired job performance to constructs believed to underlie the behavior necessary to perform the job well. Each job-relevant construct should be clearly identified and distinct from other constructs. Selection tests then need to specifically measure applicants on only these job relevant constructs. If the test is purported to only measure job relevant constructs then it has construct validity; this can be assessed by high correlations with other tests known to measure the same constructs. Any constructs that are not job-relevant measured by a selection test reduces the test's construct validity and increases the chance of adverse impact. Doing so improves the chance that a test can be focused on ensuring each portion of a desired behavior is measured adequately. This specificity is necessary to ensure the test is measuring what it purports to measure, only the knowledge, skills, and abilities directly linked to successful job performance.

For some very specific and clearly defined knowledge, skills, and abilities, requirements for construct validity may be relaxed. For example, an information technology test that purports to measure C++ programming skills can be fairly obvious that it

measures what it intends to measure. However, some form of construct validity evidence is still necessary. This may take the form of large relationships with established measures of C++ programming, or it may take the form of expert ratings that the test sufficiently covers the domain of C++ programming and does not leave out important elements. Some of this evidence may also be available under the term “content validity,” but the same principle holds that evidence needs to be presented that the test does measure well what it intends to.

ASSESSMENT OF THE CONSULTANT

If your organization is working with a consultant, you may have additional questions that can be answered by the consultant. The consultant is likely to be the person most familiar with the test, and can help give advice to avoid risks associated with AI. However, ultimately the user, not the consultant, is responsible for any adverse impact that may occur.

Questions for the Consultant

The first and clearly the most important question to ask the consultant is for the technical manual for the test; this will address the questions discussed above. This is also the time to ask any of those questions that are not addressed in the technical manual. If the consultant does not have access to the test manual, ask the consultant about the questions mentioned previously. Regardless of how they are gathered, answers to the above questions are key to informing the HR practitioner when making a decision to help avoid legal risk from AI. Use this time to gather information for making a beneficial decision about the test.

In addition to questions answered by the technical manual, it is important to examine the test itself; go through the test and assess how the test is measuring what it claims to measure. In other words, examine your impressions of the face validity of the test. As a whole, it should be evident that the test does measure

what it claims to. However, it is important to note that unique single items are generally not cause for concern; it is often the case that some items which are less obviously related to the intended construct have been shown through empirical evidence to relate to the same construct. But if a large proportion of items appear not to relate to the focal construct, this may be cause for concern. If this is a paper-and-pencil or static online test (*i.e.*, not adaptive) it will be possible to examine all the items. If the test is simulated, adaptive, or otherwise relies on a progression based on the ability of the user, it may not be possible to view all items in one pass. Take the opportunity to go through the test several times to gain an understanding of how the test works. Face validity, or being clear about what the test is measuring, is an important factor in anticipating future applicants' reactions to the test.

If the test has been used in similar environments, inquire about past HR managers and or recruiter reactions. Are they satisfied with what the test purports to measure? Do they feel that it helps them make better selection/promotion decisions? Additionally, ask about past applicant reactions. Are applicants satisfied with the test and do they understand how it relates to the job they are applying for? Is there any dropout associated with the assessment, and if so, is there a particular subgroup that is affected by this dropout (*i.e.*, non-random)? Do applicants dislike the test? Genuine dissatisfaction amongst applicants may lead to problems later, and unfavorable impressions of the company.

Assess, and ask for input from the consultant on, how this test or assessment battery could complement and enhance current human resource practices. Is the test a beneficial addition to the

assessment procedures currently in place, or is it a replacement? If adverse impact analyses are not currently conducted in your organization, can the consultant provide assistance with these calculations before and/or after the adoption of a new assessment procedure? Evaluation of current practices may reveal unknown issues or benefits that may affect the final decision made about the new assessment procedure(s).

The consultant can also be asked to help link the assessment to the current competency model for the organization's hiring practices; this will help show the job-relatedness of the test. Asking the consultant to conduct a job analysis to link the assessment to the job requirements is an important component of establishing the usefulness of the test. Conducting a job analysis will allow the consultant to provide a list of job responsibilities and requirements, which can then be connected to desired job performance. Both of these strategies help to illustrate the business necessity of using the assessment, decreasing legal risk of AI. The selection test or assessment chosen should ultimately be predictive of the desired job performance.

CONCLUSION

Taking the time to critically evaluate the test manual and/or garner information from the consultant will help ensure the test chosen is satisfactory. Satisfaction with the test or assessment can stem from several areas, such as validity of the measure, the expense and utility associated with the test, and if the test mitigates AI. Avoiding AI is important not only for legal reasons, but for moral and social grounds as well. Ensuring any new assessment will not negatively impact current HR practices yields a hiring practice the organization can be proud of and confident in the avoidance of adverse impact and legal risk. ☒

REFERENCES

- Foldes, H. J., Duehr, E. E., & Ones, D. S. (2008). Group differences in personality: Meta-analyses comparing five U.S. racial groups. *Personnel Psychology, 61*, 579-616.
- Guion, R. M. (1998). The legal context for personnel decisions. In R. M. Guion (Ed.) *Assessment, Measurement, and Prediction for Personnel Decisions*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Roth, P. L., Bevier, C. A., Bobko, P., Switzer, F. S. III, & Tyler, P. (2001). Ethnic group differences in cognitive ability in employment and educational settings: a meta-analysis. *Personnel Psychology, 54*, 297-330.
- Sackett, P. R., DeCorte, W. & Lievens, F. (2010). Decision aids for addressing the validity-adverse impact tradeoff. In J. L. Outtz (Ed.), *Adverse Impact: Implications for Organizational Staffing and High Stakes Selection* (pp. 3-28). New York: Routledge.
- Schmidt, F. L. & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*, 262-274.
- Zedeck, S. (2010). Adverse impact: History and evaluation. In J. L. Outtz (Ed.), *Adverse Impact: Implications for Organizational Staffing and High Stakes Selection* (pp. 3-28). New York: Routledge.

How Croson Studies can be used to
EFFECTIVELY DEVELOP DEFENSIBLE DIVERSITY INITIATIVES IN A POST-RICCI WORLD

Dan A. Biddle, Ph.D. | *Biddle Consulting Group, Inc.*
 Richard E. Biddle, Founder | *Biddle & Associates, Inc.*

INTRODUCTION

On June 29, 2009, the U.S. Supreme Court handed down the first Title VII ruling that answers the difficult question: “Under what circumstances can an employer subject to Title VII implement otherwise prohibited disparate-treatment discrimination to avoid disparate impact liability?” When answering this question, the U.S. Supreme Court adopted a “strong-basis-in-evidence standard” as a matter of statutory construction for courts to use as a means of resolving conflicts between Title VII’s disparate-treatment and disparate-impact provisions—“allowing violations of one in the name of compliance with the other only in certain, narrow circumstances” (*Ricci*, p. 23).

Twenty years prior to the *Ricci v. DeStefano* decision, the U.S. Supreme Court codified the “strong-basis-in-evidence standard” in *Richmond v. Croson*¹—a concept that first originated three years earlier in *Wygant v. Jackson Board of Education*² by Justice Powell. Both of these cases laid down legal concepts that were relevant to the *Ricci* context in ways that will be fully explored in this article—defining how the *strong basis in evidence standard* applies

to various personnel actions and diversity initiatives in a Post-Ricci world.

CASE BACKGROUND

The *Ricci v. DeStefano*³ case involved the City of New Haven’s (Connecticut) firefighter promotional practices. The foundation for the City’s selection procedure is an agreement with its firefighters’ union that only a written test and an oral interview would be used as the selection devices, weighted 60% and 40% respectively. Although the weighting came about as a result of negotiations with the union, it was apparently not based on job analysis or validation research. Candidates were required to be placed on the eligibility list in rank order of their scores, even though the measurement properties of the tests could not distinguish between candidates in such a finite manner. The City’s charter required promoting only from the highest scoring three candidates, called the “rule of three,” even though there could be other candidates just as qualified, but only a few decimal points away in score. The examinations were administered in November and December 2003. The results of the exams are shown in Table 1.

Table 1: Ricci Exam Results

Exam/Status	# in Group (Passing Rate %)			
	Total	White	Black	Hispanic
Lieutenant Exam (taking)	77	43	19	15
Lieutenant Exam (passing)	34	25 (58%)	6 (32%)	3 (20%)
Captain Exam (taking)	41	25	8	8
Captain Exam (passing)	22	16 (64%)	3 (38%)	3 (38%)

Under the “rule of three,” the top 10 candidates were eligible for an immediate promotion to the 8 open Lieutenant positions. All 10 were white. Subsequent vacancies would have allowed at least 3 black candidates to be considered for promotion to Lieutenant. Seven (7) Captain positions were vacant at the time of the examination. Under the “rule of three,” 9 candidates were eligible for an immediate promotion to Captain (7 whites and 2 Hispanics).

After giving the tests and learning that there was substantial adverse impact on minority candidates (80% rule violations⁴ and various levels of statistical significance based on each exam and how the groups could be combined), the City discarded the test results. 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 faced two choices for a lawsuit: 1) if the City discarded the test results, a lawsuit from whites who would have likely been appointed had the lists been adopted; or 2) a lawsuit 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 elected to discard the exam results and face the possibility of a disparate treatment type of lawsuit.

The Court ruled that the City’s action was impermissible in this case: “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 to justify such drastic action: “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” (*Ricci*, pp. 33–34). The City did not provide the required strong basis in evidence—only unsupported arguments were presented by the City.

EVALUATING THE RELEVANCE OF RICCI IN BOTH PUBLIC AND PRIVATE SECTORS

Because *Ricci* involved a promotional exam given to city firefighters (government employees), many might question the relevance of the ruling to private sector employers. However, both public and private employers are subject to Title VII. Further, *Ricci* was decided on statutory grounds, not constitutional grounds. “We hold only 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” (*Ricci*, p. 25).

Because state and local governments are *public sector employers*, their hiring, promotional, and contracting practices are covered by Title VII and are subject to the *strict scrutiny* theory of constitutional law. The Court used principles from constitutional cases for guidance: “Our cases discussing constitutional principles can provide helpful guidance in this statutory context” (*Ricci*, p. 23).

The cornerstone case that pertains to the strict scrutiny and strong basis in evidence standards in the public sector is *Richmond v. Croson*.⁵ In *Croson*, the Supreme Court codified the strong-basis-in-evidence standard that originated three years earlier in *Wygant v. Jackson Board of Education*.⁶ Both of these cases involved the strong-basis-in-evidence standard in ways that were relevant to the framework later used in *Ricci*.

The *Croson* case 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 are 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 or 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 other gender or race/ethnic groups that are not part of the program.

Because private employers are not covered by the Fourteenth Amendment,⁷ they may be held to a lesser standard than the strict scrutiny standard. Nevertheless, 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” (*Ricci* p. 20). This admonition appears to speak to both public and private employers alike.

Clearly, the Court used concepts from prior

constitutional cases and applied them in *Ricci*: “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” (*Ricci*, pp. 22-23). The basic concept put forth by the Court in *Ricci* applies to private and public employers—before an employer can implement intentional discrimination on a protected group (race, color, religion, sex, national origin) for the purpose of avoiding a potential disparate impact liability, the employer must have a “strong basis in evidence” that it will be subject to disparate impact liability if it fails to take the discriminatory action. The “strong basis in evidence” is the common thread for both public and private employers, and, since it is now a statutory standard, must be applied in Title VII cases involving both public and private employers.

THE STRONG-BASIS-IN-EVIDENCE STANDARD

In the *Ricci* case, the Supreme Court stated: “We hold only 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” (*Ricci*, p. 26). The Court’s analysis begins with the premise that the City’s actions would violate the disparate-treatment provisions of Title VII unless there was *some valid defense rooted in a strong basis in evidence*, and reasons: “All the evidence demonstrates that the City chose not to certify the examination results because of the statistical disparity based on race. (*i.e.*, how minority candidates had performed when compared to white candidates). Without some other justification, this express, race-based decision making violates Title VII’s command that employers cannot take adverse employment actions because of an individual’s race” (*Ricci*, p. 19).

The Court clarified that the primary question being asked in the case was not whether the City’s conduct was discriminatory but whether it had a lawful justification for its race based action: “We consider, therefore, whether the purpose to avoid disparate-impact liability excuses what otherwise would be prohibited disparate-treatment discrimination . . . Our task is to provide guidance to employers and courts for situations when these two prohibitions could be in conflict absent a rule to reconcile them” (*Ricci*, p. 20). The Court continued and clarified that “Fear of litigation alone cannot justify an employer’s reliance on race to the detriment of individuals who passed the examinations and qualified for promotions” (*Ricci*, p. 33) and concluded, “[e]ven if respondents were motivated as a subjective matter by a desire to avoid committing disparate-impact discrimination, the record makes clear there is no support for the conclusion that respondents had an objective, strong basis in evidence to find the tests inadequate, with some consequent disparate-impact liability in violation of Title VII” (*Ricci*, p. 26).

Regarding making changes to the tests *after they had been given*—up to and including discarding the results altogether—the Court ruled that a strong basis in evidence had to be somehow demonstrated on the record to justify any such changes. This was not done by the City. Had the City conducted a *Croson Study*⁸ (defined in detail below) for the jobs at issue *prior to the development of the selection processes*, it may have been justified in implementing less extreme race-conscious remedies to identify substantially equally qualified candidates for the lists with less adverse impact. Examples of less intrusive race-conscious remedies include, but are not limited to competency-based cutoff scores, test weightings less likely to adversely impact minorities identified by subject-matter experts to produce substantially equally qualified candidates, and banding based upon the reliability of the tests to group substantially equally qualified candidates together, etc.

In the public sector, rules are often set up by a

“merit system,” negotiation with a union, tradition, or some other reason that can influence the type of selection procedures to be administered and their use. Here are some examples:

- a written test might be specified as required (rather than considering other options such as an assessment center),
- a 70% cutoff score might be required (rather than a competency-based cutoff score),
- certain weightings on tests might be required (rather than using job analysis data or having subject-matter experts establish the weightings based upon their opinions of relevance to the at-issue jobs, perhaps combined with a competency-based cutoff),
- the rule of three might be required (rather than identifying substantially equally qualified candidates based upon the measurements of the tests),
- a rank-ordered list might be required (rather than grouping candidates who are substantially equally qualified together), or
- a rigid banding based upon the same score might be required (rather than grouping substantially equally qualified candidates together).

Different practices, procedures, and tests measure different competencies—each with varying levels of psychometric precision—and such measurements should be developed and used with respect to how they relate to the actual job requirements (more so than just tradition). These tradition-based constraints are not necessarily geared to identify the most qualified candidates based upon competency. The constraints outlined above are typically established so everyone knows the ground rules going into the selection process. The emphasis is not based on competency and finding the most qualified applicants, a system more fair to candidates. If a public entity has any of these constraints, the *Ricci* case presents a blueprint of what is needed to establish a strong basis in evidence before considering race-conscious remedies.

Even more, a competency-based system can often reduce unnecessary adverse impact against groups

protected by Title VII in the first instance. Tests, such as assessment centers, often create less adverse impact than traditional paper-and-pencil tests and can provide very qualified candidates.⁹ Assessment centers can measure important or critical parts of jobs not measurable with paper-and-pencil tests; however, the cost is usually higher, except when the number of candidates is small. Use of competency-based cutoff scores often provides more qualified candidates with less adverse impact than an arbitrary, fixed cutoff score, such as the seemingly arbitrary 70%. Using job analysis data or having subject-matter experts determine weightings to be used by different parts of the selection process often gives greater flexibility to a public entity for obtaining very qualified candidates with less adverse impact than fixed weightings not based on job analysis data or subject-matter expert opinions. Using a banding process to group substantially equally qualified candidates based upon the reliability of the tests (how consistently the test measures what it is supposed to measure) used in the selection process can provide many more competent candidates with less adverse impact than a fixed “rule of three.”

As seen in *Ricci*, any changes made after test administration for the sole purpose of lowering adverse impact, such as altering cutoff scores, alternate use of qualified weightings, etc. may invite challenges. This is especially the case if the changes are outside of the “substantially equally valid” or “substantially equally qualified” doctrines or are not justified on the “strong basis of evidence” standard—see the section below regarding applying the *Ricci* standard to common testing situations. *Ricci* holds that these changes should be made, based upon a strong basis in evidence, prior to the administration of the selection process (hence the need for a Croson Study prior to the implementation of a selection procedure):

Title VII does not prohibit an employer from considering, before administering a test or practice, how to design that test or practice in order to provide a fair

opportunity for all individuals, regardless of their race. And when, during the test-design stage, an employer invites comments to ensure the test is fair, that process can provide a common ground for open discussions toward that end. We hold only 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 (Ricci, pp. 25-26).

The strong-basis-in-evidence standard does not have to be evidence that is provable in court:

“Applying the strong-basis-in-evidence standard to Title VII gives effect to both the disparate-treatment and disparate-impact provisions, allowing violations of one in the name of compliance with the other only in certain, narrow circumstances. The standard leaves ample room for employers’ voluntary compliance efforts, which are essential to the statutory scheme and to Congress’s efforts to eradicate workplace discrimination... And the strong-basis-in-evidence standard appropriately constrains employers’ discretion in making race-based decisions: It limits that discretion to cases in which there is a strong basis in evidence of disparate-impact liability, but it is not so restrictive that it allows employers to act only when there is a provable, actual violation” (Ricci, pp. 23-24).

BACKGROUND OF CROSON STUDIES: ESTABLISHING THE NEED AND JUSTIFICATION FOR TAKING REMEDIAL ACTION

The Supreme Court’s ruling in *Richmond v. Croson* and the related *Wygant v. Jackson Board of Education* case established the basic principle that a governmental employer must provide a strong basis in evidence for the need for remedial action before taking the action. To support an affirmative action program that requires remedial action for minorities under *strict scrutiny*, there must exist a “strong basis in evidence” of past

discrimination by the specific entity to support the conclusion that remedial action is necessary.¹⁰

A generalized assertion that there has been past discrimination in an entire industry will not be enough to justify a race-conscious program under strict scrutiny.¹¹ Applying this rule, however, has produced conflicting results based on various state regulations surrounding affirmative action.¹² *Croson* did not provide guidance as to the amount and type of facts necessary to show a strong basis in evidence that discrimination existed in a particular industry.¹³

The most probative type of evidence in government contracting cases is statistical data showing “gross statistical disparities between the proportion of minorities hired . . . and the proportion of minorities willing and able to do the work.”¹⁴ In government contracting cases, this is often shown through the use of a *disparity index*, which is a comparison between the share of contracts awarded to minority contractors and the percentage of qualified minority-owned firms in the local population that are qualified to do such work.¹⁵ In the employment context, an “availability comparison” or a “utilization analysis” would be a similar comparison. In addition, while the combination of “convincing anecdotal and statistical evidence is potent,”¹⁶ anecdotal evidence, by itself, will rarely suffice to justify a race-conscious remedy evaluated under the strict scrutiny standard.¹⁷ Specific analyses of underutilization for the at-issue jobs, past adverse impact for the practices, procedures, and tests used for the at-issue jobs, analysis of the job relatedness of prior selection procedures used for the at-issue jobs, and evaluation of alternate employment practices is needed. These items make up a *Croson* Study, but are not necessarily facts ready to be proved in court. The strong-basis-in-evidence standard is not so restrictive, according to the Supreme Court, as to require employers to act only when there is a provable, actual violation (*Ricci*, p. 24).

In the *Ricci* case, the Supreme Court stated that the

City of New Haven’s race-based rejection of the test results based only on statistical adverse impact against minorities cannot satisfy the strong-basis-in-evidence standard (*Ricci*, pp. 26–34). The racial adverse impact in this case was acknowledged by the City Attorney and the head of Human Resources. There was no dispute that the City was faced with at least a *prima facie* case of disparate-impact liability, but a *prima facie* case alone was not enough to cancel the eligibility 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—falls short of being a strong basis in evidence that the City would have been 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 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*. And, based on the record the parties developed through discovery, there was “. . . no substantial basis in evidence that the test was deficient in either respect” (*Ricci*, pp. 26–28).

The City made claims that the exams were *not* job related and consistent with business necessity, but failed 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 parts that were not, if any. Of particular interest was the fact that the City Attorney asked the test developers not to provide the validation study contemplated in the contract between the test developer and the City (*Ricci*, p. 6). Several people 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 not having a validation study on the record, the only evidence before the Court demonstrated that detailed *steps* were taken to develop and administer

the tests and there were *painstaking analyses* of the questions on the test to assure their relevance to the Captain and Lieutenant positions. The testimony also showed that complaints that certain examination questions were contradictory or did not specifically apply to firefighting practices in the City were fully addressed, even while the City turned a blind eye to evidence supporting the exams' validity (*Ricci*, pp. 28–29).

A Croson Study would have identified that a validation study was needed to provide the required strong basis in evidence. The validation study would have provided the City with information showing the lack of validation support for the adverse impact created by the minimum qualifications, the 70% arbitrary cutoff score, the weightings negotiated with the union rather than set by job analysis or by subject-matter experts, the use of a rank-ordered list that presented candidates in a way not supported by the psychometric properties of the written test and oral interview, and the certification process of the rule of three, which exceeded the psychometric properties of the tests used to create the artificial groupings of three. As a result, the City was unable to provide a strong basis in evidence that the exams were not job related to support its decision to toss out the test results.

Likewise, the City also made a futile attempt at showing that alternative selection procedures were available with less adverse impact. The same strong basis in evidence requirement was needed to support each of the City's proposed alternate employment selection procedures. Instead, the City asked for an opinion from a testing expert who was not involved in the selection process, and who did not review the selection procedures at issue before providing advice to the City. The City asserted that the use of an assessment center to evaluate candidates' behavior in typical job tasks would have had less adverse impact than written exams, but this assertion was contradicted by other statements in the record from

firefighter Ricci indicating that the City could not have used assessment centers for the exams at issue, because it would have taken too long to develop. A Croson Study would have identified that a testing expert was needed to provide alternate selection procedure options and an expert could have rebutted firefighter Ricci's assertions regarding how long an assessment center takes to develop.

It was inevitable that the City's arguments would fail without a strong basis in evidence. The City referred to testimony that a different composite-score calculation would have allowed the City to consider black candidates for the open positions, but it produced no evidence to show that the weighting actually used was indeed arbitrary, or that different weightings would be a substantially equally valid way to determine whether candidates were qualified for promotions. A Croson Study would have identified that alternate weightings required subject-matter experts. These subject-matter experts would have provided the strong basis in evidence needed to provide an option for the City to consider with substantially equally qualified candidates with less adverse impact.

The City additionally argued that it could have adopted a different interpretation of its charter provision limiting promotions to the highest scoring applicants, and that an alternative interpretation would have produced less discriminatory results; but the Supreme Court said the City's approach would have violated Title VII's prohibition of race-based adjustment of test results. "Here, banding was not a valid alternative for this reason: Had the City *reviewed the exam results and then* adopted banding to make the minority test scores appear higher, it would have violated Title VII's prohibition of adjusting test results on the basis of race" (*Ricci*, p. 31). A Croson Study would have identified banding as an option *before* the selection process to avoid the problem of race-based adjustments of scores after the fact. The City failed to provide a strong basis in evidence that any viable alternate employment practice was available (*Ricci*, pp.

29–33). A Croson Study would have identified for the City the proper mechanisms for creating the strong basis in evidence the City needed.

ESSENTIAL COMPONENTS OF A CROSON STUDY

The Supreme Court in *Croson* and *Wygant* ruled that certain actions taken by governmental entities to remedy past racial discrimination—actions that are themselves based on race—are constitutional only when there is a “strong basis in evidence” that the actions taken for remedial purposes are necessary. A Croson Study is a way to gather this “strong basis in evidence.” In *Ricci*, the Court stated: “. . .the standard appropriately constrains employers’ discretion in making race-based decisions: It limits that discretion to cases in which there is a strong basis in evidence of disparate-impact liability, but it is not so restrictive that it allows employers to act only when there is a provable, actual violation” (*Ricci*, p. 24). A Croson Study is a practical and prudent vehicle for employers to use to minimize the likelihood of losing a Title VII case, or perhaps even being challenged.

A Croson Study is specific to the job classification or rank for which a promotional list is being considered. The Croson Study process includes an evaluation of the utilization of minorities and women as well as the adverse impact and job relatedness of past practices, procedures, and tests used to select candidates during prior selection processes. The job relatedness evaluation includes the assistance of subject-matter experts, as well as outside experts who can provide information for alternative selection procedures. It is not enough to simply review prior statistics of practices, procedures, and tests (*i.e.*, utilization analyses and adverse impact analyses). The job relatedness of the practices, procedures, and tests used must also be addressed and some additional job relatedness work might be necessary. The statistical and job relatedness information is needed to establish a strong basis in evidence required to make temporary remedial or race-

conscious changes. Problems identified from a Croson Study do not have to reach the level required of evidentiary proof in court. However, problems found from a Croson Study *can be* used as a blueprint for some change and as a basis for limited actions to substantially reduce the likelihood of losing a Title VII suit, or in the best case, avoiding the suit altogether. As the Court stated, “Applying the strong-basis-in-evidence standard to Title VII gives effect to both the disparate-treatment and disparate-impact provisions, allowing violations of one in the name of compliance with the other only in certain, narrow circumstances. The standard leaves ample room for employers’ voluntary compliance efforts, which are essential to the statutory scheme and to Congress’s efforts to eradicate workplace discrimination” (*Ricci*, p. 23). Table 2 below summarizes how various levels of justification for using race-conscious remedies for current/future selection processes can be amassed through a Croson Study.

For example, an employer could conduct a Croson Study that would include an evaluation of how test scores were *used* (*e.g.*, ranked, banded, weighted) in a certain selection process for a given position. The purpose of the evaluation would be to determine whether tests scores should be used in a different way for future selection processes, or whether previously-administered tests that are pending certification (as in *Ricci*) should be used in the same way they were previously, or used differently based on the results of the Croson Study.

If the Croson Study determined that the tests were used and weighted in a way that matches how skills are required for the position (*e.g.*, based on a job analysis or opinions from qualified subject-matter experts) and ranking was only used when justified, no strong basis in evidence would be gathered for using the test results in ways that could minimize adverse impact. However, if the tests were *not* used and weighted in a way that matched how skills are required for the position and ranking was conducted on high adverse impact tests with only limited

Table 2: Elements of a Comprehensive Croson Study

Element	Description	Level of Justification Gathered for Using Race-Conscious Remedies for Future/Current Selection Process	
		Low Range	High Range
Recruiting	Applicant outreach, job fairs, job postings, etc.	Good faith outreach toward all subgroups within relevant labor market (both local and regional).	Narrow outreach only to majority rich sources/areas.
Selection Process	Tests used in hiring/promotional process.	Reasonably balanced and job related/validated tests.	Use of high adverse impact tests not related to the job.
Test Use	Pass/fail cutoffs, banding, ranking, or weighting/combining.	Tests are used/weighted in a way that matches how skills are required for the position. Ranking used only when justified.	Ranking on high adverse impact tests with limited justification (e.g., low reliability, bunched scores, not differentiating traits).
Utilization Analysis	Comparison between current workforce composition and qualified labor pool.	No statistically significant imbalances.	Gross imbalances that are statistically and practically significant.
Adverse Impact	Historical adverse impact analyses on tests.	Low/no adverse impact or adverse impact on job related/validated tests.	Long history of unjustified adverse impact on non-validated tests.
Alternate Employment Practices	Tests that are substantially equally valid but have lower adverse impact.	Good faith efforts made to investigate and evaluate alternatives.	Alternatives not investigated and evaluated considered.

Note: "Tests" in Table 2 represent practices, procedures, or tests (either formal or informal) as defined by the Uniform Guidelines 18 (1978).

justification (e.g., low reliability, bunched scores, not differentiating skills), then a strong basis in evidence would be gathered for using the test results in ways that could minimize adverse impact. Obviously, a stronger basis in evidence would be required for using the test results differently than originally presented to applicants (i.e., in a post-exam situation with pending score results) than would be required for changing how the test scores would be used for future selection processes.

WHAT CONSTITUTES A DEFENSIBLE CROSON STUDY?

How defensible is a Croson Study? Ultimately, that depends on the extent to which the race-conscious remedies based on the Croson Study are exclusive versus inclusive. Race-conscious interventions that have the effect of being exclusive towards certain groups at the expense of being inclusive to others will be more highly scrutinized. After the landmark *Adarand Constructors, Inc. v. Peña*¹⁹ decision, a case

where the U.S. Supreme Court held that all racial classifications, imposed by whatever federal, state, or local government actor, must be analyzed by a reviewing court under the “strict scrutiny” standard, the U.S. Department of Justice authored a memorandum to general counsels²⁰ regarding the implications of the case relevant to the “narrowly tailored,” and “strict scrutiny” requirements surrounding race-conscious remedies (such as those that might be justified based on a Croson Study). This memorandum included several evaluation criteria regarding the use of race-conscious remedies which are summarized below:

- Before using race-conscious remedies, consider whether *race-neutral* alternatives could be used instead. This consideration should be obvious, but is nonetheless often overlooked by many well-meaning employers. Both public- and private-sector employers should consider race-neutral measures before resorting to race-conscious action.

- What is the *scope* of the program (*i.e.*, in what *manner* is race used)? Is the race-conscious remedy a hard, fixed number (*e.g.*, a quota)? Is race used as a “plus factor,” or is it included in a static formula? Is it a soft or hard standard, variable or fixed? Quota systems typically require judicial instatement and enforcement, whereas “plus” factors can sometimes be used with less scrutiny. For example, Justice O’Connor’s opinion for the Court in *Croson* criticized the scope of Richmond’s original 30% minority subcontracting requirement as a “rigid numerical quota.” In the Supreme Court case *Gratz v. Bollinger*,²¹ the Court reviewed the University of Michigan’s affirmative action policies that used a strict “race = +20 points” program and ruled that race can be *one of many factors* considered by colleges when selecting their students, but cannot be used in such a strict formulaic manner.

- What statistical evidence exists that substantiates possible discrimination? Evaluating this factor can include two types of statistical evidence: utilization analyses and classical adverse impact evidence. Utilization analyses can be used to evaluate the group’s current representation in the workforce (*e.g.*, their numbers in the at-issue position) to their availability in the relevant labor market (using either the qualified applicant flow data or census EEO data). The two-tail exact binomial test can be used to evaluate whether the group is underutilized, with a statistically significant finding constituting a threshold test²² of possible discrimination. Classic disparate impact evidence (where the success rates of two groups are compared using a statistical test such as the Chi-Square or the two-tail Fisher Exact Test)²³ exists if the practices, procedures, and tests administered in the past have exhibited adverse impact and have not been validated.

- The duration of the program. Race-conscious remedies should only be utilized for a sufficient time period needed to remedy the ill effects of previous discrimination.

- The extent to which the race-conscious remedy has an exclusionary effect on whites or men. The most ideal race-conscious remedy is one that does not unnecessarily trammel the rights of any willing competitor to the process.

Many of these criteria have also been evaluated and upheld when private employers have used race-conscious remedies as part of their Affirmative Action Plans (AAPs) required under Executive Order 11246 (which requires private employers that have contracts with the federal government exceeding \$50,000 to maintain AAPs). For example, in *Sharkey v. Dixie Electric Membership*,²⁴ the employer’s use of race-conscious remedies was challenged in a “reverse discrimination” action. Sharkey, who was white, alleged that Dixie electric engaged in discriminatory hiring practices in violation of Title VII when it

failed to hire him for a vacant position even though he was qualified and instead hired an African American applicant (in part because the at-issue job was identified in Dixie’s AAP as underutilized to a statistically significant degree). Race was only to be used as a “plus” factor if the applicant was otherwise qualified for the position. The Fifth Circuit supported Dixie’s action, ruling that the action was substantially justified by means of a *valid AAP*.

In reviewing the validity of the AAP, the Court noted that the plan 1) was created on an annual basis, 2) properly identified *specific positions that were underutilized*, 3) made accurate comparisons to the relevant labor pool, 4) did not unnecessarily trammel the rights of non-African Americans, and 5) did not present an absolute bar to employment of other groups.

Similar rulings have been made by other Appellate Courts. For example, in *Doe v. Kamehameha School/ Bernice Pauahi Bishop Estate*,²⁵ the Ninth Circuit ruled that private employers’ AAPs: 1) must respond to a manifest imbalance in the work force, 2) must not unnecessarily trammel the rights of members of the non-preferred class or “create an absolute bar to their advancement,” and 3) must do no more than is necessary to attain a balance.

Following the guidelines offered by these two cases can help safeguard both public sector employers and federal contractors who integrate goals into their hiring and promotional processes.

EVALUATING SPECIFIC PRACTICES, PROCEDURES, AND TESTS IN A CROSON STUDY—SOME EXAMPLES

Evaluating weightings in a Croson Study.

Examining the weightings used in prior selection processes to create a ranked list is one practice a Croson Study would evaluate. For example, a public safety department may have used a 60% weighting assigned to the written test based upon a collective bargaining agreement. The Croson Study may

uncover that while the 60% weighting had been negotiated, it had no job relatedness or validation support. The Croson Study might find that the 60% weighting actually caused more adverse impact than lower weightings suggested by subject-matter experts, and that the lower weightings would have provided substantially equally qualified candidates, especially if a job-related competency-based cutoff was used first on the written test before the weighting was applied.

As implied above, one step that could be involved in a Croson Study is to gather data from 7-10 subject-matter experts for a mid-sized department regarding weightings. These subject-matter experts would be provided information on what each test (written and oral) measures and would be asked for independent opinions regarding the weightings for each test that best represent the proper qualification mix necessary for the job. This effort could lead to, for example, a conclusion that the written test weighting should be 30% and be used with a minimum competency cutoff score, and this mix would provide substantially equally qualified candidates.

In the *Ricci* case, however, the City demonstrated that weightings of 30% / 70% would substantially reduce adverse impact but provided no evidence that the scores were substantially equally valid: “Nor does the record contain any evidence that the 30% / 70% weighting would be an equally valid way to determine whether candidates possess the proper mix of job knowledge and situational skills to earn promotions. . . . On this record, there is no basis to conclude that a 30% / 70% weighting was an equally valid alternative the City could have adopted” (*Ricci*, p. 30). Further, the Court ruled that changing the weightings for the sole purpose of reducing adverse impact could likely violate the race-norming prohibition of the 1991 Civil Rights Act: “Changing the weighting formula, moreover, *could well have violated* Title VII’s prohibition of altering test scores on the basis of race. See §2000e–2(f)” [emphasis added] (*Ricci*, p. 30). The “could have” caveat offered by the Court was in no

doubt offered because of the battle that could have been waged (although not in the *Ricci* case) over the suitability of various weighting combinations relative to their comparative degrees of adverse impact.

Clearly, a strong basis in evidence is needed to show that the different weighting options, perhaps with a job-related cutoff, would produce substantially equally qualified candidates. In the *Ricci* opinion, the Supreme Court explains its concerns *before and after* testing:

Nor do we question an employer's affirmative efforts to ensure that all groups have a fair opportunity to apply for promotions and to participate in the process by which promotions will be made. But once that process has been established and employers have made clear their selection criteria, they may not then invalidate the test results, thus upsetting an employee's legitimate expectation not to be judged on the basis of race. Doing so, absent a strong basis in evidence of an impermissible disparate impact, amounts to the sort of racial preference that Congress has disclaimed, §2000e-2(j), and is antithetical to the notion of a workplace where individuals are guaranteed equal opportunity regardless of race.

Title VII does not prohibit an employer from considering, before administering a test or practice, how to design that test or practice in order to provide a fair opportunity for all individuals, regardless of their race [emphasis added] (Ricci, p. 25).

Another option from a Croson Study could be to use subject-matter expert opinions and data from prior exams (either given by that department, similar departments, or reported in the testing literature) with the currently proposed exam to develop weightings that are likely to minimize adverse impact, then set these weightings in advance of administering the tests. Using a competency-based cutoff on the written test could sometimes allow the test to be used as a pass-fail device, rather than a ranking device. Using a job-related competency-based cutoff score on the written test could be used to ensure candidates had a minimum competency, then the rest of the selection process could be used to rank or

band candidates, using factors that differentiate more between marginally successful and higher levels of job performance (ideally, ranking should be based on qualification factors that differentiate job performance, rather than factors that are only needed at a minimum level for successful job performance—see further discussion later).

A Croson Study also evaluates the practice of how the math was done on developing the weightings prior to creating the ranked list. Written test scores and oral interview scores may be normally distributed but they often come from different distributions. If the employer multiplied the raw score from the written test by 60%, and the raw score from the oral interview score by 40%, and then added these two scores together, it is very possible the *effective weights* for the written test could be very different than the *intended weights* (e.g., 70% rather than the 60% advertised to the candidates, needlessly increasing adverse impact). When we combine scores and desire to give them a certain intended weighting, we need to consider each person's score expressed as the deviation of their score from the mean score in each distribution in units of standard deviation. This process is referred to as standard scoring—a practice that has been used in testing since at least 1921.²⁶ When we do not transform each candidate's score into a score in relation to the standard deviation of that distribution, we do not know about the relative performance of the candidate compared to others on this test.

Raw scores on written tests and oral interview tests are usually quite different. A written test may have a mean of 55 with a standard deviation of 15 (i.e., 68% of the scores fall between 40-70; 95% of the scores fall between 25-85; and 99% of the scores fall between 10-100). An oral interview may have a mean of 84 and a standard deviation of 5 (i.e., 68% of the scores fall between 79-89; 95% of the scores fall between 74-94; and 99% of the scores fall between 69-99). Sometimes, oral interviewers will give a candidate a score of 69 as a “failure,” not using the whole distribution of failing scores.

A quick look at the mean raw score for the written test and the mean raw score for the oral interview may detect a difference. If the standard deviations for the written test and oral interview are different, almost certainly the *effective weightings* are not going to be the same as the *intended weightings*. Standard scoring is the process used to ensure the intended weightings become the effective weightings. With standard scoring, the candidate's raw score is subtracted from the mean score then divided by the standard deviation of the test. This conversion puts the candidate's score in perspective with the distribution of scores. After each candidate's score is converted to a standard score, the weightings can be multiplied by the converted standard score to ensure the intended weighting becomes the effective weighting. This process was not completed by the City of New Haven, which resulted in the creation of a final list being issued where 43 of the 77 Lieutenants (56%) were out of their appropriate mathematical ranking (some as many as four positions out of their correct rank order) and 15 of the 41 Captains (37%) also being out of correct rank order.²⁷

Relative Weight Analysis²⁸ can be used to evaluate the effective weightings if raw scores have been used. Re-analyzing the raw score data in the *Ricci* case reveals that the actual effective weightings of the Lieutenant's oral interviews and written tests were 37.2% and 62.8% respectively, and 38.2% for and 61.8% for the Captains oral interviews and written tests, respectively. While only a 2-3% difference from the intended weightings, small differences are of course meaningful to those being tested.

Use of 70% cutoff score in a Croson Study.

Another practice evaluated in a Croson Study is the cutoff score used in prior examinations for the rank. If a 70% cutoff score was used, the basis for its use may be a civil service rule, charter provision, or a provision negotiated with a union. If this cutoff is found to have had adverse impact on past exams for

the rank and no job-relatedness data is available to support it, then the public employer may wish to use these facts to establish a strong basis in evidence and implement a cutoff score based upon job-related competency.²⁹ In footnote 16 of Justice Ginsburg's dissenting opinion in *Ricci* she states:

It appears that the line between a passing and failing score did not accurately differentiate between qualified and unqualified candidates. A number of fire-officer promotional exams have been invalidated on these bases. See, e.g., Guardians Assn., 630 F. 2d, at 105 ("When a cutoff score unrelated to job performance produces disparate racial results, Title VII is violated.") (Ricci, p. 32).

If a Croson Study finds that the 70% cutoff score has created unnecessary adverse impact in the past and is not supported by any job-relatedness or validation data, then a strong basis in evidence is available to the public employer to make a change. The fact that a city charter, merit system, or even state law specifies the use of a 70% rule does not provide adequate justification. The Supreme Court warns against this: "A state court's prohibition of banding, as a matter of municipal law under the charter, may not eliminate banding as a valid alternative under Title VII. See 42 U.S.C. §2000e-7" (*Ricci*, p. 31). Not using a competency-based minimum score can needlessly increase adverse impact. In other situations, using a competency-based cutoff may actually increase adverse impact, although in such situations the adverse impact would be justified because the cutoff score represents the competency level necessary for success.

If an employer desires to determine a job-related, competency-based cutoff score, the steps are straightforward. One such process (known as the modified Angoff technique) can be completed by convening a panel of 7-10 qualified subject-matter experts to rate every item on the test with respect to the percentage of minimally-acceptable candidates who would likely answer the question

correctly. These ratings are then averaged in a “critical score” that is reduced by deducting 1, 2, or 3 Conditional Standard Errors of Measurement (CSEMs) after the test is administered and the reliability becomes known.

The classical Standard Error of Measurement (just “SEM”) is calculated using the standard deviation and reliability of the test. It represents the variance in a score resulting from factors other than the ability being measured by the test. The concept is based on the premise that the underlying ability being measured by the test cannot be measured exactly without a perfectly reliable test. Factors cause the candidates taking the test to score higher or lower than their true score due to chance error, differential testing conditions, imperfect test measurement, and other factors unrelated to the applicant’s true ability. The SEM provides an estimate of the average test score error for all candidates regardless of their ability level. The SEM assumes that errors are the same at every score throughout the distribution of candidate scores and represents a weighted average of errors associated with each score.

Because the SEM varies across the range of candidate proficiencies, and individual candidate score levels on any specific test could have different degrees of measurement error, the SEM as well as individual score level estimates, commonly referred to as *Conditional* Standard Errors of Measurement (CSEM), are typically reported.³⁰ The CSEM gives the estimate of reliability (error estimate) at each score point. The SEM is considered a test-level statistic and the CSEM is considered a score-level statistic. Both can be used to compute a confidence band around an observed score to determine a score range in which the *true score* (i.e., a score absent measurement error) probably lies. Unlike the SEM, the CSEM takes the variation in measurement accuracy across the score scale into consideration. That is why the CSEM can offer a more precise error band around any one individual score.

With the Supreme Court’s warning about section 2000e-2(l) of Title VII, which specifies “It shall be an unlawful employment practice for a respondent, in connection with the selection or referral of applicants or candidates for employment or promotion, to adjust the scores of . . . or otherwise alter the results of, employment related tests on the basis of race, color, religion, sex, or national origin,” it might be prudent to predetermine whether to use 1, 2, or 3 CSEMs to adjust the critical score. To properly determine the CSEMs, the test must first be administered, but then the predetermined number of CSEMs can be deducted to arrive at the final cutoff for the test (there are, however, some methods for determining the CSEM prior to administration—see later discussion).

While predetermining the number of CSEMs to use for the final adjustment may lower risk, determining the number to use after the test is a risk if the decision is based solely upon race. Lowering a validated cutoff score by 1, 2, or 3 CSEMs to reduce adverse impact (as well as factoring in other relevant factors) is merely accounting for the lack of reliability of the test (using 1, 2, or 3 CSEMs as confidence intervals for including true scores); it does not *alter the results or use different cutoff scores* for different groups, but merely *uses the existing results in a different way*. It is unclear how a conservative Supreme Court would view using race as one factor. Therefore, while it would be safer to make the decision prior to test administration, there are some advantages as well as risks associated with waiting until after test administration to decide on the number of CSEMs to use.

There are several case examples where reducing cutoff scores within substantially equally qualified ranges to reduce adverse impact has been supported. Perhaps the first example is given in the Supreme Court case, *U.S. v. South Carolina*.³¹ In the *South Carolina* case, the Court permitted the consideration of five statistical and human factors when choosing whether to deduct 1, 2, or 3 SEMs:

- The size of the SEM. Large SEMs indicate low test reliability and/or high levels of variance in the applicant pool;
- The possibility of sampling error in the study (this relates to the number of subject-matter experts who served on the cutoff development panel). Panels with only a few subject-matter experts raise concern based on this factor, especially if there are a large number of incumbents in the workforce);
- The consistency of the results (internal comparisons of the panel results). Panels that included biased subject-matter experts raise concern here (only if they were not removed based on inter-rated reliability and/or being extreme outliers);
- The supply and demand for incumbents in the target position (pertaining to the demand for workers needed in the work force); and
- The racial composition of the workforce/levels of adverse impact on each of the cutoff options should be considered.

Deducting SEMs (or preferably, CSEMs) in this fashion preemptively addresses concerns or attacks that plaintiff groups may bring by arguing that lower cutoff scores are within the range of the “substantially equally valid” requirement of the Uniform Guidelines (Section 3B) and the third burden (“alternate employment practices”) discussed in the 1991 Civil Rights Act. Without deducting CSEMs from validated critical scores, employers could find themselves wide open to *alternate employment practice* disparate impact discrimination arguments brought by enforcement agencies or plaintiff groups. Further, if the critical score has statistically significant adverse impact and 1 CSEM below does not, the employer is faced with a situation where two cutoff options that are substantially equally valid result in very different liability outcomes (one is enforceable under Title VII, the other is not).

A more recent example occurred in *Isabel v. City of*

Memphis (2005),³² where the City of Memphis used a written test for the sergeant promotional process. The City had negotiated a 70% cutoff for the test which had disparate impact against blacks, and the Court ruled that the cutoff score was invalid, stating: “To validate a cutoff score, the inference must be drawn that the cutoff score measures minimal qualifications . . .” The district court found that the cutoff score was “nothing more than an arbitrary decision and did not measure minimal qualifications.” Given this context, the Court supported lowering the (arbitrary) 70% cutoff score by 4% to reduce adverse impact.

Use of score lists in rank order in a Croson Study.

Another area investigated in a Croson Study is how the prior lists have been used. If the lists have been used in strict rank order, then an investigation is made using a distribution type of adverse impact analysis to see if the rank order process itself adversely impacted any groups protected by Title VII. If so, then the investigation continues to the job-relatedness support to see the reasons for using the list in rank order. It is not enough to provide job-relatedness data using content validity to support a rank-ordered list with adverse impact. The employer must provide evidence that the test measures those *aspects of performance which differentiate among levels of job performance*. The Uniform Guidelines (1978, Section 14C9) state:

Where a selection procedure supported solely or primarily by content validity is used to rank job candidates, the selection procedure should measure those aspects of performance which differentiate among levels of job performance. [emphasis added].

This requirement has been affirmed by numerous courts. For example, in *Vulcan Pioneers v. NJ Department of Civil Service*,³⁴ the Court ruled that the tests were not appropriate for ranking because “subject-matter experts were not even asked whether the knowledges, skills, and abilities were appropriate for ranking; rather, they were asked to indicate whether the knowledges, skills, and abilities ought to

be ‘qualifying’ . . . several knowledges, skills, and abilities, such as knot-tying or first aid, appear to be abilities common to all experienced firefighters, and as to which only a insignificant range of performance ought to have been possible.” The Court also noted that the reliability of the examinations was poor and that the test placed a premium on test-taking ability rather than the relevant knowledges, skills, and abilities required for the job.

An additional requirement surrounding ranking that has been applied in litigation is the extent to which the scores are adequately dispersed around the range of interest. For example, in *Guardians v. Civil Service Commission of New York*,³⁵ the Court ruled that the scores were “too bunched” to be used in rank order. The Court remarked:

Rank-ordering satisfies a felt need for objectivity, but it does not necessarily select better job performers. In some circumstances the virtues of objectivity may justify the inherent artificiality of the substantively deficient distinctions being made. But when test scores have a disparate racial impact, an employer violates Title VII if he uses them in ways that lack significant relationship to job performance.

The Court ruled that rank ordering was only permissible when a demonstration could be made of “such substantial test validity that it is reasonable to expect one- or two-point differences in scores to reflect differences in job performance.” Even though the Court approved the validity of the test itself, it viewed the evidence required to use the test as a separate and distinct issue: “Our prior conclusion that the test itself may have had enough validity to be used does not, therefore, lead to approval of using its results for rank-ordered selections.”

In *Ricci*, the differentiating evidence required to demonstrate the job relatedness of the rank-ordered lists was not addressed by the Supreme Court majority.³⁶ The plaintiffs did not attack this process because the top scorers on the lists *were* the plaintiffs. Not surprisingly, the City did not criticize its own

rank-ordered process, though the validation report would have, had the City asked for it!

When lists are used in a strict rank-ordered manner, they can needlessly increase adverse impact. The reliability of tests and consistency of measurement used in personnel selection is not perfect. Anyone who has taken a typing test knows that you may score 45 words a minute, then, five minutes later, re-take the test and score 43 words a minute or 48 words a minute. You are unlikely to score exactly 45 words a minute again. However, if we repeat this process with many candidates, it is possible to find the reliability of the test. There are others ways to find reliability, but reporting the reliability of tests is a requirement of the Uniform Guidelines, whenever it is feasible.³⁷

Test scoring programs typically produce different types of reliability calculations. The reliability can be used to establish the reasonable range within which candidates are likely to obtain their true score. This range reflects the error of measurement for the test and can be used to group scores so that substantially equally qualified candidates are banded together based upon the test(s) used. See the next section for a discussion of the common mechanics used for the banding process.

A Croson Study could evaluate the distribution of scores of prior exams to see if the rank-order process contributed to adverse impact. If so, the Croson Study would evaluate the justification for the use of the test in a rank-ordered way to see if there is differentiating job relatedness support that adequately addresses the Uniform Guidelines requirements. Also, the Croson Study would evaluate whether the reliability of the tests has been addressed in the manner the rank-ordered list had been used. If there are problems in this area, the Croson Study can be used to establish the strong basis in evidence needed to make changes to help the employer reduce unnecessary adverse impact and more properly use these selection tools, relying on the reliability of the test to group scores of substantially equally qualified candidates.

It is important that the decision to band is set up *prior* to the next test administration. Changes in the scoring process after test administration may place the public entity at a major risk. The Supreme Court has stated that banding would not be an option after the fact (if it was used solely to reduce adverse impact), but should rather be decided in advance as a practice the employer will adopt: “Had the City reviewed the exam results and then adopted banding to make the minority test scores appear higher, it would have violated Title VII’s prohibition of adjusting test results on the basis of race. §2000e–2(l)” (*Ricci*, p. 31). If adverse impact is shown with the rank-ordered process after test administration, adequately addressing the differentiating validity requirements is the only hope of success in court.

Rule of three in a Croson Study.

Another practice that could be evaluated in a Croson Study is the practice of how top candidates are certified to an appointing authority for final appointment. The appointment certification procedure called the “rule of three” used by a public employer may adversely impact groups protected by Title VII. Any practice, procedure, or test that causes adverse impact on a group protected by Title VII in a hiring or promotion process is subject to a disparate-impact discrimination challenge. If an adverse impact analysis of the rank-ordered distribution of scores indicates adverse impact throughout the appointing process, the chances are very good that the rule of three will itself have adverse impact (as it is used to select three people at a time from the rank-ordered lists). If the Croson Study finds that the practice of the rule of three has adversely impacted a group or groups protected by Title VII, then the job relatedness of the rule of three needs to be demonstrated. This will be hard to do when the psychometrics of a written test and oral interview will *almost never* be reliable enough to justify strict rank order in groups of just three candidates.

For example, in the *Ricci* case, using the known standard deviations of each test—typical reliability estimates of .90 for the written test (internal consistency) and .60 for the interview (inter-rater reliability)—and the actual correlations between these two tests ($r = .35$ and $r = .40$ for the Lieutenant and Captain process, respectively), the composite reliability of the two combined measures can be estimated ($r = .84$ and $r = .85$ for the Lieutenant and Captain process, respectively). This value can be used with the standard deviation of scores on each list to compute a Standard Error of Difference (SED) that can be used to band applicants that are deemed *substantially equally qualified*. Multiplying the SED in this example by 1.96 provides a 95% confidence interval that applicants who are within a 9.12 point spread (for the Lieutenant list) and 9.06 point spread (for the Captain list) possess scores that are statistically indistinguishable. In other words, score differences within the broader extremes of this range can be considered the product of measurement error with these tests, rather than reliably different ability levels between the candidates. With these ranges of substantially equally qualified candidates, there might be as few as two substantially equally qualified candidates for the appointing authority to consider or as many as nine or more. It is unlikely that there will always be three. The reliability of the test is used to establish the measurement range of the test. The number of substantially equally qualified candidates will vary throughout the range of scores. But that’s what a job-related selection procedure should be doing—basing the grouping of candidates *according to how their competency levels may reliably differ*.

Remember that the Supreme Court has stated that banding would not be an option after the fact if the sole reason for doing so is race-based. The decision to band needs to be made before test administration begins. “Had the City reviewed the exam results and then adopted banding to make the minority test

scores appear higher, it would have violated Title VII's prohibition of adjusting test results on the basis of race. §2000e-2(l)" (*Ricci*, p. 31). If a public entity performs a Croson Study, evaluates the rule of three as a part of that study, and finds adverse impact in the past, it is likely to happen again.

Changing to a banding process *after test administration* will probably be too late, according to the Supreme Court, if the reason is race-based. Therefore, stating *before* test administration, that banding will be used based upon the reliability of the tests involved in making the list, rather than the rule of three, is making adjustments on the basis of the test reliability, not race, and is a more prudent decision than waiting for a case to be filed after testing is done. Banding should be done based upon the measurement properties of the tests involved and not based upon an artificial number of three candidates at a time. The Croson Study can help act as the change agent to avoid disparate-impact discrimination liability because, once adverse impact is shown on a current list with the rule of three, it is probably too late (see discussion on this specific topic below).

An inherent conflict exists between a Civil Service requirement that the selection process be job related and a requirement to use a rule of three. In the 1971 *Griggs v. Duke Power*³⁸ case, we learned from the Supreme Court that the rule of three is a practice that can violate Title VII: "Under the Act, practices, procedures, or tests neutral on their face, and even neutral in terms of intent, cannot be maintained if they operate to 'freeze' the status quo of prior discriminatory employment practices."

In *Ricci*, the city's charter established a merit system. The merit system required the use of job-related exams:

When the City of New Haven undertook to fill vacant Lieutenant and Captain positions in its fire department (Department), the promotion and hiring process was governed by the city charter, in addition to federal and

state law. The charter establishes a merit system. That system requires the City to fill vacancies in the classified civil service ranks with the most qualified individuals, as determined by job-related examinations [emphasis added] (Ricci, p. 3).

The *Griggs* decision warns that "examinations" must be interpreted as "practices, procedures, or tests." The City's merit system states the requirement to use the rule of three:

After each examination, the New Haven Civil Service Board (CSB) certifies a ranked list of applicants who passed the test. Under the charter's "rule of three," the relevant hiring authority must fill each vacancy by choosing one candidate from the top three scorers on the list (Ricci, p. 3).

Perhaps the conflict between being required to use job-related practices and also being required to use the rule of three has been anticipated when the Supreme Court in *Ricci* issued a clear warning to the public sector that relying upon a state court's prohibition of banding may not be enough: "A state court's prohibition of banding, as a matter of municipal law under the charter, may not eliminate banding as a valid alternative under Title VII. See 42 U. S. C. §2000e-7" (*Ricci*, p. 31). Title VII states in part:

Nothing in this subchapter shall be deemed to exempt or relieve any person from any liability, duty, penalty, or punishment provided by any present or future law of any State or political subdivision of a State, other than any such law which purports to require or permit the doing of any act which would be an unlawful employment practice under this subchapter. (42 U.S.C. §2000e-7).

A Croson Study can gather the data regarding adverse impact of the rule of three and crystallize the issue for public sector management to address before being required to by a court. With a Croson Study completed, more options are available to the public employer to avoid disparate impact in the first instance.

HAS RICCI CREATED A CONFLICT BETWEEN SECTIONS 2000E-2(L) AND 2000E-2(K)(1)(A)(II) OF THE 1991 CIVIL RIGHTS ACT?

As a result of *Ricci*, two sections of Title VII may now stand in conflict. Section 2000e-2(l)³⁹ is the section dealing with the “Prohibition of discriminatory use of test scores.” Section 2000e-2(k)(1)(A)(ii)⁴⁰ deals with the “Burden of proof in disparate impact cases,” specifically the “alternate employment practice” provision. Simply reading these two sections can give rise to some apparent conflicts. However, reading the Uniform Guidelines’ interpretation of the alternate employment practices application (see section 60.3B)⁴¹ confirms a potential conflict and raises questions among practitioners regarding how these two sections can be reconciled.

Upon close examination, there is a way to minimize the conflict created by *Ricci*, which added substantial constraints to the interpretation published by the Equal Employment Opportunity Commission, the Department of Justice, the Civil Service Commission and the Department of Labor in section 3B of the Uniform Guidelines on Employee Selection Procedures (1978).

In *Ricci*, the Supreme Court states regarding section 2000e-2(l): “If an employer cannot rescore a test based on the candidates’ race, §2000e-2(l), then it follows *a fortiori* that it may not take the greater step of discarding the test altogether to achieve a more desirable racial distribution of promotion-eligible candidates—absent a strong basis in evidence that the test was deficient and that discarding the results is necessary to avoid violating the disparate impact provision” (*Ricci*, p. 24). When addressing the proposed alternate employment practice of changing the weightings from 60% on the written test and 40% on the oral interview to 30% / 70%, the *Ricci* majority stated: “Nor does the record contain any evidence that the 30% / 70% weighting would be an equally valid way to determine whether candidates

possess the proper mix of job knowledge and situational skills to earn promotions. Changing the weighting formula, moreover, *could* well have violated Title VII’s prohibition of altering test scores on the basis of race. See §2000e-2(l). *On this record*, there is no basis to conclude that a 30% / 70% weighting was an equally valid alternative the City could have adopted” [emphasis added] (*Ricci*, p. 30).

In the *Ricci* decision, the Supreme Court addressed section 2000e-2(l) again dealing with the alternative employment practice of banding: “Here, banding was not a valid alternative for this reason: Had the City reviewed the exam results *and then* adopted banding to make the minority test scores appear higher, it would have violated Title VII’s prohibition of adjusting test results on the basis of race. §2000e-2(l)...” [emphasis added] (*Ricci*, p. 31).

The alternative employment practice provisions of Title VII in the Civil Rights Act of 1991, gives a specific day for an important reason. It states in section 2000e-2(k)(1)(A)(ii): “the complaining party makes the demonstration described in subparagraph (C) with respect to an alternative employment practice and the respondent refuses to adopt such alternative employment practice.” The (C) section states: “The demonstration referred to by subparagraph (A)(ii) shall be in accordance with the law as it existed on June 4, 1989, with respect to the concept of ‘alternative employment practice’” [emphasis added]. On June 5, 1989, *Wards Cove v. Atonio*⁴² was decided by the Supreme Court in a manner that was objectionable to a majority of Congress. The objectionable part of the *Wards Cove* decision states:

Finally, if on remand the case reaches this point, and respondents cannot persuade the trier of fact on the question of petitioners’ business necessity defense, respondents may still be able to prevail. To do so, respondents will have to persuade the factfinder that “other tests or selection devices, without a similarly undesirable racial effect, would also serve the employer’s legitimate [hiring] interest[s]”; by so demonstrating,

respondents would prove that “[petitioners were] using [their] tests merely as a ‘pretext’ for discrimination.” *Albemarle Paper Co.*, *supra*, at 425; see also *Watson*, 487 U.S., at 998 (O’Connor, J.); *id.*, at 1005-1006 (Blackmun, J., concurring in part and concurring in judgment). If respondents, having established a *prima facie* case, come forward with alternatives to petitioners’ hiring practices that [490 U.S. 642, 661] reduce the racially disparate impact of practices currently being used, and petitioners refuse to adopt these alternatives, such a refusal would belie a claim by petitioners that their incumbent practices are being employed for nondiscriminatory reasons.

Of course, any alternative practices which respondents offer up in this respect must be equally effective as petitioners’ chosen hiring procedures in achieving petitioners’ legitimate employment goals. Moreover, “[f]actors such as the cost or other burdens of proposed alternative selection devices are relevant in determining whether they would be equally as effective as the challenged practice in serving the employer’s legitimate business goals.” *Watson*, *supra*, at 998 (O’Connor, J.). “Courts are generally less competent than employers to restructure business practices,” *Furnco Construction Corp. v. Waters*, 438 U.S. 567, 578 (1978); consequently, the judiciary should proceed with care before mandating that an employer must adopt a plaintiff’s alternative selection or hiring practice in response to a Title VII suit. [emphasis added]

The specific words regarding alternate employment practices from *Albemarle* (prior to June 4, 1989) were: “. . . the respondents have not until today been specifically apprised of their opportunity to present evidence that even validated tests might be a “pretext” for discrimination in light of alternative selection procedures available to the Company.”⁴³ The *Watson* decision (prior to June 4, 1989) stated the following regarding alternate employment practices: “Factors such as the cost or other burdens of proposed alternative selection devices are relevant in determining

whether they would be equally as effective as the challenged practice in serving the employer’s legitimate business goals. The same factors would also be relevant in determining whether the challenged practice has operated as the functional equivalent of a pretext for discriminatory treatment.”⁴⁴

While Congress obviously wanted to place less constraints on how alternate employment practices had been interpreted by the Supreme Court, the 1991 Civil Rights Act could only do so much by reversing any impact the *Wards Cove* case could have. The Supreme Court’s narrowing to employment goals from business goals was erased. The warning that “the judiciary should proceed with care before mandating that an employer must adopt a plaintiff’s alternative selection or hiring practice in response to a Title VII suit” was clearly reversed by the 1991 Civil Rights Act.

APPLYING THE RICCI STANDARD TO COMMON TESTING SITUATIONS

What are the implications for employers of the Supreme Court’s analysis of the test-related evidence that was available in the record in *Ricci*? 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 against minorities and having whites sue on the basis of disparate treatment). The classic Title VII burdens established under the USSC’s unanimous ruling in *Griggs* 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*,⁴⁵ the court clarified that the *Ricci* case does not change the law as it relates to the burden-shifting requirements outlined in Title VII. In addition, Federal Enforcement Agencies such as the Office of Federal Contract Compliance Programs (OFCCP) have clarified that *Ricci* does not alter the standard enforcement processes that are in place when federal contractors use testing practices that exhibit adverse impact.⁴⁶

While the major tenets 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 it can be proven, using the strong-basis-in-evidence standard, that the test was likely not valid or that other substantially equally valid (lower adverse impact) options were available but were overlooked.)

Considering 1) the *Ricci* admonishment regarding changes after a selection process has been administered absent a “strong basis in evidence” that not acting will cause a disparate impact situation, 2) the “Prohibition of discriminatory use of test scores” in Title VII’s section 2000e-2(l), and 3) the alternate employment practices in Title VII’s 2000e-2(k)(1)(A)(ii) (and related doctrine codified in the Guidelines, Sections 3B, 5G, 14B[5] and [6], and 14C[8] and [9] with respect to “alternate use with less adverse impact”), below is a list of practical examples of how employers can address real employment situations. Each example will be discussed in the context of the employers in situations with and without *a strong basis in evidence* (*i.e.*, evidence that the test(s) used were not sufficiently valid) as well as *before or after* the test administration.

Changing weightings. For purposes of this discussion, the process of changing weightings means changing from union-negotiated or arbitrary weightings to a set of job-related weightings based on job analysis data or ratings from a panel of subject-matter experts. Because developing test weightings based upon job research results in weightings that accurately reflect job requirements, such a process is likely to produce a more qualified applicant list than union-negotiated or arbitrary weightings, and may produce more or less adverse impact.

Changing weightings before test administration with or without a strong basis in evidence. This practice is acceptable, even if adverse impact was a motivating factor behind starting the process to research and identify the actual job-related weightings (based on both the “alternate employment

practice” requirement of the 1991 Civil Rights Act and the related “substantially equally valid” doctrine of the Guidelines). For example, an employer, in an attempt to both maximize the validity of the process and reduce undesirable adverse impact, may conduct research and determine that assessment centers are an effective means at reducing adverse impact while measuring competencies that are typically untapped by written tests, and decide to include an assessment center as part of the overall weighted testing process. Another example would be an employer that uses both a written test and an assessment center (weighted arbitrarily 50% / 50%), and decides to conduct a study to re-weigh the tests according to their *relative importance* to the job. Yet another example would be using the written test with a validated cutoff score as a pass/fail device only (especially if it only measures abilities needed at a baseline level, rather than differentiating abilities), and then weighting the assessment center 100% (especially if it measured differentiating abilities).

Changing weightings after test administration without a strong basis in evidence. This practice would not be defensible *post-Ricci*.

Changing weightings after test administration with a strong basis in evidence. The Supreme Court ruled in *Ricci* that “Changing the weighting formula, moreover, *could well have violated* Title VII’s prohibition of altering test scores on the basis of race. See §2000e-2(l)” [emphasis added] (*Ricci*, p. 30). The Court’s key caveat to this statement was, “*On this record*, there is no basis to conclude that a 30% / 70% weighting was an equally valid alternative the City could have adopted” [emphasis added] (*Ricci*, p. 30). If, however, the City presented a set of weightings (based on job analysis data or qualified subject-matter expert opinions) that were substantially equally valid—or perhaps even *more valid*—than the original set used by the City, but reduced adverse impact, they might have been accepted. For example, if the City used job analysis data or subject-matter expert opinions to determine that a 50% / 50% weighting

scheme was substantially equally valid (or even more valid) than the 60% / 40% weighting that was used, and the job related set reduced adverse impact, such a weighting scheme could have been adopted. However, merely suggesting alternative weightings *without also demonstrating* that they were substantially equally valid to sufficiently address the “alternate employment practice” requirement of the 1991 Civil Rights Act and the related “substantially equally valid” doctrine of the Guidelines would obviously not be acceptable, as they were rejected in the *Ricci* case.

Lowering cutoffs. Because cutoffs split the entire applicant group into two groups—passing and failing—they should be set in a way that maximizes the differentiation between the qualified versus unqualified applicants. The most effective cutoff scores have the highest levels of Decision Consistency Reliability—a type of reliability that evaluates how consistently the test classifies “qualified” and “non-qualified,” or those who pass the test versus those who fail.⁴⁷ Cutoffs that are set too high will screen out too many applicants who are actually qualified. Cutoffs that are set too low have the opposite effect—they will screen in too many unqualified applicants. Cutoffs that are accurately set using a job-related process like the Angoff method are likely to have higher levels of Decision Consistency Reliability than arbitrary cutoffs. Even if a cutoff is set using a job-related method like the Angoff technique, the test used to measure applicant qualification levels along the continuum of scores will have a certain degree of score consistency (reliability) with respect to measuring the same. As discussed above, accounting for this lack of perfect measurement can be done by reducing the critical score by 1-3 CSEMs. Also discussed below is the possibility of lowering cutoffs without respect to the consideration of such measurement error.

Lowering cutoffs before test administration without a strong basis in evidence. Although this example implies lowering cutoffs in a way that is not

based on the psychometric properties of the test (*e.g.*, without considering the reliability or CSEM), it is likely defensible because it was done before the administration of the test. However, lowering cutoffs without respect to the psychometric properties of the test (or the ability continuum inherent within the score distribution) may possibly result in setting a standard that is below the qualification levels needed for the job. It is recommended to use the psychometric properties of the test and a job-related cutoff procedure.

Lowering cutoffs before test administration with a strong basis in evidence. If an employer determines a job-related cutoff before the test is given and desires to account for the measurement error of the test, the employer can defensibly lower the cutoff using CSEMs and determine the exact number (either 1, 2, or 3) beforehand. Most CSEM methods require post-administration statistics for computation. However, the Lord-Keats method can be used for estimating CSEMs before a test has been administered, although it is more advisable to simply state up front whether 1, 2, or 3 CSEMs will be deducted and then use the most accurate CSEM computations after test administration.

Lowering cutoffs after test administration with a strong basis in evidence. If an employer administers a test with an arbitrary cutoff, and subsequently determines that neither the test nor the cutoff score are sufficiently valid (establishing a strong basis in evidence), lowering the cutoff score to reduce or eliminate adverse impact may be permissible. In fact, the Court ruled in *Ricci* that having a strong basis in evidence may even justify a *more severe* action—retracting the test results altogether. Lowering the cutoff score with consideration of the psychometric properties of the test would be helpful to support this situation, as would knowing the test score level associated with the minimum levels needed for the job (*e.g.*, a cutoff score set using the modified Angoff technique).

Lowering cutoffs after test administration without a strong basis in evidence. Provided that the employer lowers the cutoff based upon the psychometric limitations of the test, lowering the cutoff by 1-3 CSEMs may be permissible based on both the “alternate employment practice” requirement

of the 1991 Civil Rights Act and the related “substantially equally valid” doctrine of the Guidelines. It should be noted here that the race-norming prohibition of the 1991 Civil Rights Act (§2000e-2[I]) expressly states that *using different cutoff scores* for different groups is prohibited; however, this should be

Table 3: The Strong-Basis-in-Evidence Requirement Applied to Several Testing Scenarios

Practice	Strong-Basis-in-Evidence Requirement (i.e., evidence of non-validity and/or no Alternate Employment Practice)			
	No Risk	Some Risk	Moderate Risk	High Risk
Changing weights.	Before administration.	After administration based on correct math (e.g., using standard scores v. raw scores to weight tests). ⁽¹⁾	After administration based on updated/ corrected job analysis data or SME panel opinions. ⁽²⁾	After administration changing weightings without job research just to reduce adverse impact ⁽³⁾
Lowering cutoffs.	Deciding to lower a job-related cutoff using 1-3 CSEMs before or after administration. ⁽⁴⁾	Deciding to lower an arbitrary cutoff before testing (not related to using 1-3 CSEMs).	Lowering an arbitrary cutoff score after administration just to reduce adverse impact.	
Banding.	Deciding to band before test administration.	Banding based on arbitrary or weak statistical methods	Widening pre-determined bands or deciding to band after administration just to reduce adverse impact. ⁽⁵⁾	
Taking part or whole test actions or changes after identifying adverse impact.	Considering future alternate employment practices with lower adverse impact, making changes, or validating for subsequent administrations.	Removing test items based on qualified Differential Item Functioning (DIF) analyses. ⁽⁶⁾	Retracting or not certifying test results.	

Notes: (1) Changing from raw scores to standard scores will typically have an effect of changing the effective weightings. (2) Using accurate and reliable job analysis data and/or input from a panel of qualified subject-matter experts (such a change could result in higher or lower adverse impact). (3) Note that the Court in *Ricci* ruled that changing the weights for the sole purpose of reducing adverse impact could likely violate the race-norming prohibition of the 1991 Civil Rights Act (*Ricci*, p. 30). Given the “could” caveat offered by the court, it is safe to say that weights that were not “substantially equally valid” would mostly likely constitute a violation; whereas weights within a substantially equally valid range “may” cause a violation. (4) This process is supported and based upon the psychometric characteristics of the test. (5) In the *Ricci* decision, the Supreme Court addressed section 2000e-2(l) again dealing with the alternative employment practice of banding: “Here, banding was not a valid alternative for this reason: Had the City reviewed the exam results and then adopted banding to make the minority test scores appear higher, it would have violated Title VII’s prohibition of adjusting test results on the basis of race. §2000e-2(l)...” [emphasis added] (*Ricci*, p. 31). (6) Removing test items based on DIF studies is a defensible practice provided that the levels of DIF are high and the psychometric and validity properties of the items removed are limited (see D. Biddle, 2006, pp. 68-72).

differentiated from lowering a cutoff score, which is applied to *all* test takers in the distribution, within a range that is within the substantially equally valid confines. If the employer arbitrarily lowers the cutoff for race reasons without a strong basis in evidence, this practice would not be supported post *Ricci*.

Banding. As discussed above, test score banding is widely practiced in the testing field and has been endorsed far more times than denied in litigation settings.⁴⁹ However, if banding is adopted after a test has been administered for the sole purpose of reducing adverse impact, the foundations set forth in *Ricci* apply as follows:

Banding before test administration with or without a strong basis in evidence. Before a test is administered, employers can predetermine the use of banding, even though the span of the bands can only be determined after the test(s) have been administered (because the psychometric properties of the tests cannot be known until afterwards).

Banding after test administration without a strong basis in evidence. This practice is not permissible after *Ricci*.

Banding after test administration with a strong basis in evidence. Provided that this process is based upon the psychometric properties of the test(s) involved, banding using 1-3 Standard Errors of Difference⁵⁰ may be permissible based on both the “alternate employment practice” requirement of the 1991 Civil Rights Act and the related “substantially equally valid” doctrine of the Guidelines.

In sum, Table 3 provides a summary of each of the above recommendations for employer compliance in a post-*Ricci* context.

APPLYING RICCI TO OTHER TITLE VII SITUATIONS (OUTSIDE OF TESTING)

Within a few short weeks after the Supreme Court’s ruling in *Ricci*, the human resources and legal field began exploring the decision’s potential to reach

situations outside of the typical disparate impact/disparate treatment framework. Some stretched the *Ricci* precedent to mean that meeting the strong-basis-in-evidence rule prevented employers from taking *any* actions based on their adverse impact findings—sending messages to the EEO compliance community that proactive adverse impact analyses were either moot (at best) or impermissible (at worst). Others offered cautionary guidance, implying that HR and legal practitioners should tread lightly when it comes to applying disparate impact theory to transactions such as layoffs or proactive compensation analyses.⁵¹ Some courts have even felt the need to clarify the fact that *Ricci* does not change the standard Title VII disparate impact theory applied in testing cases.⁵²

When evaluating the span of *Ricci*, it is important to recognize the context under which the case was tried. In *Ricci*, the strong-basis-in-evidence standard was applied within a set of extreme circumstances—candidates who had competed in a public-sector promotional process (where the competitive ground rules were laid down *before the process began*) were faced with losing their promotional opportunities *after* being previously-deemed qualified. In addition, the context was the public sector; it dealt with promotional processes that were judged by the Court as acceptable (albeit based on a very limited evidence record); and it weighed the permissibility of redacting promotional eligibility based on publically announced criteria. Given these circumstances, it is safe to say that the facts in *Ricci* sets the case far to one side of the continuum of HR practices that might be subjected to Title VII scrutiny. Table 3 above discusses the different testing practices that may fall along this same continuum.

The extent to which the *Ricci* precedent may apply to other (non-testing) situations is explored below. The discussion is limited to adverse impact pertaining to layoff (Reduction in Force, or “RIF”) decisions and compensation analyses because these are the two

most common situations where disparate impact theory is applied outside of the testing context.

Applying Ricci to Reduction in Force (RIF) decisions.

The primary practical implication from *Ricci* is that if employers are using practices, procedures, and/or tests that are fair and valid, they are likely on safe ground—even when such practices, procedures, and/or test have adverse impact, unless an alternate employment practice that is substantially equally valid with less adverse impact has been presented to the employer and the employer rejected it. The *Ricci* ruling squarely stated that 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 undergo a RIF process, they ought be certain 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 disparate impact 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 by 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 in such a RIF case could prevail under the “alternate employment practice” theory (*i.e.*, by successfully arguing RIF practices or criteria that were “substantially equally valid,” but had

less adverse impact were available to the employer and the employer refused to use the alternate employment practice(s)).

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 Supreme Court has ruled in *Smith v. City of Jackson* that there are important textual differences between the Age Discrimination in Employment Act (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 requirements under Title VII. 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 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 uses unvalidated criteria, they leave themselves vulnerable in a Title VII lawsuit.

There are a few differences that specifically apply when comparing RIFs to the redacted promotional process that occurred in *Ricci*. *Ricci* dealt with a promotional process where all of the candidates knew the promotional process and criteria in *advance* and then competed within this pre-defined framework to earn their slots on a promotion slate. By competing in such a process, the employees developed certain “legitimate expectations” that were upset by the redacting of the eligibility list:

“Examinations like those administered by the City create legitimate expectations on the part of those who took the tests. As is the case with any promotion exam, some of the firefighters here invested substantial time, money, and personal commitment in preparing for the tests” (*Ricci*, p. 24). Only with a strong basis in evidence could the employer consider tossing the lists generated by the selection process, if adverse impact existed. By contrast, RIF decisions are competitive per se, but are typically based on pre-existing factors that are both individually (*e.g.*, seniority, performance reviews) and employer based (*e.g.*, needs within a specific department, desire to keep certain core functions). By keeping initial RIF decisions confidential, an employer is able to proactively analyze their *proposed* decisions for adverse impact prior to making their final decisions known. The intent of this initial analysis is to (perhaps) allow the employer to adjust their RIF decisions based upon the results of the analyses and the existence of adverse impact. Absent a strong basis in evidence, this adjustment of RIF decisions would likely not withstand legal scrutiny, but because the initial RIF decisions are confidential, it would be much harder for a potential plaintiff to realize what is happening. The best advice for employers making RIF decisions is to ensure that RIF decisions themselves are job-related and consistent with business necessity and to stand by their initial decisions.

Applying Ricci to compensation analyses.

Since the passage of the 1963 Equal Pay Act and the 1964 Civil Rights Act, the courts and federal government have been involved in making sure the compensation practices of employers are free of pay discrimination. While all subgroups are protected under these Acts, women and/or minorities are typically the plaintiffs in such actions. The U.S. Equal Employment Opportunity Commission (EEOC) and the U.S. Department of Labor’s enforcement arm, OFCCP proactively regulate and

enforce policies for analyzing the compensation practices of thousands of U.S. corporations, covering both federal- and non-federal contractors.

While active in their enforcement efforts, the government’s tools for identifying systemic compensation discrimination for the first two decades following the Equal Pay Act and Civil Rights Act have been limited. For example, from the 1990s to 2004 (when the draft version of the now final compensation standards were released), the typical “red flag” analyses used for identifying possible compensation discrimination consisted simply of making average salary comparisons between men and women and whites and minorities by job grouping variables such as job title or job group.⁵⁴ In the courtrooms, more sophisticated multiple regression procedures had been used but with mixed response, being called incomprehensible and deemed questionable in some courtrooms while being accepted and considered necessary in others.⁵⁵ The vast majority of high-level litigation, however, adopted this use of multiple regression for analyzing compensation data for possible discrimination.

On November 16, 2004, however, the landscape changed when the OFCCP released a new set of draft guidelines that solidified the role of multiple regression analysis in evaluating compensation discrimination. These draft guidelines went through a public review process and were released in final form on June 16, 2006, in the form of two sets of documents. The first set, titled “Interpreting Nondiscrimination Requirements of Executive Order 11246 with Respect to Systemic Compensation Discrimination,”⁵⁶ essentially constitutes OFCCP’s formal guidelines regarding analyzing compensation data received from federal contractors (“Interpretive Standard”). The second set is titled, “Voluntary Guidelines for Self-Evaluation of Compensation Practices for Compliance With Nondiscrimination Requirements of Executive Order 11246 With Respect to Systemic Compensation Discrimination,”⁵⁷

and is provided as a set of useful guidelines that federal contractors can utilize proactively to ensure that their compensation analyses will be done using procedures and techniques that will address the Interpretive Standards (“Voluntary Guidelines”).

The Voluntary Guidelines provide both a strong recommendation that federal contractors use multiple regression techniques to proactively identify pay disparities between groups, as well as proactively provide “make whole relief” when pay discrimination is identified. In fact, under federal affirmative action regulations, employers that are federal contractors are required to conduct pay analyses *annually* to ensure that minorities and women are being fairly compensated.⁵⁸ This is a similar function to what Croson Studies are designed to do—identify problems and prepare a path for correction. In addition, when contractors are scheduled for an on-site review for their compensation practices, Item 11 of the Scheduling Letter requests that the employer provide “annualized compensation data (wages, salaries, commissions, and bonuses) by either salary range, grade, or level showing total number of employees by race and gender and total compensation by race and gender.”⁵⁹ When supplied by the contractor, this information is analyzed by OFCCP for possible statistical indicators of pay discrimination. To encourage contractors to use the multiple regression and related techniques described in the Voluntary Guidelines, the OFCCP offers a program called “compliance coordination” which in effect states that the OFCCP will forego independent review of a contractor’s raw compensation data and simply review the contractor’s compliance with the Guidelines themselves.

While nearly four decades of enforcement in the courts and the federal enforcement guidelines certainly lay a substantial foundation for conducting proactive pay analyses, the outer boundaries surrounding pay investigations are currently unknown. Does *Ricci* weigh into where these lines are drawn? It should first be stated that while minorities

and women are most often the targeted group for inquiry in a pay review, the 1963 Equal Pay Act and the 1964 Civil Rights Act also protect whites and men and, to the extent that pay studies reveal significant pay discrepancies against these groups, they should be (and have been) enforced.

But what about situations—like in *Ricci*—where efforts to comply with civil rights laws based on a possible violation against one group apparently creates a situation where discrimination can occur against another? Such situations are not limited to just promotional testing circumstances—they have also occurred in pay analysis cases. For example, in *Rudebusch v. Hughes*⁶⁰ a federal judge ruled that Northern Arizona University violated the civil rights of 40 white male faculty members by giving salary raises exclusively to female and minority professors.

In *Rudebusch*, the University implemented a “pay equity” plan that gave one-time pay increases that averaged \$3,000 for minorities and \$2,400 for women. White males, however, received no increases. Such changes were made based on a (limited) regression model that was used to make \$207,613 of pay increases to *just female and minority male faculty members* whose salaries fell below the predicted salary of a similarly situated white male faculty member. While this practice (*i.e.*, making pay adjustments based upon significant disparities) is common (in fact, mandatory for compliance in some cases), the problem in *Rudebusch* was that no non-minority men were granted salary increases—*even those with salaries below their predicted levels*. In addition, the court noted three major inadequacies with the pay study used to justify making changes to minority and female pay levels.

The first inadequacy was that the study indicated that the *highest single pay disparity* was only 2.0 standard deviations from the predicted salary. This is not to be confused with an entire group’s pay triggering the 2.0 standard deviation test required for showing statistical significance. The highest single pay disparity in *Rudebusch* being “2.0 standard deviations from the

predicted salary” pertains to the distance of a single individual away from their predicted salary level—not an entire group factor (e.g., gender) being statistically significant at the 2.0 standard deviation level.

The next inadequacy was the fact that, “although over half the minority faculty made less than the predicted salary, a very large percentage of white male faculty also made less than the predicted amount.” The court deemed this as some degree of evidence that other factors besides discrimination may be to blame in explaining the imbalances.

The third limitation pointed out by the Court had to do with the regression study itself. The Court deemed the regression model “inadequate” because it did not include critical performance factors that were fundamental to predicting pay for the at-issue position (teachers):

As a final point, we have concerns about the way in which Hughes calculated and made the adjustments in this case. Particularly when, as was the case here, adjustments depend upon a regression analysis that does not account for performance factors such as academic credentials, performance, merit, teaching, research, or service—factors that are the major criteria for faculty compensation on campuses across the country—the failure to make some sort of more individualized

determination of what sort of adjustments are warranted in any given case will not satisfy strict scrutiny. (Rudebusch, para. 33).

Given this pay disparity-specific legal guidance, coupled with the more general guidance given in *Ricci* regarding the strong-basis-in-evidence standard, the following points should be considered when conducting proactive pay investigations:

- Never make pay adjustments without statistically significant evidence (e.g., comparing means only) that accounts for realistic qualification differences that may exist between groups. Using multiple regression is the only acceptable way to complete this process, and has enjoyed decades of the support and recent endorsement from federal enforcement agencies.
- Never make pay adjustments prior to performing a “cohort” analysis whereby additional variables (i.e., those not included within the regression analysis) are investigated.
- Conduct multiple regression studies under attorney-client privilege.
- Make sure that pay investigations look for possible pay disparities that may exist for *any group* (whites and men included), and when pay

Table 4: The Strong-Basis-in-Evidence Requirement Applied to RIFs and Compensation Analysis

Practice	Strong-Basis-in-Evidence Requirement (i.e., evidence of non-validity and/or no Alternate Employment Practice)			
	None	Some Risk	Moderate Risk	High Risk
Layoffs (RIFs).	Evaluating adverse impact of possible RIF decisions.	Changing with validated RIF criteria to lower adverse impact after learning about adverse impact.	Changing moderately-validated RIF criteria to lower adverse impact after learning about adverse impact.	Making RIF decisions based entirely on adverse impact results.
Compensation analyses.	Making pay changes to <i>any</i> statistically significantly impacted group based on a <i>strong</i> regression model.	Making pay changes to <i>any</i> statistically significantly impacted group based on a <i>moderate</i> regression model.	Making pay changes to <i>any</i> statistically significantly impacted group based on a <i>weak</i> regression model.	Making pay changes to only minorities or women based on weak regression models (or non-regression methods).

adjustments are made, be sure that the criteria and rules are *uniformly applied* across all gender and race/ethnic groups.

■ When pay adjustments are made based on a multiple regression model, ensure that the model itself is statistically significant, the strength of the regression model is sufficiently adequate for making predictions, and the fundamental factors that are relevant to pay are included.⁶²

All employers—especially federal contractors—are encouraged to invest in expert statistical and legal expertise to ensure their compensation evaluations and pay adjustment methodologies are sound. Table 4 summarizes the post-*Ricci* recommendations for RIFs and compensation analysis.

CONCLUSIONS

The *Ricci* case delivers the unmistakable mandate that employers need to perform a Croson Study to establish a *strong basis in evidence* to support a decision that remedial action is necessary *before* implementing race-conscious remedies that are as extreme as throwing out the results of a list

because the results adversely impact a group protected by Title VII of the Civil Rights Act when they do not feel it is job related enough to withstand a court challenge. Conducting a Croson Study under attorney-client privilege to determine if race-conscious remedies may be appropriate is a prudent step for any employer thinking of implementing race-conscious remedies as a part of their diversity initiatives. A Croson Study can provide a strong basis in evidence in determining the type of test, the weightings used, the cutoff scores, and alternate selection procedures that should be developed and administered to substantially reduce unnecessary adverse impact while selecting qualified candidates.

And finally, employers that evaluate alternate employment practices under a Croson Study should make sure to compile evidence, not just rhetoric, to support the position that the alternate employment practice (such as banding, a different set of weightings, a different cutoff score, or the use of an assessment center) will provide substantially equally qualified candidates with less adverse impact. ☒

ENDNOTES

1. *Richmond v. Croson*, 488 U.S. 469, (1989).
2. *Wygant v. Jackson Board of Education*, 476 U.S. 267 (1986).
3. *Ricci et al., v. DeStefano et al.*, USSC, No. 07–1428 (June 29, 2009).
4. An 80% violation occurs when the selection rate of the focal group is less than 80% of the reference group's selection rate. The 80% test is only sometimes used as a practical evaluation of possible adverse impact (statistical significance tests are the most definitive standard—see Biddle, 2006).
5. *Richmond v. Croson*, 488 U.S. 469, (1989).
6. *Wygant v. Jackson Board of Education*, 476 U.S. 267 (1986).
7. U.S. Constitution, Amendment 14: “All persons born or naturalized in the United States, and subject to the jurisdiction thereof, are citizens of the United States and of the State wherein they reside. No State shall make or enforce any law which shall abridge the privileges or immunities of citizens of the United States; nor shall any State deprive any person of life, liberty, or property, without due process of law; nor deny to any person within its jurisdiction the equal protection of the laws.” The Equal Protection Clause limits only the powers of government bodies, and not the private parties on whom it provides equal protection.
8. “Croson Studies” come from a 1989 Supreme Court decision that decided local governments could not establish preferences based upon race and gender unless there is proof of prior discrimination. Race is a suspect classification that is subject to strict judicial scrutiny. *Richmond v. Croson*, 488 U.S. 469, (1989). A “Croson Study” is used to justify minority business participation that is narrowly tailored to remedy past discrimination.
9. Dean, M. A., Bobko, P., & Roth, P. L. (2008). Ethnic and gender subgroup differences in assessment center ratings: A meta-analysis. *Journal of Applied Psychology*, 93, 685-691.
10. *Croson*, 488 U.S. at 500.
11. *Id.* at 498.
12. Donze, P. L. (2000). Supreme Court's denial of certiorari in Dallas fire fighters leaves unsettled the standard for compelling remedial interests. *Case Western Reserve Law Review*, 50, 759-796.
13. Alphan, D. M. (2003). Proving discrimination after Croson and Adarand: If it walks like a duck, *U.S.F. Law Review*, 37, Rev. 887, 892-93.
14. *Engineering Contractors Ass'n v. Metropolitan Dade County*, 122 F.3d 895, 907 (11th Cir. 1997), cert. denied, 118 S. Ct. 1186 (1998) (quoting *Ensley Branch, NAACP v. Seibels*, 31 F.3d 1548, 1565 (11th Cir. 1994); see also *Contractors Ass'n v. City of Philadelphia.*, 6 F.3d 990, 1004 (3d Cir. 1993).
15. *Contractors Ass'n*, 6 F.3d at 1005.
16. *Id.* at 1003 (quoting *Coral Constr. Co. v. King County*, 941 F.2d 910, 919(9th Cir. 1991).
17. *Contractors Ass'n*, 6 F.3d at 1003; *Coral Constr.*, 941 F.2d at 919; *Engineering Contractors*, 122 F.3d at 925; *Concrete Works, Inc. v. City & County of Denver*, 36 F.3d 1513, 1521 (10th Cir. 1994); *Croson*, 488 U.S. at 509.
18. Uniform Guidelines – Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, and Department of Justice (August 25, 1978), 43 Federal Register, 38,290-38,315.

19. *Adarand Constructors, Inc. v. Peña* (63 U.S.L.W. 4523, U.S. June 12, 1995).
20. U.S. Department of Justice (June 28, 1995). Legal guidance on the implications of the Supreme Court's decision in *Adarand Constructors, Inc. v. Peña*: Memorandum to general counsels (author).
21. *Gratz v. Bollinger*, 539 U.S. 244 (2003).
22. See, for example, pp. 22-24 in Biddle, D. (2006). *Adverse impact and test validation: A practitioner's guide to valid and defensible employment testing*, Second Edition. Burlington, VT: Gower Publishing Company; Biddle, R. E., (November 1995). Disparate impact reference trilogy. *Labor Law Journal*; Biddle, R.E., (April 1996). The role of two statistical approaches in EEO cases. *Labor Law Journal*.
23. The Lancaster mid-p correction should be used when computing the Fisher Exact Test. See <http://www.disparateimpact.com> for a web-based tool that can be used for calculating these adverse impact statistics.
24. *Sharkey v. Dixie Electric Membership Corporation*, No. 06-31199, 5th Cir., (Jan. 23, 2008).
25. *Doe v. Kamehameha School/Bernice Panabi Bishop Estate*, 470 F. 3d 827, 840 (9th Cir., 2006).
26. <http://www.merriam-webster.com/dictionary/standard+score>. See also: Francis, R. W. (Fall, 2006). Common errors in calculating final grades, *Thought & Action*.
27. The means and standard deviations of the oral interview and written were 63.43 / 12.35 and 71.44 / 10.79 respectively for the Lieutenant process and 69.45 / 11.72 and 72.05 / 10.43 respectively for the Captain process.
28. Johnson, J. W. & Lebreton, J. M. (2004). History and use of relative importance indices in organizational research. *Organizational Research Methods*, 7 (3), 238-257.
29. Biddle, R. E. (Spring, 1993). How to set cutoff scores for knowledge tests used in promotion, training, certification, and licensing. *Public Personnel Management*.
30. A computer program for calculating the conditional SEM is available from the first author. Standard 2.14 of the APA Standards requires that conditional standard errors of measurement are used and reported if constancy of measurement error cannot be assumed (which is normally the case) and states, "Where cut scores are specified for selection or classification, the standard errors of measurement should be reported in the vicinity of each cut score" (American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education (1999), *Standards for Educational and Psychological Testing*. Washington DC: American Educational Research Association.
31. *U.S. v. South Carolina*, 445 F. Supp. 1094 (D.S.C. 1977), aff'd U.S. Supreme Court (434 US, 1026, 1978).
32. *Isabel v. City of Memphis*, 404 F. 3d 404 (6th Cir. 2005).
33. For example, the Mann-Whitney U test can be used for evaluating whether the effect of ranking caused adverse impact against a group (see, for example, *Bridgeport Guardians v. City of Bridgeport*, 933 F.2d 1140, 1145 (2nd Cir. 1991); *Bryan v. Koch*, 492 F. Supp. 212, 220 (S.D.N.Y., 1980), aff'd, 627 F.2d 612 (2d Cir.1980); *Fritz v. Baker*, No. 87 Civ. 5662 (S.D.N.Y. Jan. 17, 1990) (1990 WL 3921), aff'd without opinion, 914 F.2d 239 (2d Cir. 1990). See an illustrated approach in Biddle, R. E. (Spring, 1993). How to set cutoff scores for knowledge tests used in promotion, training, certification, and licensing. *Public Personnel Management*.
34. *Vulcan Pioneers v. NJ Department of Civil Service*, 625 F. Supp. 527, 539 (D.NJ 1985).
35. *Guardians v. CSC of New York*, 630 F.2d 79 (2d Cir. 1980).

36. In Justice Ginsburg’s dissenting opinion in *Ricci*, she states: “. . .Notably, the exams were never shown to be suitably precise to allow strict rank ordering of candidates. A difference of one or two points on a multiple-choice exam should not be decisive of an applicant’s promotion chances if that difference bears little relationship to the applicant’s qualifications for the job” (FN 16).
37. The reliability of selection procedures justified on the basis of content validity should be a matter of concern to the user. Whenever it is feasible, appropriate statistical estimates should be made of the reliability of the selection procedure. (See 41 CFR Part 60-3.14[C][5])
38. *Griggs v. Duke Power*, 401 U.S. 424 (1971)..
39. Title VII Section 2000e-2(l) Prohibition of discriminatory use of test scores: It shall be an unlawful employment practice for a respondent, in connection with the selection or referral of applicants or candidates for employment or promotion, to adjust the scores of, use different cutoff scores for, or otherwise alter the results of, employment related tests on the basis of race, color, religion, sex, or national origin.
40. Title VII Section 2000e-(k)(1)(A)(ii)(k) **Burden of proof in disparate impact cases.** (1) (A) An unlawful employment practice based on disparate impact is established under this subchapter only if— (i) a complaining party demonstrates that a respondent uses a particular employment practice that causes a disparate impact on the basis of race, color, religion, sex, or national origin and the respondent fails to demonstrate that the challenged practice is job related for the position in question and consistent with business necessity; or (ii) **the complaining party makes the demonstration described in subparagraph (C) with respect to an alternative employment practice and the respondent refuses to adopt such alternative employment practice.** (B) (i) With respect to demonstrating that a particular employment practice causes a disparate impact as described in subparagraph (A)(i), the complaining party shall demonstrate that each particular challenged employment practice causes a disparate impact, except that if the complaining party can demonstrate to the court that the elements of a respondent’s decision making process are not capable of separation for analysis, the decision making process may be analyzed as one employment practice. (ii) If the respondent demonstrates that a specific employment practice does not cause the disparate impact, the respondent shall not be required to demonstrate that such practice is required by business necessity. **(C) The demonstration referred to by subparagraph (A)(ii) shall be in accordance with the law as it existed on June 4, 1989, with respect to the concept of “alternative employment practice.”**
41. Uniform Guidelines section 60.3B: **Consideration of suitable alternative selection procedures.** Where two or more selection procedures are available which serve the user’s legitimate interest in efficient and trustworthy workmanship, and which are substantially equally valid for a given purpose, the user should use the procedure which has been demonstrated to have the lesser adverse impact. Accordingly, whenever a validity study is called for by these guidelines, the user should include, as a part of the validity study, an investigation of suitable alternative selection procedures and suitable alternative methods of using the selection procedure which have as little adverse impact as possible, to determine the appropriateness of using or validating them in accord with these guidelines. If a user has made a reasonable effort to become aware of such alternative procedures and validity has been demonstrated in accord with these guidelines, the use of the test or other selection procedure may continue until such time as it should reasonably be reviewed for currency. Whenever the user is shown an alternative selection procedure with evidence of less adverse impact

and substantial evidence of validity for the same job in similar circumstances, the user should investigate it to determine the appropriateness of using or validating it in accord with these guidelines. This subsection is not intended to preclude the combination of procedures into a significantly more valid procedure, if the use of such a combination has been shown to be in compliance with the guidelines.

42. *Wards Cove Packing Co. v. Atonio*, 490 U.S. 642 (1989).
43. *Albemarle*, supra, at 425.
44. *Watson v. Fort Worth Bank & Trust*, 487 U.S. 977 (1988).
45. *U.S. v. City of New York*, 07-cv-2067, NGG, RLM (2009).
46. http://www.dol.gov/ofccp/regs/compliance/faqs/Ricci_FAQ.htm.
47. See Biddle, 2006.
48. Most CSEM methods require post-administration statistics for computation. However, the Lord-Keats method can be used for estimating CSEMs before a test has been administered (see Lord, F. M. [1984]. Standard errors of measurement at different ability levels. *Journal of Educational Measurement*, 21, 239-243).
49. See, for example, Aguinis, H. (2004). *Test-score banding in human resource selection: legal, technical, and societal issues*. Praeger Publishers.
50. See Biddle, D. (2008). Overview of C-SEM Methods. In Hurtz, G. M. (Chair), Integrating conditional standard errors of measurement into personnel selection practices. Symposium presented at the Annual Conference of the Society for Industrial and Organizational Psychology. San Francisco, CA, April 2008. See also, Biddle, D., Kuang, D. C. Y., & Higgins, J. (2007, March). Test use: ranking, banding, cutoffs, and weighting. Paper presented at the Personnel Testing Council of Northern California, Sacramento.
51. See, for example: Cave, B. (2009, July 15). Ricci v. DeStefano Supreme Court finds that city discriminated against white employees. *Labor and Employment Bulletin*. Retrieved from <http://www.bryancave.com/files/Publication/b1149e33-b667-436b-9cc4-d0d6f8ecbd27/Presentation/PublicationAttachment/352c8bfc-5f1f-4e53-8dd1-d1cb9e97cb56/LaborAlert7-15-09.pdf>. Hammell, J.W. & Curtin, Z. M. (2009, July 10). Ricci v. DeStefano: Supreme Court holds employer liable for trying to avoid claims of adverse-impact discrimination. Dorsey & Whitney LLP. Retrieved from http://www.dorsey.com/ricci_analysis/
52. *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.
53. *Smith v. City of Jackson* (03-1160, 544 U.S. 228, 2005, 351 F.3d 183, affirmed).
54. Truesdell, W. H. (2003). *Secrets of Affirmative Action Compliance* (6th ed.). The Management Advantage, Inc. (Author). p. 311.
55. Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction* (3rd ed). Wadsworth.
56. OFCCP—U.S. Department of Labor (DOL), Employment Standards Administration, Office of Federal Contract Compliance Programs (November 16, 2004). *Interpreting Nondiscrimination Requirements of Executive Order 11246 With Respect to Systemic Compensation Discrimination* (Notice). Federal Register, Vol. 69, No. 220.
57. OFCCP—U.S. Department of Labor (DOL), Employment Standards Administration, Office of Federal Contract Compliance Programs (June 16, 2006). *Interpreting Nondiscrimination Requirements of Executive Order*

- 11246 With Respect to Systemic Compensation Discrimination* (Notice). Federal Register, Vol. 71, No. 116.
58. 41 CFR 60.2.17(b)(3).
59. Interpretive Standards, p. 35125.
60. *Rudebusch v. Hughes*, 313 F.3d 506 (9th Cir. 2002).
61. Teske, J. (Fall 2003). The significance of statistical significance: Ninth Circuit clarifies usefulness of statistical evidence when implementing pay equity adjustments in *Rudebusch v. Hughes*. *Loyola of Los Angeles Law Review*. Rev. 153.
62. Checking to see whether the model itself is significant can be accomplished by evaluating the ANOVA associated with the model; the strength of the regression model can be evaluated using the Adjusted R^2 value; the predictive efficiency of the model can also be weighed by evaluating the Adjusted R^2 as well as the degree of multicollinearity among the variables (which tends to inflate standard errors associated with predictions), as well as making sure that the most relevant (fundamental) pay factors are included in the model.

ADDITIONAL RESOURCES

AFFIRMATIVE ACTION PLANNING AND AUDIT RESOURCES

- OFCCP Website <http://www.dol.gov/ofccp/index.htm>
- OFCCP Frequently Asked Questions <http://www.dol.gov/ofccp/regs/compliance/faqs/offaqs.htm>
- Census Data (2000 EEO File, Details) <http://www.eeoc.gov/stats/census/index.html>
- Census Data (2000 EEO File, Live Data) <http://www.census.gov/eo2000>
- Census data for Puerto Rico <http://www.dol.gov/ofccp/regs/compliance/PRicoeod.htm>
- Zip Code/County Search <http://www.zip-codes.com/search.asp>
- Construction AAP Guide <http://www.dol.gov/ofccp/TAguides/ctaguide.htm>
- Uniform Guidelines (Regulation Format) http://www.dol.gov/dol/allcfr/ESA/Title_41/Part_60-30/toc.htm
- Uniform Guidelines (User-Friendly Format) ... <http://www.uniformguidelines.com>
- AAP Fact Sheet <http://www.dol.gov/ofccp/regs/compliance/aa.htm>
- Compliance Manual <http://www.dol.gov/federalregister/DocumentList.aspx?AgencyId=10&DocumentType=3>
- State Job Bank Information <http://www.jobbankinfo.org>

FEDERAL REGISTERS, LAWS, AND REGULATIONS

- Title VII of the Civil Rights Act of 1964 <http://www.eeoc.com/policy/laws/title-vii-of-the-civil-rights-act-of-1964>
- E-Verify http://www.dhs.gov/xprevprot/programs/gc_1185221678150.shtm
- Americans with Disabilities Act <http://www.dol.gov/ofccp/regs/statutes/ada.htm>
- Federal Register Notices <http://www.dol.gov/federalregister/DocumentList.aspx?AgencyId=10&DocumentType=2>
- Regulations for AAPs (41CFR Chapter 60) http://www.dol.gov/dol/allcfr/ESA/Title_41/Chapter_60.htm
- Executive Order 11246 <http://www.dol.gov/ofccp/regs/statutes/eo11246.htm>
- Section 503 of the Rehabilitation Act of 1973, as amended <http://www.dol.gov/ofccp/regs/compliance/sec503.htm>
- Vietnam Era Veterans' Readjustment Assistance Act of 1974, as amended <http://www.dol.gov/ofccp/regs/statutes/4212.htm>
- VEVRAA Fact Sheet <http://www.dol.gov/vets/programs/fact/vet97-5.htm>
- Policy Directives <http://www.dol.gov/ofccp/regs/statutes/eo11246.htm>
- G-FIVE <http://www.dol.gov/ofccp/regs/compliance/directives/dir282.htm>
- Federal Contractor's Online Application Selection System <http://www.dol.gov/ofccp/regs/compliance/directives/dir281.htm>
- Corporate Management Reviews <http://www.dol.gov/ofccp/regs/compliance/directives/dir202.htm>
- Active Case Management (ACM) rules <http://www.dol.gov/ofccp/regs/compliance/directives/dir285.htm>
- Investigative procedures when a test(s) is one cause of adverse impact in hiring <http://www.dol.gov/ofccp/regs/compliance/directives/dir267.htm>

Retention of Documentshttp://www.dol.gov/ofccp/regs/compliance/directives/dir77_26.htm
Calculating Interest on Back Pay<http://www.dol.gov/ofccp/regs/compliance/directives/dir280.htm>
Separate Facility Exemptions<http://www.dol.gov/ofccp/regs/compliance/directives/dir260.htm>

ADDITIONAL AGENCY AND ORGANIZATION RESOURCES

EEOC Website<http://www.eeoc.gov>
EEO-1<http://www.eeoc.gov/eo1survey>
VETS-100<https://vets100.vets.dol.gov>
National Industry Liaison Group (NILG)<http://www.nationalilg.org/main.html>
National Opinion Research Center (NORC)
 data for graduate surveys<http://www.norc.org/homepage.htm>
Employment Resource Directory
 (OFCCP Regional Outreach)<http://www.dol.gov/ofccp/ERRD/errsrvs.htm>
National Employment Law Institute (NELI)<http://www.neli.org>
Society for Industrial & Organizational
 Psychology (SIOP)<http://www.siop.org>
American Psychological Association<http://www.apa.org>

CALENDAR *of* EVENTS

DATE	ORGANIZATION	EVENT	ADDITIONAL DETAILS	CONTACT FOR MORE INFORMATION
3/21– 3/24/10	National Employment Law Institute	Employment Law Briefing	Las Vegas, NV	www.neli.org
3/30/10	Arizona Affirmative Action Association	Monthly Meeting	Affirmative Action Plans A to Z: Compliance Evaluations	John Garza, President JGQuadA@aol.com
4/8/10	Personnel Testing Council of Northern California	Luncheon Meeting	Sacramento, CA	www.ptcnc.org
4/8– 4/9/10	National Employment Law Institute	ADA & FLMA Compliance Update	Chicago, IL	www.neli.org
4/8– 4/10/10	Society for Industrial/Organizational Psychology	Annual Conference	Atlanta, GA	www.siop.org
4/15– 4/16/10	National Employment Law Institute	ADA & FLMA Compliance Update	San Francisco, CA	www.neli.org
4/20/10	Arizona Affirmative Action Association	Annual Conference	Location TBD	John Garza, President JGQuadA@aol.com
4/21/10	BCG Institute for Workforce Development	Webinar - Record Keeping For Transactions With An Emphasis On Adverse Impact Analysis	11:00AM - 12:30PM PST 2:00PM - 3:30PM EST	www.bcginstitute.org
4/22– 4/23/10	National Employment Law Institute	ADA & FLMA Compliance Update	Washington, DC	www.neli.org
4/26– 4/28/10	Northern California Human Resources Association	HR West Conference	South San Francisco	www.hrwest.org
4/26– 4/28/10	Society for Human Resources Management	Staffing Management Conference & Exposition	Orlando, FL	www.shrm.org
5/18/10	Arizona Affirmative Action Association	Monthly Meeting	Applicant Tracking	John Garza, President JGQuadA@aol.com
5/19/10	BCG Institute for Workforce Development	Webinar - Compensation 101 For Federal Contractors—Proactive Analysis for 2010	11:00AM - 12:30PM PST 2:00PM - 3:30PM EST	www.bcginstitute.org

* Dates and event information are subject to change. If you would like to submit an event for inclusion in the calendar, please e-mail editor@eeoinsight.com.



DATE	ORGANIZATION	EVENT	ADDITIONAL DETAILS	CONTACT FOR MORE INFORMATION
5/20/10	Minnesota Affirmative Action Compliance Council	May Meeting	Legal and Practical Issues of Applicant Tracking Systems	Christi Anthony Canthony@biddle.com
6/27– 6/30/10	Society for Human Resources Management	Annual Conference & Exposition	San Diego Convention Center	www.shrm.org
8/3– 8/6/10	Industry Liaison Group	National Conference	Red Rock Resort Las Vegas, Nevada	www.ilgnvegas2010.com