4-26-2011

Disparate Impact Realism

Amy L. Wax
University of Pennsylvania Law School, awax@law.upenn.edu

Disparate Impact Realism

by
Amy L. Wax
Robert Mundheim Professor of Law
University of Pennsylvania Law School
(forthcoming *William and Mary Law Review*, Fall 2011)

DRAFT: *do not cite or quote without permission*

I. *Introduction and summary*

In *Ricci v. DeStefano*, 129 S. Ct. 2658 (2009), the Supreme Court recently reaffirmed the doctrine, first articulated by the Court in *Griggs v. Duke Power Company*, 401 U.S. 424 (1971), that employers can be held liable under Title VII of the 1964 Civil Rights Act for neutral policies with a disparate impact on minority workers. The Court has further held that employers can escape liability by showing that the policy is job related or consistent with business necessity.[1]

In the interim since *Griggs*, social scientists have generated a substantial body of research designed to help employers comply with the mandates of the doctrine. This evidence has undermined two key elements of *Griggs* that have informed the application of the disparate impact rule more generally. First, *Griggs* and its progeny rest on the implicit assumption that fair and valid staffing practices will result in workers from each race being hired or promoted in rough proportion to their numbers in the background population or in an otherwise appropriately defined pool of candidates. The so-called 4/5 rule, under which an employer is presumptively liable if the percentage of minority applicants hired is less than 80% of those selected from the majority white population, reflects this assumption.[2] Second, the Court in *Griggs* noted the

---

[1] *Griggs*, 401 U.S. 431 ("The touchstone is business necessity. If an employment practice which operates to exclude Negroes cannot be shown to be related to job performance, the practice is prohibited.")

[2] See discussion of 4/5 rule infra.

absence of evidence that the screening criteria in that case (a high school diploma and performance on a "professionally developed aptitude test")[3] were related to subsequent performance of the service jobs at issue, and expressed doubt about the existence of such a relationship.[4]

Social science research casts doubt on both these aspects of *Griggs*. First, research in industrial and organization psychology (IOP) has repeatedly documented that, despite their imperfections, tests and criteria such as those at issue in *Griggs* (which are heavily "g"-loaded and thus dependent on cognitive ability)[5] remain the best predictors of performance for jobs at all levels of complexity. Second, work in psychometrics, educational demography, and labor economics indicates that blacks, and to a lesser extent hispanics, currently lag behind whites both in cognitive ability test performance and in the skills needed for success on the job. These gaps are reflected in lower scores on the types of g-loaded job screens that best predict job success. The combination of well-documented racial differences in cognitive ability and the consistent link between ability and job performance generates a pattern that experts term "the validity-diversity tradeoff": the most effective job selection criteria consistently generate the smallest number of minority hires. Indeed, the evidence indicates that most valid screening devices will have a significant adverse impact on blacks and will also violate the 4/5 rule under the law of disparate impact.

In sum, the IOP literature demonstrates that the empirical and demographic premises

---

[3] *Griggs*, 401 U.S. 426-428

[4] *Griggs*, 401 U.S. 428 (Noting that neither of the aptitude tests required for hiring and promotion in that case were "directed or intended to measure the ability to learn to perform a particular job or category of jobs.") See also 401 U.S. 431-432 (noting the absence of a demonstrable connection between the tests at issue and job performance and observing that workers at the company who had not graduated from high school or taken the test were performing in a satisfactory manner).

[5] The tests at issue in *Griggs* were the Wonderlic Personnel Test (which is a standard type of intelligence test) and the Bennett Mechanical Comprehension Test. 401 U.S. 428.

behind the disparate impact rule do not match reality and have turned out to be myths. As a consequence, most *legitimate* job selection practices, including those that predict productivity better than alternatives, will routinely trigger liability under the current rule. Although the Supreme Court in *Griggs* and subsequent cases has repeatedly stated that disparate impact doctrine is consistent with a rigorously competitive meritocracy, employers seeking to maintain such a meritocracy among a diverse population will run a high risk of being sued for violations of the rule. Such lawsuits will put employers to the onerous, uncertain, and sometimes impossible task of justifying their job selection practices, which may result in unwarranted liability or induce undesirable, self-protective strategies.

To help alleviate these consequences, this paper proposes to modify the doctrine of disparate impact to adopt a new standard of "disparate impact realism." The disparate impact rule should be revised by making two changes in the standard that triggers potential liability. First, the target 4/5 ratio of minority to majority hires should be relaxed to reflect the empirically documented gap in actual productivity between whites and minority workers. Second, the fixed 4/5 threshold ratio should be abandoned in favor of a sliding scale relationship, documented in the IOP literature, that pegs expected group staffing ratios to measured disparities in group performance and the selectivity of particular positions.

Although not altogether relieving employers of the burdens imposed by the disparate impact rule, disparate impact realism compares favorably with the current regime. By shrinking the number of employment practices that can potentially trigger liability, realism lessens the pressure to hire a racially balanced workforce, especially for highly selective jobs. Moreover, the uncertainties and potential Constitutional difficulties generated by the Supreme Court's recent decision in *Ricci v. DeStefano* make it desirable to cut down on the number of situations that can generate disparate impact claims. Finally, disparate impact realism functions as an information- forcing device. By making it easier for employers to satisfy the rule and aligning expectations with current labor demographics, realism enhances employers' incentives to devise

personnel practices that maintain productivity while achieving maximally feasible diversity.

Alternatively, this paper proposes repealing the disparate impact rule altogether. The principal argument for repealing disparate impact is that, under present social conditions, racial imbalances in employment are exceedingly weak evidence of discrimination, either in the form of race-based disparate treatment or through unlawful disparate impact. The IOP data indicate that differences in the distribution of skill and human capital, not race-based exclusion or arbitrary barriers to employment, are the principal factors behind racial imbalances on the job. In light of these realities, the disparate impact rule is fatally overbroad and ensnares far too much conduct in its net. Under current social conditions, the vast majority of commonly employed selection procedures are valid and job related, and thus do not actually violate the disparate impact rule. Yet most valid personnel practices will routinely show enough adverse impact to create a prima facie case of discrimination, thereby shifting the burden of justification to employers. Given the legal uncertainties and practical difficulties of defending disparate impact claims, employers run a significant risk of being found liable regardless of whether their defense is valid, and even though they are not actually violating the rule. Virtually no aspect of the business necessity defense is settled law, so employers still face the prospect of protracted, expensive, uncertain, and resource intensive litigation to defend their practices. This encourages them to engage in perverse, inefficient, and evasive tactics, including de facto affirmative action. In sum, the overbreadth of the disparate impact rule is both inefficient and fundamentally unfair. Racial preferences are also directly at odds with the meritocratic goals of the disparate impact rule.

The data currently reveal that most jobs are more diverse than disparate impact doctrine actually requires. Indeed, blacks lag behind whites in performance on the job in many categories. This indicates that employers are not arbitrarily excluding minorities from the workforce, but rather bending over backwards to include them. In addition, disparate impact litigation does nothing to correct the underlying skill disparities reflected in these on-the-job

gaps, and distracts from the task of addressing them. The doctrine represents a costly, misplaced effort that could better be directed at addressing the root causes of workforce racial imbalance.

II. *Disparate impact employment discrimination: the doctrine and its uncertainties*

In *Griggs v. Duke Power Company*, the Supreme Court ruled for the first time that job requirements with a disparate impact on minorities, despite being "neutral on their face, and even neutral in terms of intent," could be unlawful under Title VII of the Civil Rights Act.[6] The Court further held that an employer could escape liability if job criteria at issue were shown to have "a manifest relationship to the employment in question."[7] As the Court stated, "The touchstone is business necessity."[8] The Court ruled in *Griggs* that the job requirements at issue in that case – a high school diploma or a minimum score on an IQ test – were impermissible because they screened out too many black applicants and were not shown to be "reasonably related to job performance" for the positions at issue.

In *Griggs* and subsequent cases expanding on the disparate impact (DI) doctrine, the Supreme Court has repeatedly stated that disparate impact rules do not mandate a particular racial balance or ethnic makeup in the workplace.[9] Rather, the objective is to "achieve equality

---

[6] The Supreme Court has also recognized that Title VII forbids disparate treatment, defined as adverse action against an employee motivated by or taken "because of" a protected characteristic, such as race or sex. See, e.g., David Sherwyn and Michael Heise, The *Gross* Beast of Burden of Proof: Experimental Evidence on How the Burden of Proof Influences Employment Discrimination Case Outcomes, 42 Arizona St. L. J. 901, 905 (Fall 2010)(noting that "[m]ost employment discrimination cases fall into two general categories: disparate treatment (intentional discrimination) and disparate impact" which occurs when "a company has a policy or practice that, while neutral on its face, adversely affects a protected class"). See also McDonnell Douglas Corp. V. Green, 411 U.S. 792 (1973)(describing unlawful disparate treatment as adverse treatment motivated by race).

[7] *Griggs*, at 432.

[8] *Griggs*, at 431.

[9] See, e.g., 401 U.S. at 430-431. See also *Watson v. Fort Worth Bank and Trust*, 487 U.S. 977 (1988). See also Civil Rights of 1991.

of employment opportunities and remove barriers that have operated in the past to favor an identifiable group."[10] The doctrine's stated goal of equal opportunity is consistent with a competitive meritocracy in which businesses are assumed to have a legitimate stake in selecting the best and most productive workers, and in developing and adopting personnel practices that best accomplish this goal. On this view, employers are entitled to hold all job seekers to uniform requirements, as long as those criteria are work-related.[11] The goal of the disparate impact doctrine is therefore at odds with race-conscious double standards or forms of affirmative action that selectively maintain less exacting criteria for some social or racial groups.[12]

---

[10] 401 U.S. at 429-430.

[11] See, e.g., Kenneth R. Davis, Wheel of Fortune: A Critique of the "Manifest Imbalance" Requirement for Race-Conscious Affirmative Action under Title VII, 43 Georgia L. Rev. (Summer 2009), 995-1057, 1037 ("The Supreme Court has identified equality of opportunity and meritocracy as goals of Title VII . . . in [*Griggs*], the Court articulated these objectives when creating the legal standard for disparate impact cases."); see also Harris, Cheryl I., and Kimberly West-Faulcon, Reading *Ricci*: Whitening Discrimination, Racing Test Fairness, 58 UCLA Law Rev. 73, 156-157 (October 2010) ("Title VII promotes selection that is more merit-based and thus is not a mechanism to enact racial preferences.")

[12] That the goals of disparate impact enforcement are inconsistent with race-based affirmative action has been widely and consistently recognized. See generally Rutherglen, George, Ricci v. DeStefano: Affirmative Action and the Lessons of Adversity, 2009 *Supreme Court Review* 83. See also, e.g., Harris, Cheryl I., and Kimberly West-Faulcon, supra., at 156-157; Carle, Susan, A Social Movement History of Title VII Disparate Impact Analysis, 63 Florida L. Rev. 251, 258-259 (2011) (noting that "popular perception sometimes conflates disparate impact analysis with affirmative action," but that "the two anti-discrimination concepts are actually quite different"); id. At 296 (noting that disparate impact is often confused with "bugbears such as quotas, strong race-conscious mandates, and harsh forms of affirmative action").

Attempts to use disparate impact doctrine to engage in race-conscious racial balancing have been turned back by Congress. So-called "race norming," or adjusting scores or selection methods based on race, was outlawed in the 1991 Civil Rights Act. See Civil Rights Act of 1991, Sections 106, 107(a) ("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"). For more discussion of race-norming, and the distinction between the disparate impact theory and affirmative action, see infra.

In the wake of *Griggs*, the EEOC and the courts have been charged with crafting rules consistent with the doctrine's stated goals and creating guidelines for their application. The Supreme Court has addressed adverse impact in only a few cases, and its decisions are vague or silent on key questions surrounding its application.[13] In light of this lack of guidance, the lower courts have grappled with defining staffing patterns that trigger application of the rule, and with identifying the appropriate population or pool against whom adverse impact is measured. Courts have also been asked to determine the scope of the business necessity defense that employers can invoke once adverse impact is demonstrated. All of these issues have generated considerable uncertainty and none has received a definitive resolution.

On the question of the hiring patterns that trigger liability, the EEOC, through the Uniform Guidelines on Employee Selection Procedures, has adopted what is known as the four-fifths (4/5), or 80% rule, as a touchstone for determining adverse impact.[14] Hiring or promoting minorities at less than 80% of the rate for the majority group gives rise to a potential violation and suffices to establish a prima facie case of a disparate impact discrimination. Partly

---

[13] See, e.g., David Sherwyn and Michael Heise, The *Gross* Beast of Burden of Proof: Experimental Evidence on How the Burden of Proof Influences Employment Discrimination Case Outcomes, 42 Arizona St. L. J. 901, 905 (Fall 2010)(noting that the Supreme Court "has addressed adverse impact cases on a few occasions").

[14] See 29 C.F.R. 1607.4D (" a selection rate for any race, sex, or ethnic group, which is less than four-fifths (4/5) of the rate for the group with the highest rate will generally be regarded . . as evidence of adverse impact"). Courts are not bound by the EEOC Guidelines, but the Supreme Court has said that they should receive "great deference," *see Griggs*, 401 U.S. at 433-434 (1971). The 4/5 rule was embraced by the Supreme Court in *Connecticut v. Teal*, 457 U.S. 440 (1982) and subsequent decisions, and has been repeatedly applied by the lower courts in disparate impact cases. For recent reviews of the standard for a prima facie case see, e.g., Rutherglen, George, Ricci v. DeStefano: Affirmative Action and the Lessons of Adversity, 2009 *Supreme Court Review* 83-114, 105-106; Harris, Cheryl I., and Kimberly West-Faulcon, Reading *Ricci*: Whitening Discrimination, Racing Test Fairness, 58 UCLA Law Rev. 73, 135-136 (October 2010).

in response to critiques of the 4/5 rule as insufficiently nuanced and statistically naive,[15] the courts have not relied exclusively on this method, and some have scrutinized workplace diversity using commonplace tests of statistical significance. Although the courts and the EEOC permit recourse to alternative methods,[16] the 4/5 rule remains an important benchmark for assessing disparate impact.

In applying the 4/5 rule, the courts have confronted the problem of defining the applicable baseline population against which to assess unlawful impacts. Questions arise as to whether the 4/5 rule should be gauged against a broader baseline, such as the adult work-eligible population, or defined more narrowly to include, for example, the local adult population, actual job applicants, or work-eligible persons possessing threshold qualifications.[17] All of these

---

[15] See, e.g., Sheldon Zedeck, Adverse Impact: History and Evolution, in Outtz, James L.(ed.). (2010). *Adverse Impact: Implications for Organizational Staffing and High Stakes Selection* [hereinafter Outtz, *Adverse Impact*]. See also Nancy T. Tippins. Adverse Impact in Employee Selection Procedures From the Perspective of an Organizational Consultant, in Outtz, *Adverse Impact* (suggesting that, under some circumstances, a 4/5 rule violation could easily occur by chance).

[16] See Guidelines at – (stating that smaller differences in selection rate than dictated by the 4/5 rule may constitute adverse impact, where they are "significant in both statistical and practical terms."); Jennifer L. Peresie, 84 Indiana L. J. 773 (2009) at 777 ("Plaintiffs generally prove [disparate impact] causation by comparing selection rates of majority and minority applicants for a position and then showing that the disparity is statistically significant or that it violates the four-fifths rule.") See also, e.g., *Hazelwood School Dist. v. United States*, 433 U.S. 299, 309 n.14 (1977) (noting a "general rule" in employment discrimination cases with sufficiently large samples that "if the difference between the expected value and the observed number is greater than two or three standard deviations, then the hypothesis that [employees] were hired without regard to race would be suspect"). See generally Philip Bobko and Philip L. Roth. An Analysis of Two Methods for Assessing and Indexing Adverse Impact: A Disconnect Between the Academic Literature and Some Practice. pp. 29-49, in Outtz, *Adverse Impact*, supra (noting the use of the 4/5 rule and as well as various tests of statistical significance in disparate impact cases); id. At 30-31 & 35 (noting that the EEOC endorses of the 4/5 rule but also permits the use of recognized tests of statistical significance in conjunction with or in place of that criterion in appropriate cases.)

[17] See, e.g., Jennifer L. Peresie, Toward Coherent Test for Disparate Impact Discrimination, 84 Ind. L. J. (2009) at 778 (footnotes omitted)("Plaintiffs must establish

approaches are problematic and some are open to challenge on the very principles underlying the disparate impact rule. Using applicants is suspect because employers who discriminate may discourage minorities from applying, thus skewing the baseline pool. Threshold requirements of any kind can end up screening out minority applicants, and thus using such requirements to define the baseline is itself vulnerable to challenge under the disparate impact rule.[18] Unfortunately, the Supreme Court has set no clear standard for identifying the population against which workplace disparate impact should be assessed and the lower courts vary in their approach.[19] This aspect of disparate impact doctrine is in serious disarray.

---

disparate impact with respect to the pool of qualified persons in the relevant labor market for the give position. Most often, plaintiffs present statistics from the actual applicant pool for the position. Plaintiffs might also choose to use national population statistics; state data, as in Griggs,; or data from a smaller geographic area").

[18] See, e.g., Scott Baker, Defining "Otherwise Qualified Applicants": Applying an Antitrust Relevant-Market Analysis to Disparate Impact Cases, 67 Chicago L. Rev. (2000), 725-747, 732 ("The obvious flaw" in using job requirements to define the job applicant pool "is that it ignores the effect that a hiring requirement that is known to have a disparate impact would have on a potentially qualified applicant's decision whether or not to apply for the job in the first place. It is unlikely that a potentially qualified applicant would take the time to apply for a job if she knew that a particular hiring requirement would prevent her from getting the job.")

It can be argued that the point of disparate impact scrutiny is to question the job relatedness of every requirement or selection factor that might produce an adverse impact. The assumption implicit in this view is familiar from *Griggs*: that all persons, regardless of group identity, is presumed equally qualified to do every job, and that every qualification, requirement, and job hurdle that generates a departure from racial balance must be justified as job-related. The actual practice does not always adhere to this analysis. See, e.g., Ian Ayres, Testing for Discrimination and the Problem of Included Variable Bias (draft on file with author – Penn Law and Economics Seminar series, October 6, 2010). On the problem of identifying the baseline population for purposes of DI analysis, see generally, e.g., Martha Chamallas, Evolving Conceptions of Equality under Title VII, 31 UCLA Law Rev. 305 (1982). See also *Wards Cove Packing v. Antonio*, 490 U.S. 642 (1989) (raising the question of how to define the eligible pool for disparate impact scrutiny).

[19] See, e.g., Joseph Gastwirth, Employment Discrimination: A Statistician's Look at Analysis of Disparate Impact Claims, 11 Law & Inequality151 (1992-1993) (discussing the uncertainty surrounding identifying the majority and minority labor market pool available for particular jobs). See Defining "Otherwise Qualified Applicants": Applying an Antitrust Relevant-Market Analysis to Disparate Impact Cases, 67 Chicago L. Rev. (2000), 725-747, 732

Another crucial issue in disparate impact doctrine is the standard for establishing the defense of job-relatedness consistent with business necessity. Showing job relatedness involves demonstrating a relationship between a screening device and ability to do the job -- a process known as validation. Drawing on standards developed by the courts as well as work in the IOP field, the EEOC guidelines recognize three principal methods by which employers can justify their selection procedures: content, criterion, and construct validation.[20] Both construct and criterion validation require demonstrating a formal and statistically valid relationship between job selection methods and either specified job-related skills (construct validation), or workers' actual performance on the job (criterion validation). As the most rigorous and demanding process, criterion validation is considered the "gold standard" and is the focus of considerable study by IOP experts. In contrast, content or "facial" validation is regarded as less exacting. Content validation does not generally require a formal demonstration that a job criterion actually predicts superior job performance or productivity. Rather, it depends on showing a manifest

---

("Because the Supreme Court has not laid out specific guidelines for defining the scope of the qualified applicant pool, district courts have developed various methods for making this determination."). See e.g., NAACP v. Town of East Haven, 998 F Supp 176 (D. Conn 1998 (applicant pool deemed to consist of qualified teachers from defined geographical area). See also Scott Baker, supra, discussing other cases; Jennifer Peresie, 84 Ind. Law J., supra, at 778.

[20] For a description of current EEOC guidelines, see e.g., Brief for Industrial-Organizational Psychologist as *Amici Curiae* in Support of Respondents, Ricci v. DeStefano, 129 S.Ct. 2658 (2009) (Nos. 07-1428 & 08-328) at 4-7. See also id. at 7, note 2, citing the EEOC Uniform Guidelines, and noting that the inference of validity rests on "evidence . . . that a test identifies those who are qualified to do the job," and explaining that "content validity supports [an inference of validity] by showing that the test's content matches the essential content of the job, while criterion validity supports the inference by showing that test results successfully predict job performance." In addition "construct validity is more abstract and is shown by evidence that the test measures the degree to which candidates have characteristics, or traits, that have been determined to lead to successful job performance." See also Robert Belton, the Unfinished Agenda of the Civil Rights Act of 1991, 45 Rutgers L. Rev. (1992); Harris, Cheryl I., and Kimberly West-Faulcon, Reading *Ricci*: Whitening Discrimination, Racing Test Fairness, 58 UCLA Law Rev. 73, 144-147 (reviewing validation standards); Rutherglen, George, Ricci v. DeStefano: Affirmative Action and the Lessons of Adversity, 2009 *Supreme Court Review* 83, 106-107 (same).

relationship or plausible match between the abilities assessed and the tasks that must actually be performed on the job.[21]

The burden is compounded by the intricacies of the validation rules. The details of the Guidelines are complex and ambiguous, and courts' practices erratic. Although the Guidelines recommend formal statistical validation, they also acknowledge content validation (the demonstration that selection criteria are closely geared to job tasks) as an accepted method for satisfying the job-relatedness requirement without clarifying which methods are appropriate in particular cases. In the same vein, the Supreme Court in *Watson v. Fort Worth Bank*, made clear that rigorous statistical validation of selection methods is not required in all cases, but did not elaborate further.[22] This lack of guidance means that the courts vary widely in the standards they apply and retain wide discretion in deciding what kind of evidence satisfies the business necessity defense. Courts sometimes show a deferential attitude towards defendants practices, while others express skepticism, reminiscent of the Supreme Court's doubts in *Griggs* about the relationship between competency tests and job performance, especially for jobs requiring lower levels of skill. Some courts accept relaxed forms of content validation resting on appeals to

---

[21] See, e.g., Kelman, Mark (1991). Concepts of Discrimination in "General Ability" Job Testing. *Harvard Law Review*, 104(6), 1157-1247 (discussing forms of validation); Roth, Philip L., Philip Bobko, and Lynn A. McFarland. (2005). A Meta-Analysis of Work Sample Test Validity: Updating and Integrating Some Classic Literature. *Personnel Psychology*, 58(4), 1009-1037 (same).

[22] See, e.g., *Watson v. Fort Worth Bank* , 487 U.S. 977, 998 (1988)(noting that formal validation studies are infeasible in some cases, and stating that "[e]mployers are not required, even when defending standardized or objective tests, to introduce formal 'validation studies' showing that particular criteria predict actual on-the-job performance."); see also Menjoge, Sujata S. (2003). Testing the Limits of Anti-Discrimination Law: How Employers' Use of Pre-Employment Psychological and Personality Tests Can Circumvent Title VII and the ADA. *North Carolina Law Review*, 82(1), 326-365, at 360 (stating that "despite recommendation by the EEOC that defendants use validation studies to determine whether a test is job related (see 29 CFR 1607.5 (2002)), defendants do not need to provide any formal validation study that psychological or personality criteria predict actual on-the-job performance.")

common sense[23] or to the finding of a "manifest relationship" between pre-employment criteria and successful job performance,[24] while others demand rigorous statistical evidence. It is unclear, for example, how strongly predictive a hiring or promotion criterion must be to survive scrutiny. The law does not specify the required correlations, and the courts have provided no

---

[23] Compare, e.g., *Douglas v. Hampton*, 512 F. 2nd 976 (Dc Cir. 1975) (in black college graduates' challenge to the use of the federal service entrance examination (FSEE), demanding empirical validation in the form of "identifying criteria indicating successful job performance," and then showing "a correlation between test scores and those criteria.") to *Stender v. Lucky Stores, Inc.*, 803 F. Supp. 259, 321-22 (N.D. Cal. 1992) (interpreting 1991 Civil Rights Act to require employer 'to show that its selection criteria bear 'a manifest relationship to the employment in question'").

[24] See *Ass'n of Mexican-American Educators v. California*, 231 F.3d 572, 585 (9th Cir. 2000) (using manifest relationship test to evaluate employer's business necessity defense in disparate impact claim); *Bullington v. United Air Lines, Inc.*, 186 F.3d 1301, 1315 n.10 (10th Cir. 1999) (using manifest relationship test); *NAACP v. Town of East Haven*, 70 F.3d 219, 225 (2d Cir. 1995) (mandating manifest relationship between employment practice and job screens); *Zamlen v. City of Cleveland*, 906 F.2d 209, 217 (6th Cir. 1990) (following standard that uses manifest relationship test); *Davis v. City of Dallas*, 777 F.2d 205, 211 (5th Cir. 1985) (applying manifest relationship standard); *Robinson v. Lorillard Corp.*, 444 F.2d 791, 798 (4th Cir. 1971) (adopting manifest relationship standard).

Some courts have adopted the so-called "Spurlock doctrine," also known as the "demonstrably necessary" test, which has been interpreted as requiring a somewhat higher degree of relationship between criterion and job, although stopping short of demanding actual statistical validation. See, e.g., *Bew v. City of Chicago*, 252 F.3d 891, 894 (7th Cir. 2001) (outlining requirement for business necessity defense, using demonstrably necessary standard); *Anderson v. Zubieta*, 180 F.3d 329, 342 (D.C. Cir. 1999) (employing demonstrably necessary standard for disparate impact cases); *Fitzpatrick v. City of Atlanta*, 2 F.3d 1112, 1118-19 (11th Cir. 1993) (employing demonstrably necessary standard); *Banks v. City of Albany*, 953 F. Supp. 28, 35 (N.D.N.Y. 1997) (adopting demonstrably necessary standard); *Donnelly v. R.I. Bd. of Governors*, 929 F. Supp. 583, 594 (D.R.I. 1996) (using similar standard).

Legal scholars have criticized content validation as too easy to satisfy, while also recognizing that disparate impact doctrine presents the courts with "the unpalatable choice of either requiring businesses to conduct expensive validation studies to establish business necessity or watering down the defendant's burden of proof to the point of meaninglessness." Charles Sullivan, Disparate Impact: Looking Past the *Desert Palace* Mirage. *William & Mary Law Review*, *47*(3) (2005) , 912-1002, 994 (2005), quoting Linda Krieger, The Content of Our Categories, 47 Stanford Law Review 1161, 1232 (1995). See also Sullivan, op cit. at 994 (recognizing that formal validation is often impracticable, but also that formal validation is not required "across the spectrum of disparate impact cases").

guidance on the magnitude of the relationship (or correlation co-efficient) between a job screen and subsequent job performance that satisfies the "job-relatedness" standard. Likewise, neither the Supreme Court nor the EEOC has ever squarely addressed the question of whether an employer can use the most predictive filter or standard available, regardless of whether that choice increases disparate impact compared to a less predictive screen. Relatedly, the law is unclear on whether employers are free to create as competitive a process for staffing the workplace as the market will bear, or whether they must give equal consideration to candidates who have demonstrated a minimum level of ability to do the job.[25] Finally, some courts have compounded the uncertainty by adopting a more exacting approach, based on language in some Supreme Court opinions, ambiguously codified in the 1991 Civil Rights Act, that suggests that an employer can establish a business necessity defense only if no equally valid or job-related selection method with less adverse impact is available.[26] Because satisfying this requirement

---

[25] See, e.g., Samuel Issacharoff and Erin Scharff, Antidiscrimination in Employment: The Simple, the Complex, and the Paradoxical, New York University School of Law, Center for Law, Economics, and Organization, Working Paper No 10-10 (paper on file with author). The authors suggest that disparate impact can be construed to"reduc[e] hiring criteria to the minimum level of competence rather than the most credentialed employees," thereby "den[ying] employers the option of taking advantage of a surplus of overqualified workers and demanding higher-level credentials in their workforce." They cite no support in the statute or the case law for this view of disparate impact's requirements. In fact, Title VII permits employers to "give and . . act upon the results of any professionally developed ability test" that is "not designed, intended, or used to discriminate," and does not otherwise limit the use of such criteria by, for example, forbidding the top-down hiring of the best performers. See Susan Carle, A Social Movement History of Title VII Disparate Impact Analysis, 63 Florida L. Rev. at 285 & note 208 (describing the Title VII language permitting the use of ability tests).

[26] See, e.g. Title VII, Section 105 [703(k)(A)] (1991) stating that an employer violates the disparate impact rule if the plaintiff demonstrates that there is another job-related device available with less adverse impact and the defendant "refuses to adopt such alternative employment practice." Despite language in the 1991 Civil Rights Act suggesting that the plaintiff must demonstrate the existence of the alternative selection method, the courts have been divided on who must prove the elements relevant to this showing. Compare International Brotherhood of Electrical Workers v. Mississippi Power and Light Co., 442 F. 2nd 313 (5th Cir. 2006)(burden on plaintiff) with Bradley v. Pizzaco of Nebraska, 7 F. 3d 795 (8th Cir 1993) (stating that the employer must show a "compelling need ... to maintain that practice ... and that there is no alternative to the challenged practice."). See also Harris, Cheryl I., and Kimberly

amounts to proving a negative, courts that impose this standard make the job-relatedness defense significantly harder to establish.

The vagueness of the law on these key questions adds to the confusion surrounding the business necessity and job-relatedness defense and gives the courts considerable leeway to decide how strictly to apply the job relatedness or business necessity defense.[27] This confusion is reflected in the cases challenging job testing in general, and civil service tests for public servants in particular.[28] In addition, the technical nature of the validation process dictates that establishing the job-relatedness of personnel methods requires parties to present testimony from

---

West-Faulcon, Reading *Ricci*: Whitening Discrimination, Racing Test Fairness, 58 UCLA Law Rev. 73, 160-161 (October 2010) (discussing the "less discriminatory alternative" requirement).

[27] See, e.g., Rutherglen, George, Ricci v. DeStefano: Affirmative Action and the Lessons of Adversity, 2009 *Supreme Court Review* 83, 106-107 (noting ambiguities surrounding the business necessity defense, and stating that "[t]he particular language adopted by Congress [in the 1991 Civil Rights Act] just perpetuates the ambiguity that can be found throughout the opinion in *Griggs* and, it is fair to say, in every opinion of the Supreme Court" on the precise contours of the business necessity defense, which leaves unresolved whether "the employer's burden of justifying a practice with adverse impact is a light burden . . or a heavy burden.")

[28] As a general matter, courts tend to be more forgiving of requirements geared closely to job tasks and more suspicious of tests of general ability (such as the intelligence exams at issue in *Griggs*), but approaches vary. For example, in *Ricci*, the Supreme Court accepted a form of content validation of the firefighter supervisors' test. See Rutherglen, supra 2009 Supreme Court Review, at 107 (with respect to the decision in Ricci, "on the spectrum between heavier and lighter burdens of justification, the Court came down decidedly in favor of a lighter burden)." The lower courts in that case relied largely on expert testimony and evidence concerning the test development process, which was carefully geared to investigating and then assessing the knowledge and skills needed to perform a firefighter supervisors' job. The courts did not demand a statistical demonstration of the tests's ability to predict superior performance. In contrast, lower courts in other cases have faulted the paucity of data on firefighter tests' statistical validity. See e.g.,United States v. Vulcan Society, 637 F. Supp. 77 (E.D. N.Y. 2009) discussed in Heather MacDonald, Fighting Fire with Quotas, Manhattan Institute website, http://www.city-journal.org/2010/eon1024hm.html (noting the district court's grant of summary judgment on the plaintiffs' disparate impact challenge to New York City's firefighters' exams after finding the City's evidence of job-relatedness inadequate). See also ; see also Lewis v. City of Chicago (ND Ill. March 22, 2005)(invalidating firefighters' exam), aff'd 130 S. Ct. 2191 (2010). For cases invalidating firefighter screening requirements and written civil service tests, see Helen Norton, The Supreme Court's Post-Racial Turn Towards a Zero-Sum Understanding of Equality, 52 William and Mary Law Rev. 197, 254 notes 233 & 234 (collecting cases) (2010)

experts in the IOP field, which often issues in a protracted "battle of the experts."[29] All of these factors ensure that defending a business practice with adverse impact will, by definition, be a costly, burdensome and risky process.

In the midst of these uncertainties, two key features of disparate impact doctrine, originating in *Griggs*, have informed the courts' subsequent application of the rule. First, *Griggs* proceeds from the assumption, albeit implicit, that fair and valid personnel practices will result in workers from each race being hired or promoted in rough proportion to their numbers in the population or the appropriately defined pool of job candidates. That assumption finds expression in the subsequent development and wide acceptance of the 4/5 rule for defining adverse impact. By holding an employer presumptively liable if minority hires are less than 80% of hires from the majority group, the 4/5 rule effectively embodies the understanding that the workforce will reflect the racial composition of the background working-age population. Second, in applying the business necessity defense, the Court in *Griggs* noted the absence of evidence that the screening criteria in that case (a high school diploma and scores on an intelligence test) were related to job performance. And the Court's discussion expresses skepticism about whether those requirements would bear any relationship to performance in the relatively unskilled jobs at issue.

As the discussion below shows, however, subsequent work in social science reveals that both these assumptions are incorrect. First, IOP research, as well as work in labor economics and educational demography, indicate that blacks, and to a lesser extent hispanics, currently lag behind whites in the skill sets that predict proficiency in most jobs. Direct measures of job performance ratings also reveal that, on average, these groups perform less well than whites at work. Second, the data indicate that, despite their imperfections, tests and criteria such as those at issue in *Griggs*, which are heavily "g"-loaded – that is, dependent on cognitive ability --

---

[29] See generally the opinions *Ricci v. DeStefano* (discussing voluminous testimony for both sides concerning the content validity of the firefighters' test in that case) and discussion below.

remain the best predictors of performance for jobs at all levels of complexity. The g-dependency of job selection methods both contributes to their usefulness and accounts for their adverse impact. These realities are inconsistent with the key understandings articulated in *Griggs* and carried forward in subsequent disparate impact cases. The tension between the evidence and the founding myths of *Griggs* create practical problems for applying the disparate impact doctrine, and argue for a substantial revision in the rules for disparate impact litigation.

III. *Evidence on disparate impact: industrial and organizational psychology (IOP) research*

In confronting the challenges of selecting and managing their workforce without running afoul of the disparate impact rule, employers have enlisted the assistance of experts trained in industrial and organizational psychology (IOP). In the years since *Griggs*, an explosion of studies and scholarly articles has appeared to address the problems employers face in seeking to comply with the disparate impact's strictures, including avoiding liability under the 4/5 rule and establishing a job-relatedness or business necessity defense in the event of legal challenge. Accordingly, research in this field has become increasingly focused on two goals: first, identifying and refining job selection devices that predict job productivity and thus satisfy the courts' definition of job-relatedness; and developing effective personnel practices that minimize adverse impact on racial minorities.

IOP experts operate on the assumption that the workplace is a competitive meritocracy. Employers' staffing decisions are routinely made under conditions of scarcity: there are usually more applicants for jobs than there are positions available. This means that employers must be selective, and are in a position to pick and choose among available workers. Another convention of the field is that employers have an interest in creating the most productive workforce possible. IOP experts thus assume that employers are interested in finding the best workers for available jobs regardless of race or background. Finally, researchers proceed from the understanding that pursuit of this interest is legitimate and lawful, and that selection methods that are both valid and unbiased – in being equally predictive of productivity for persons from all groups – best advance the meritocratic idea of equal opportunity embodied in law of disparate

impact.

In helping businesses meet disparate impact requirements in light of these goals, IOP experts and psychometricians have generated a large body of empirical research and statistical analysis concerning the validity and adverse impact of various personnel screening devices. In analyzing the implications of their research, IOP experts have generally taken the 4/5 rule as an important benchmark for triggering liability under disparate impact, although many also work with conventional tests of statistical significance. Much effort is devoted to correlating screening scores with ratings and assessment scores of workers in various jobs. Research has also accumulated evidence on group differences in performance on a range of commonly used screens and on the job, as reflected by ratings and evaluations commonly performed in the workplace. In fact, psychometricians and demographers have long been interested in identifying methods for selecting good workers, and research on predicting job performance long predates the decision in *Griggs*. In assessing selection devices, IOP experts try to demonstrate measurable, reproducible, and statistically significant correlations with actual employment outcomes – that is, they focus on formal validation of job screening methods. Establishing the predictive validity of job selection devices rests on the ability accurately to measure job performance. Thus, IOP experts are preoccupied with investigating and developing sound methods for rating workers on the job, with the ultimate aim of devising better instruments for screening job candidates and establishing the predictive validity of those screens.

A. *Job screening methods and predictions of job performance*:

In managing the workplace, employers must routinely decide whom to hire and whom to promote into various jobs.[30] Depending on the nature and selectivity of the jobs at issue,

---

[30] For a review of job selection methods, see Borman, Walter C., Mary Ann Hanson, and Jerry W. Hedge. (1997). Personnel Selection. *Annual Review of Psychology*, 48, 299-337; Schmidt, Frank L. and John E. Hunter. (1998). The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 85 Years of Research Findings. *Psychological Bulletin*, 124(2), 262-274; Hunter, John E. and Ronda F. Hunter. (1984). Validity and Utility of Alternative Predictors of Job Performance. *Psychological Bulletin*, 96(1).

employers rely on a range of methods for recruiting applicants, and for screening and selecting candidates for jobs. In developing a pool of eligible candidates, some use informal methods like word of mouth referrals and personal recommendations, while others make use of more formalized protocols, including extensive advertising and posting of jobs. When it comes to evaluating candidates for hiring or promotion, employers rely heavily on years of education, type of educational experience, and specialized training (collectively known in the field as "biodata"), and then use devices such as job interviews, personality or skill tests, recommendation letters, and other specialized screens to choose among applicants. Entry to higher level jobs is often restricted to persons who have completed high school, college, or graduate degrees. Selection for some positions depends on scores on written or paper and pencil exams, ranging from standardized tests of intelligence, aptitude or cognitive ability, to specialized assessments of job knowledge, competence, or skill (including civil service and professional qualifying exams), to personality or "integrity" tests.[31] Recently, prompted partly by concerns about the adverse impact of conventional paper and pencil tests and assessments that focus on reading and verbal ability, experts have developed various alternative instruments, administered at so-called "assessment centers" that are intended more precisely to mirror actual job requirements. These

---

See also Philip Bobko and Philip L. Roth, An Analysis of Two Methods for Assessing and Indexing Adverse Impact: A Disconnect Between the Academic Literature and Some Practice. pp. 29-49, in Outtz, *Adverse Impact*.

[31] On general ability, or intelligence, tests in personnel selection, see Outtz, James L. (2002). The Role of Cognitive Ability Tests in Employment Selection. *Human Performance, 15*(½), 161-171; Kelman, Mark (1991). Concepts of Discrimination in "General Ability" Job Testing. *Harvard Law Review*, 104(6), 1157-1247; On personality measures, see Barrick, Murray R. and Michael K. Mount. (1991). The Big Five Personality Dimensions and Job Performance: A Meta-Analysis. *Personnel Psychology*, 44(1), 1-26. On job screening interviews, see Cortina, Jose M., Nancy B. Goldstein, Stephanie C. Payne, H. Kristl Davison, and Stephen W. Gilliland, Incremental Validity of Interview Scores over and Above Cognitive Ability and Conscientiousness Scores, *Personnel Psychology* 52 2000, 325-351; Huffcut, Allen I. And Winfred Arthur, Jr., Hunter and Hunter (1984) Revisited: Interview Validity for Entry-Level Jobs, *J. of Applied Psychology* 79(2) 1994, 184-190. On job knowledge or job skills tests, see e.g., Brief for Industrial-Organizational Psychologists as *Amici Curiae* in Support of Respondents, *Ricci v. DeStefano,* 129 S.Ct. 2658 (2009) (Nos. 07-1428 & 08-328).

include task or job simulations, real-time problem-solving exercises, work sample or "in-box" evaluations, and exams administered orally or using audio-visual techniques.[32]

By collecting data on screening methods in a variety of contexts, industrial psychologists have identified the factors most predictive of job performance over a wide range of occupations. Although estimates of the magnitude and relative power of correlations are somewhat sensitive to methodology and parameters specific to jobs at issue, a general consensus has emerged that measures of general cognitive ability – also designated as "g" or IQ -- are generally the best predictors of job performance for all types of jobs, with correlations in the range of approximately .5 or more with measured outcomes.[33] General intelligence has been observed to

---

[32] On methods of job screening, including use of so-called "assessment center" (AC) protocols to gauge applicant's ability to deal with situations commonly encountered on the job, see, e.g., Sackett, Paul and Phillip Lievens, (2008). Personnel Selection. *Annual Review of Psychology*, *59*, 419-450; On job performance simulations and problem-solving exercises, see, e.g., Gaugler, Barbara B., Douglas B. Rosenthal, George C. Thornton III, and Cynthia Bentson. (1987). Meta-Analysis of Assessment Center Validity. *Journal of Applied Psychology*, 72(3), 493-511. On work samples, see Roth, Philip L., Philip Bobko, and Lynn A. McFarland. (2005). A Meta-Analysis of Work Sample Test Validity: Updating and Integrating Some Classic Literature. *Personnel Psychology*, *58*(4), 1009-1037. For a review of alternatives to paper and pencil tests of ability and job knowledge, see Schmitt, Neal, Catherine Clause and Elaine Pulakos, Subgroup Differences Associated with Different Measures of Some Common Job-Relevant Constructs, 11 *International Rev. Of Ind. And Org. Psychology* (1996) 115-139.

[33] See, e.g., Kuncel, Nathan R. And Sarah A. Hezlett, Face and Fiction in Cognitive Ability Testing for Admission and Hiring Decisions, 19(6) *Current Directions in Psychological Science*, 339-345, 339 (2010)("Standardized tests of cognitive abilities . . . are some of the strongest and most consistent predictors of performance in educational and work settings"). See also id. At 341 (noting the substantial correlation between cognitive ability and performance for jobs of high, medium, and low complexity); McKay, Perspectives on Adverse Impact in Work Performance, in Outtz, Adverse Impact, at 253 ("Cognitive ability is the strongest single predictor of job performance"). See also Hunter, John E. and Ronda F. Hunter. (1984). Validity and Utility of Alternative Predictors of Performance. *Psychological Bulletin*, *96*(1), 72-98, 72 (citing cumulative research showing that "for entry level jobs there is no predictor with validity equal to that of [cognitive] ability, which has a mean validity of .53."); Harold W. Goldstein, Charles A. Scherbaum, and Kenneth P. Yusko. Revisiting *g*: Intelligence, Adverse Impact, and Personnel Selection. pp. 95-134, 116 in Outtz, *Adverse Impact*, supra. (Noting reports of correlations of .51-.56 of general ability test scores with job performance, with higher correlations for demanding jobs, and stating the view that "intelligence is the single best predictor of job performance and thus should be afforded special status in the area of personnel

correlate with a variety of functions, such as "learning, memory, grasping concepts, reasoning, problem solving, and more" that virtually all jobs, whether simple or complex, draw upon or require.[34] Thus, to the extent that *Griggs v. Duke Power* suggests that intelligence tests are a poor predictor of performance in relatively unskilled jobs, the evidence belies the Court's skepticism. A similar observation holds for a high school diploma requirement, because years of education achieved are closely correlated with intelligence.[35] In fact, people with higher IQ (and

---

selection."); James L. Outtz and Daniel A. Newman. A Theory of Adverse Impact. pp. 53-94, in Outtz, Adverse Impact, at 68 ("A great deal of empirical evidence has been amassed to support the correlation between cognitive test scores and job performance"); Outtz, James L. (2002). The Role of Cognitive Ability Tests in Employment Selection. *Human Performance*, *15*(½), 161-171, 162 (noting correlation of about .5 between cognitive ability and job productivity); Hunter, John E. and Frank L. Schmidt. (1996). Intelligence and Job Performance: Economic and Social Implications. *Psychology, Public Policy, and Law*, *2*(3/4), 447-472 (expounding on the predictive power of IQ); Ree, Malcolm James and James A. Earles. (1992). Intelligence Is the Best Predictor of Job Performance. *Current Directions in Psychological Science*, *1*(3), 86-89 (discussing data showing that IQ is the best predictor of performance, based on datasets on many thousands of military recruits); Schmidt, Frank L. and John E. Hunter. (1998). The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 85 Years of Research Findings. *Psychological Bulletin*, *124*(2), 262-274, 264 (concluding, based on "thousands of studies conducted over the last nine decades" that general mental ability "has been shown to be the best available predictor of job-related learning,", "job knowledge," and "performance in job training programs," and that "the theoretical foundation for [general mental ability as the best predictor] is stronger than for any other personnel measure."); Harold W. Goldstein, Charles A. Scherbaum, and Kenneth P. Yusko. Revisiting *g*: Intelligence, Adverse Impact, and Personnel Selection. pp. 95-134, 116 in Outtz, *Adverse Impact*, supra at 100 ("A central point of the psychometric perspective is that g is the most important quality that determines success of all types, including at work.") See generally Sackett, Paul R. and Filip Lievens. (2008). Personnel Selection. *Annual Review of Psychology*, *59*, 419-4; Thorndike, Robert L., The Central Role of General Ability in Prediction, 20 Multivariate Behavioral Research (1985) 241-254, 253 ("In the context of practical prediction, 'g' appears to be alive and well."). See also discussion infra comparing intelligence test with other job performance predictors.

[34] See Harold W. Goldstein, Charles A. Scherbaum, and Kenneth P. Yusko. Revisiting *g*: Intelligence, Adverse Impact, and Personnel Selection. pp. 95-134, 97 in Outtz, *Adverse Impact*, supra.

[35] See, e.g., David Rowe, Wendy Vesterdal, and Joseph Rodgers, Herrnstein's syllogism: genetic and shared environmental influences on IQ, education, and income, 26 Intelligence (1998) 405-423 (noting a correlation of .63 between years of education and IQ in a large sample

also more education) will, on average, perform better in any job than people with less measured intelligence or schooling.

A related observation is that job screens that depend more on intellectual ability tend to predict subsequent job performance better than selection methods that place less emphasis on "g". Devices such as integrity tests or personality measures, which are relatively uncorrelated with g, are significantly less predictive of outcomes than cognitive ability tests. Measures of conscientiousness – the personality trait with the highest link to job success – shows a correlation with job performance in the range of .18 to .37, in contrast to the correlation of .5 or more for IQ.[36] In fact, most job selection devices that have proven somewhat useful tap into general cognitive ability. As already noted, the acquisition of educational credentials, including degrees completed and scores on academic tests, are dependent on general intelligence. Interviews, tests of job knowledge or competence, and situational problem-solving exercises also draw on cognitive skills. This is not surprising. Intelligence is highly correlated with the ability

---

of subjects from the National Longitudinal Survey of Youth).

[36] See, e.g., Judge, Timothy A., Chad A. Higgins, Carl J. Thoresen, and Murray R. Barrick. (1999). The Big Five Personality Traits, General Mental Ability, and Career Success Across the Life Span. *Personnel Psychology*, *52*(3), 621-652, 640, 644, 647 (acknowledging that conscientiousness is a less powerful predictor of job performance than cognitive measures of mental ability); Barrick, Murray R. and Michael K. Mount. (1991). The Big Five Personality Dimensions and Job Performance: A Meta-Analysis. *Personnel Psychology*, *44*(1), 1-26 (discussing conscientiousness as a performance predictor); Barrick, Murray R., Michael K. Mount, and Timothy A. Judge, Personality and Performance at the Beginning of the New Millennium: What Do We Know and Where Do We Go Next?, 9 *Int'l J. of Selection and Assessment*, 9-30 (March/June 2001); Ozer, Daniel J. and Veronica Benet-Martinez. (2006). Personality and the Prediction of Consequential Outcomes. *Annual Review of Psychology*, *57*, 401-421 (same); Ones, Deniz S., Chockalingam Viswesvaran, and Frank L. Schmidt. (1993). Comprehensive Meta-Analysis of Integrity Test Validities: Findings and Implications for Personnel Selection and Theories of Job Performance. *Journal of Applied Psychology*, *78*(4), 679-703; Sackett, Paul R. and James E. Wanek. (1996). New Developments in the Use of Measures of Honesty, Integrity, Conscientiousness, Dependability, Trustworthiness, and Reliability for Personnel Selection. *Personnel Psychology*, *49*(4), 787-829; Cortina, Jose M., Nancy B. Goldstein, Stephanie C. Payne, H. Kristl Davison, and Stephen W. Gilliland, Incremental Validity of Interview Scores over and Above Cognitive Ability and Conscientiousness Scores, *Personnel Psychology* 52 2000, 325-351.

to learn, and job knowledge is a function of learning through studying or through job experience. Likewise, scores on interviews are sensitive to verbal ability and analytic acumen, which both have an established correlation with general mental ability. None of these methods has proven as reliable as pure measures of general intelligence, and none predicts job success as well as IQ.[37]

B. *Job screening methods and adverse impact*: *The validity-diversity tradeoff*

In addition to identifying selection methods that predict success on the job, IOP experts have also been charged with developing practices that simultaneously comply with the strictures of disparate impact doctrine. Unfortunately, the goal of maximizing workforce productivity while reducing or eliminating disparate impact has proven elusive. Indeed, the literature consistently demonstrates that, given current social realities, the most effective job selection methods available show a substantial adverse impact on minorities over a wide range of real-world conditions, and in particular tend to screen out blacks. Moreover, the higher the predictive validity of the method used, the greater the racially disparate impact. This inverse relationship between workforce productivity and racial balance is known among IOP experts as the "validity-diversity tradeoff."[38] The consensus is that this tradeoff is a pervasive feature of personnel

---

[37] For example, the correlation between structured job interviews and job performance has been estimated as around .36. The predictive correlation with performance of other methods, including situational assessments, tests of job knowledge, and job task performance, is similar in magnitude. For a survey of methods and correlations, see, e.g., Borman, Walter C., Mary Ann Hanson, and Jerry W. Hedge. (1997). Personnel Selection. *Annual Review of Psychology*, *48*, 299-337; Hoffman, Calvin C. and George C. Thornton III. (1997). Examining Selection Utility Where Competing Predictors Differ in Adverse Impact. *Personnel Psychology*, *50*(2), 455-470; Sackett, Paul R. and Filip Lievens. (2008). Personnel Selection. *Annual Review of Psychology*, *59*, 419-450. See also Schmitt, Neal, Catherine Clause and Elaine Pulakos, Subgroup Differences Associated with Different Measures of Some Common Job-Relevant Constructs, 11 *International Rev. Of Ind. And Org. Psychology* (1996) 115-139, 134 (reporting and discussing validities for various job selection methods).

[38] See, e.g., Pyburn, Keith M Jr., Robert E. Ployhart, David A. Kravitz, The Diversity-Validity Dilemma: Overview and Legal Context, Personnel Psychology 61 (2008) 143-151 ("Unfortunately many of the most predictive knowledge, skill, and ability measures produce varying degrees of mean subgroup differences, with racio-ethnic minority groups usually scoring

practice.

The difficulty of finding selection methods that predict job success while avoiding racially adverse impact is a product of two well-established social science facts. First, as noted, the IOP literature has repeatedly documented that general intellectual ability, or "g", although imperfectly correlated with success on the job, is the most powerful predictor of job performance, over a wide range of occupations, from the least to the most demanding. Thus, as a general rule, the more g-loaded a job screen, the more predictive of job success. As a result, many commonly used job screens have a significant correlation with cognitive ability, and are somewhat "g-loaded."

Second, there is a longstanding gap in the average performance of blacks and whites on measures of general intelligence, or IQ. The magnitude of this disparity, which has been repeatedly documented within the IOP field and has been fairly stable for decades, stands at about one standard deviation from the mean.[39] This gap has important consequences for

---

lower than majority groups."); Keith Hattrup and Brandon G. Roberts. What Are the Criteria for Adverse Impact? pp. 161-197 in Outtz, *Adverse Impact* (describing the validity-diversity tradeoff as a pervasive empirical feature of IOP research and practice); Wilfried De Corte, Weighing Job Performance Predictors to Both Maximize the Quality of the Selected Workforce and Control the Level of Adverse Impact, *J. of Applied Psychology* 84(5) (1999) 695-702, 700 (noting the difficulty of identifying a job selection approach that controls the level of racially adverse impact without "neglect[ing] the goal of maximizing the quality of the selected workforce."). See also Schmitt, Neal, Catherine Clause and Elaine Pulakos, Subgroup Differences Associated with Different Measures of Some Common Job-Relevant Constructs, 11 *International Rev. Of Ind. and Org. Psychology* (1996) 115-139, 134 (describing the pervasive problem in personnel practice of the "conflict between equal representation of different groups and the achievement of maximum expected productivity.")

[39] See, e.g., Hough, Leaetta M., Frederick L. Oswald, and Robert E. Ployhart. (2001). Determinants, Detection and Amelioration of Adverse Impact in Personnel Selection Procedures: Issues, Evidence and Lessons Learned. *International Journal of Selection and Assessment*, *9*(½), 152-194, 153 ("Regarding general intelligence, the commonly accepted mean difference between blacks and whites is about one standard deviation, with blacks scoring lower than whites."); Mark Kelman, Harv. Law Rev., supra, at 1158 ("As a group, blacks score significantly lower on 'general ability' tests than do whites."); Sackett, Paul R. and Steffanie L. Wilk. (1994). Within-Group Norming and Other Forms of Score Adjustments in Preemployment Testing. *American Psychologist*, *49* (11), 929-954, 943 ("Black white differences of approx 1 SD and hispanic

personnel practice and policy. The combination of well-documented racial differences in cognitive ability and the consistent link between ability and job performance means that most valid job selection devices will show a racially adverse impact, and will disproportionately screen out blacks.

In the wake of the *Griggs* decision, cognitive ability tests have been targeted for criticism and litigation based on their adverse impact on black job applicants. Although "pure" tests of ability survive in some quarters,[40] their use has diminished significantly.[41] Nonetheless, minimizing or eliminating the use of general ability tests has not solved the adverse impact problem. Group disparities in other criteria (such as educational "biodata," outcomes of job

---

white differences of approx .6-.8 SD have been widely and consistently reported for measures of cognitive ability. "); Roth, Philip L., Craig A. Bevier, Philip Bobko, Fred S. Switzer III, and Peggy Tyler. (2001). Ethnic Group Differences in Cognitive Ability in Employment and Educational Settings: A Meta-Analysis. *Personnel Psychology*, *54*(2), 297-330 (describing data showing that Hispanics lag behind whites by about .72 standard deviations on standard cognitive tests, as compared to 1 SD for blacks). For a discussion of black-white differences in IQ and standard deviation measurements, see, e.g., Kelman, Mark (1991). Concepts of Discrimination in "General Ability" Job Testing. *Harvard Law Review*, 104(6), 1157-1247. See also Harold W. Goldstein, Charles A. Scherbaum, and Kenneth P. Yusko. Revisiting *g*: Intelligence, Adverse Impact, and Personnel Selection. 95-134, 120 in Outtz, Adverse Selection, supra (discussing black-white differences in performance on tests of general intelligence).

[40] See Philip Crewson, A Comparative Analysis of Public and Private Sector Entrant Quality, 39 American J. Of Political Science (1995), 628, 632-622 (describing the Armed Services Qualifying Test (AFQT) as a general ability test used by the military "not only as a measure of trainability and future performance but also as a general indicator of recruit quality"); Fox, Wayne L., John E. Taylor, John S. Caylor. (1969). Aptitude Level and the Acquisition of Skills and Knowledges in a Variety of Military Training Tasks. Technical Report 69-6. Washington, D.C.: George Washington University, Human Resources Research Office; Ree, Malcolm James and Thomas R. Carretta. (1996). Central Role of *g* in Military Pilot Selection. *The International Journal of Aviation Psychology*, *6*(2), 111-123. For more on armed forces screening tests and their racially disparate impact, see — infra.

[41] See Hunter and Schmidt (1996) at 466 (acknowledging a decline in the use of pure intelligence tests for job screening); see also, e.g., Kelman, supra at 1205 (discussing the General Ability Test Battery (GATB), used for many years by US Employment service to screen and rank prospective workers). See also Douglas v. Hampton, 512 F. 2nd 976 (Dc Cir. 1975)(black college graduates' challenge to the general ability federal entrance service exam (FSEE)).

interviews, or scores on tests of job skills and knowledge), although somewhat smaller than for IQ tests, are still fairly substantial.

In describing ethnic disparities, the IOP literature refers to the standardized ethnic group differences (designated as "d") associated with a given predictor of job performance.[42] Measured d values of screens vary substantially, with the largest gaps reported between blacks and white job candidates and smaller disparities for Hispanics. Values range from the black-white gap of one SD commonly reported for tests of cognitive ability, to significantly lower to negligible racial differences for some kinds of personality tests. In general, however, the d values for commonly used methods are somewhere in between, with more heavily g-loaded, or ability-dependent measures showing greater group differences.[43]

---

[42] "The d statistic is computed by subtracting the mean of the focal minority group from the mean of the majority group in the numerator," with a denominator designated as "the sample-weighted average standard deviation of the minority and majority groups." For example, a d of .5 indicates that the minority group scored on average one half of an average standard deviation lower than the majority group. See Denise Potosky, Philip Bobko, and Philip Roth, "Forming Composites of Cognitive Ability and Alternative Measures to Predict Job Performance and Reduce Adverse Impact: Corrected Estimates and Realistic Expectation," 13 international J. Of Selection and Assessment (2005), 304, 305.

[43] For a review of the racially adverse impact of commonly used screening methods, see, e.g., Schmitt, Neal, Catherine Clause and Elaine Pulakos, Subgroup Differences Associated with Different Measures of Some Common Job-Relevant Constructs, 11 International Rev. Of Ind. And Org. Psychology (1996) 115-139 (reporting a range of values for blacks and hispanics). See also, e.g., Pyburn, Keith M Jr., Robert E. Ployhart, David A. Kravitz, The Diversity-Validity Dilemma: Overview and Legal Context, Personnel Psychology 61 (2008) 143-151; Roth, Philip L., Craig A. Bevier, Philip Bobko, Fred S. Switzer III, and Peggy Tyler. (2001). Ethnic Group Differences in Cognitive Ability in Employment and Educational Settings: A Meta-Analysis. *Personnel Psychology*, *54*(2), 297-330. On the disparate impact of structured interviews, see Huffcutt, Allen I., Philip L. Roth, Racial Group Differences in Employment Interview Evaluations, *J. of Applied Psychology* 83(2) 1998, 179-189; Roth, Philip L., Chad H. Van Iddekinge, Allen I. Huffcutt, Carl E. Eidson Jr., and Philip Bobko. (2002). Corrections for Range Restriction in Structured Interview Ethnic Group Differences: The Values May Be Larger Than Researchers Thought. *Journal of Applied Psychology*, *87*(2), 369-376. On racial differences in problem solving exercises and job simulations, see, e.g., Dean, Michelle A., Philip L. Roth, and Philip Bobko. (2008). Ethnic and Gender Subgroup Differences in Assessment Center Ratings: A Meta-Analysis. *Journal of Applied Psychology*, *93*(3), 685-691, 686 (noting d values in their sample for assessment center performance of .52 for blacks, and .28 for Hispanics

A broad consensus has emerged among IOP experts that group disparities in scores on predictive job screens are not just an artifact of measurement. Rather, the numbers reflect the distribution of developed abilities and human capital in American society. In addition to group disparities in tests of cognitive ability, differences exist in other indicia of learning, aptitude and achievement. The black-white test score gap at all levels of schooling is persistent and substantial, and has been a topic of concern and discussion for some time. For example, according to scores on a 2009 national test of achievement (the National Assessment of Educational Progress, or NAEP), the average black 12th grader reads at an 8th grade level. Blacks are consistently observed to enter elementary school, high school, college, graduate school, and professional education with lower test scores, grades, academic achievement, and proficiency levels.[44]

The differences in ability and achievement have concrete consequences for how well people perform in the workplace.[45] IOP research reveals that the achievement gap does not stop at the schoolhouse door. Although there is disagreement on the precise magnitude of the disparities, with variation based on methodology, type of position, job selectivity, and other

---

compared to whites, which are noted to be smaller than cognitive ability disparities, but sufficient to "trigger disparate impact scrutiny in many cases").

[44] See Hanushek, Eric. How Well Do We Understand the Achievement Gap?, 27 FOCUS (Winter 2010) 5, 6-7 (noting "the huge difference in the achievement of students by race and background" and stating that, as of 2008, "the magnitude of the gap is stunning." Also stating that "performance appears roughly flat for almost four decades."). See also Thomas Espenshade and Alexandria Radford, No Longer Separate, Not Yet Equal: Race and Class in Elite College Admissions and Campus Life (providing data on group differences in scores and academic achievement); see also Amy L. Wax, Race, Wrongs and Remedies: Group Justice in the 21st Century (2009) (discussing racial gaps in educational performance). For a review, see, e.g., Richard C. Hunter and RoSusan Bartee, The Achievement Gap: Issues of Competition, Class, and Race, 35 Education and Urban Society (2003) 151.

[45] See Hanushek, Eric. How Well Do We Understand the Achievement Gap?, 27 FOCUS (Winter 2010) 5, 6 (observing that studies "provide very consistent estimates of the impact of test performance on earnings of young workers," with "even larger returns to achievement . . . for a more age-representative sample.")

parameters,[46] evidence from direct measures of actual job performance shows fairly consistent racial gaps. On average, black workers lag behind whites in job performance by roughly .3, or a bit less than a third of a standard deviation across a spectrum of jobs,[47] with one meta-analysis estimating the gap as between .24 and .39 standard deviations.[48]

These results are not surprising in light of race differences in academic achievement. Many jobs draw on skills learned in school, and factors such as grades and academic mastery are

---

[46] See McKay, Patrick F. and Michael A. McDaniel. (2006). A Reexamination of Black-White Mean Differences in Work Performance: More Data, More Moderators. *Journal of Applied Psychology*, *91*(3), 538-557; see also Patrick F. McKay. Perspectives on Adverse Impact in Work Performance: What We Know and What We Could Learn More About. pp. 249-270, in Outtz, Adverse Impact, 251-252 (2010).

[47] See, e.g., James L. Outtz and Daniel A. Newman. A Theory of Adverse Impact. pp. 53-94. in Outtz, Adverse Impact, at 77, (noting "average race differences around d = .3 for a variety of job performance measures"); See also Sackett, Paul R. and Steffanie L. Wilk. (1994). Within-Group Norming and Other Forms of Score Adjustments in Preemployment Testing. *American Psychologist*, *49* (11), 929-954, 934 (noting reported black-white differences in job performance as approximately .3-.4 standard deviations, with values perhaps higher when corrected for measurement error); Roth, Philip L., Allen I. Huffcutt, and Philip Bobko. (2003). Ethnic Group Differences in Measures of Job Performance: A New Meta-Analysis. *Journal of Applied Psychology*, *88*(4), 694-706 (noting that blacks lag behind whites in measured job performance ratings). See also Sackett, Paul R. and Filip Lievens. (2008). Personnel Selection. *Annual Review of Psychology*, *59*, 419-450; Sackett, Paul R., Matthew J. Borneman, and Brian S. Connelly. (2008). High-Stakes Testing in Higher Education and Employment. *American Psychologist*, *63*(4), 215-227, 223; Frank Landy. Performance Ratings: Then and Now. pp. 227-248, 252, in Outtz, Adverse Impact (summarizing range of studies of racial differences in job performance.). Most data focuses on black-white differences, although evidence on hispanics has also been collected. For data on Hispanics and job performance see, e.g., Sackett and Wilk, supra; Schmitt, Neal, Catherine Clause and Elaine Pulakos, Subgroup Differences Associated with Different Measures of Some Common Job-Relevant Constructs, 11 International Rev. Of Ind. And Org. Psychology (1996) 115-139.

[48] See McKay, Patrick F. and Michael A. McDaniel. (2006). A Reexamination of Black-White Mean Differences in Work Performance: More Data, More Moderators. *Journal of Applied Psychology*, *91*(3), 538-557; see also Patrick F. McKay. Perspectives on Adverse Impact in Work Performance: What We Know and What We Could Learn More About. pp. 249-270, in Outtz, Adverse Impact, 251-252.

predictive of job success.[49] The harsh reality is that more capable people tend to perform better on the job, and that minorities – most notably blacks and to a lesser degree hispanics – are today on average less proficient than whites and Asians across a number of related domains relevant to job success. Those differences are reflected in turn in ratings on criteria that are used to determine hiring and promotion. In other words, the reason that valid job predictors – that is, those that are correlate to some extent with actual measured job outcomes – tend to show a racially disparate impact is that there are real group differences in what is being predicted: actual job performance.

An analysis of the sources of these patterns is beyond the scope of this paper. As one commentator notes, "there are many realistic disadvantages that distinguish racial subgroups and these have some implications for job performance."[50] Factors thought to contribute to underdeveloped human capital among racial minorities include historical discrimination, conditions of upbringing, family structure, poverty, schooling quality, neighborhood, and culture.[51]

In sum, the tendency of valid job selection methods to screen out minorities is traceable

---

[49] See, e.g., Roth, Philip L. and Philip Bobko. (2000). College Grade Point Average as a Personnel Selection Device: Ethnic Group Differences and Potential Adverse Impact. *Journal of Applied Psychology*, *85*(3), 399-406 (discussing predictive power of grades and school achievement for job performance). See, e.g., Sackett, Paul R., Matthew J. Borneman, and Brian S. Connelly. (2008). High-Stakes Testing in Higher Education and Employment. *American Psychologist*, *63*(4), 215-227(noting that tests of developed intellectual ability are valid predictors of academic as well as occupational performance.) For a further analysis of racial gaps in job performance relative to gaps on job screening tests, see infra.

[50] James L. Outtz and Daniel A. Newman. A Theory of Adverse Impact. pp. 53-94, in Outtz, Adverse Impact, at 85.

[51] See, .e.g, Amy L. Wax, The Discriminating Mind: Define It, Prove It, 40 Conn. Law Review 979 (2008), at 20-22 (noting socio-demographic differences between blacks and whites that could account for disparities in skill development and thus lead to average group differences in job performance); Farkas, George. (2003). Cognitive Skills and Noncognitive Traits and Behaviors in Stratification Processes. *Annual Review of Sociology*, *29*, 541-562, at 545-548 (noting the contribution of differential family investments, economic circumstances, cultural style, and other factors to racial gaps in human capital development).

to underlying group differences that correlate with ability to function in the workplace. Regardless of the sources of these differences, the diversity-validity tradeoff reflects an unfortunate reality with predictable consequences for personnel selection – consequences that are hard or impossible to avoid. Although "personnel psychologists have devoted considerable effort to identifying test and test presentation strategies that reduce adverse impact[,]" they have found that "few strategies have eliminated adverse impact."[52]

That has not prevented IOP experts from trying, however. Attempts to circumvent the validity-diversity tradeoff have spawned a voluminous literature. The evidence suggests that supervisors value both "task performance," which relates to carrying out the core requirements of the job, and "contextual performance," or "job citizenship," which depends on cooperative and pro-social behaviors on the job.[53] Task performance is more dependent on cognitive ability, for which racial differences are relatively large, whereas citizenship is a function of other attributes, such as personality and integrity, for which group differences are negligible or non-

---

[52] Sheldon Zedeck, Adverse Impact: History and Evolution, in Outtz, Adverse Impact, at 22.

[53] See, e.g., Hattrup, Keith, Joanna Rock and Christine Scalia, The Effects of Varying Conceptualizations of Job Performance on Adverse Impact, Minority Hiring and Predicted Performance, *J. of Applied Psychology* 82(5) 1997, 656-664, 657 (discussing dimensions of measured job performance, and distinguishing between "task performance" going to the "technical core" of the job versus "contextual performance," which reflects support of the organization's "climate and culture and the display of helping, prosocial, and citizenship behaviors"); James L. Outtz and Daniel A. Newman. A Theory of Adverse Impact. pp. 53-94, 84 in Outtz, Adverse Impact (same); Bobko, Philip, Philip L. Roth, and Denise Potosky. (1999). Derivation and Implications of a Meta-Analytic Matrix Incorporating Cognitive Ability, Alternate Predictors, and Job Performance. *Personnel Psychology*, *52*(3), 561-589, 562 (noting the importance of "contextual performance factors" or "organizational citizenship behaviors"); Harold W. Goldstein, Charles A. Scherbaum, and Kenneth P. Yusko. Revisiting *g*: Intelligence, Adverse Impact, and Personnel Selection. pp. 95-134, 116 in Outtz, Adverse Impact (noting the contribution of "citizenship" to job performance ratings); Kevin R. Murphy, How a Broader Definition of the Criterion Domain Changes our Thinking about Adverse Impact, at 141, in Outtz, Adverse Impact ("The domain of job performance includes a wide range of behaviors, such as teamwork, customer service, and organizational citizenship, that are not always necessary to accomplish the specific tasks in an individual's job, but are necessary for the smooth functioning of teams and organizations").

existent.[54] In addition, minorities lag behind whites in scores on conventional paper-and-pencil tests that draw on reading and written communication skills. Based on these observations, IOP experts have sought to develop novel screens with a smaller disparate impact. These include "tests that were more interactive, behaviorally oriented, and orally or aurally oriented,"[55] such as simulations or real-time problem-solving exercises (as often employed at job assessment centers).[56] Alternatively, researchers have devised sophisticated, multi-step "composite"

---

[54] See McKay, Patrick F. and Michael A. McDaniel. (2006). A Reexamination of Black-White Mean Differences in Work Performance: More Data, More Moderators. *Journal of Applied Psychology, 91*(3), 538-557, 540 (noting larger racial differences for task performance (more ability based) than contextual performance (dependent on extra-role and prosocial behaviors)); Patrick F. McKay. Perspectives on Adverse Impact in Work Performance: What We Know and What We Could Learn More About. pp. 249-270, in Outtz, Adverse Impact, at 253-254 ("criteria that are dependent on cognitive ability will exhibit larger black-white mean disparities than those more contingent on personality").

[55] Schmitt, Neal, Catherine Clause and Elaine Pulakos, Subgroup Differences Associated with Different Measures of Some Common Job-Relevant Constructs, 11 *International Rev. Of Ind. And Org. Psychology* (1996) 115-139, 126.

[56] For a comprehensive review of attempts to develop innovative job screening protocols that maintain validity while reducing adverse impact, see Schmitt, Neal, Catherine Clause and Elaine Pulakos, Subgroup Differences Associated with Different Measures of Some Common Job-Relevant Constructs, 11 International Rev. Of Ind. And Org. Psychology (1996) 115-139.

Job assessment centers have received a good deal of play as a means to circumvent the validity-diversity tradeoff, and are the subject of an extensive literature as well as discussion in the case law. See, e.g., Brief for Industrial-Organizational Psychologists as *Amici Curiae* in Support of Respondents, *Ricci v. DeStefano,* 129 S.Ct. 2658 (2009) (Nos. 07-1428 & 08-328). The research shows that they, although they can somewhat reduce race gaps, they do not eliminate adverse impact. See Gaugler, Barbara B., Douglas B. Rosenthal, George C. Thornton III, and Cynthia Bentson. (1987). Meta-Analysis of Assessment Center Validity. *Journal of Applied Psychology, 72*(3), 493-511; Hermelin, Eran, Filip Lievens, and Ivan T. Robertson. (2007). The Validity of Assessment Centres for the Prediction of Supervisory Performance Ratings: A Meta-Analysis. *International Journal of Selection and Assessment, 15*(4), 405-411; Dean, Michelle A., Philip L. Roth, and Philip Bobko. (2008). Ethnic and Gender Subgroup Differences in Assessment Center Ratings: A Meta-Analysis. *Journal of Applied Psychology, 93*(3), 685-691; Dayan, Kobi, R. Kasten, and Shaul Fox, Entry-level Police Candidate Assessment Center: An Efficient Tool or a Hammer to Kill a Fly, *Personnel Psychology* 55 (2002) 827-849; Arthur, Winifred, Jr., Eric Anthony Day, Theresa L. McNelly, and Pamela S. Edens. (2003). A Meta-Analysis of the Criterion-Related Validity of Assessment Center Dimensions. *Personnel Psychology, 56*(1), 125-153. See also Nancy T. Tippins. Adverse

assessments that place greater weight on personality attributes like conscientiousness or integrity than on conventional measures of cognitive acumen. The goal of this research is to reduce disparate impact while maintaining the predictive power of screening instruments. Although the literature on this quest is technical and detailed, the conclusions can be summarized. As a practical matter, adverse impact cannot readily be reduced without sacrificing either accuracy or predictive validity.[57] Novel methods that de-emphasize written communication and attempt to simulate real job situations still show significant disparities by race. Adding additional predictors or fiddling with their weight has not solved the dilemma, and the hopes held out for job screens that de-emphasize ability and rely on traits like conscientiousness, agreeableness or integrity have not been realized. Although these traits are correlated with good "job citizenship," and thus have some bearing on worker performance ratings, the fact remains that they are significantly less important to overall performance than cognitive ability, and that the

---

Impact in Employee Selection Procedures From the Perspective of an Organizational Consultant. pp. 201-225, 218 in Outtz, Adverse Impact (discussing the greater costs of using assessment centers, which are cumbersome and labor-intensive devices).

[57] See, e.g., Sackett, Paul and Lawrence Roth, (1996), Multi-stage selection strategies: A Monte Carlo Investigation of Effects on Performance and Minority Hiring, *Personnel Psychology 49*(3), 549-572. (noting the limited potential of complex and multi-factorial selection practices to achieve small reductions in group differences without major sacrifices in validity, but only under specialized conditions that are difficult to predict or identify systematically ahead of time; and further noting that most of these practices would still violate the 4/5 rule); Schmitt, Neal, William Rogers, David Chan, Lori Sheppard, and Danielle Jennings, Adverse Impact and Predictive Efficiency of Various Predictor Combinations, 82 *J. of Applied Psychology* 82(5) (1997) 719-730, 723 (in analyzing alternative composite job predictors, reporting that in most instances "d remains high with the addition [to cognitively loaded factors] of predictors with smaller levels of d, and in many cases d for the composite exceeds that of cognitive ability alone"); Wilfried De Corte, Weighing Job Performance Predictors to Both Maximize the Quality of the Selected Workforce and Control the Level of Adverse Impact, *J. of Applied Psychology* 84(5) (1999) 695-702, 700 (noting the difficulty of identifying a job selection approach that controls the level of racially adverse impact without "neglect[ing] the goal of maximizing the quality of the selected workforce").

instruments for measuring them are less reliable and precise.[58] Consequently, combining more cognitively loaded with less cognitively loaded methods can modestly reduce disparate impact without compromising validity only in exceptional cases.[59] Most novel testing techniques can boost minority representation without sacrificing measurable productivity only at very high

---

[58] See Ones, Deniz S., Chockalingam Viswesvaran, and Frank L. Schmidt. (1993). Comprehensive Meta-Analysis of Integrity Test Validities: Findings and Implications for Personnel Selection and Theories of Job Performance. *Journal of Applied Psychology*, *78*(4), 679-703, 694 (noting that personality tests predict job performance with only "moderate" validity); Avis, John M., Jeffrey D. Kudisch, and Vincent J. Fortunato. (2002). Examining the Incremental Validity and Adverse Impact of Cognitive Ability and Conscientiousness on Job Performance. *Journal of Business and Psychology*, *17*(1), 87-105 (finding that selecting for conscientiousness in hiring for a large home improvement organization, although providing some incremental validity when added to cognitive ability, failed to ameliorate the adverse impact associated with that cognitive ability component); Pulakos, Elaine D. and Neal Schmitt. (1996). An Evaluation of Two Strategies for Reducing Adverse Impact and Their Effects on Criterion-Related Validity. *Human Performance*, *9*(3), 241-258, 255 (describing a composite broad-based skill assessment with less disparate impact than standard job tests and other more cognitively based screens, but which still yields significant adverse impact at selective hiring ratios below 80% of applicants; also reporting that "one cannot expect substantial reductions in subgroup differences even when one adds new measures for which there are no or minimal subgroup differences to a [valid] measure that exhibits large subgroup differences."); Paul R. Sackett, Wilfried De Corte, and Filip Lievens. Decision Aids for Addressing the Validity-Adverse Impact Trade-Off. 453-472 in Outtz, Disparate Impact (noting the limited potential for composite and novel predictors combining low and high adverse impact instruments to reduce overall disparate impact and maintain predictive validity); Hattrup, Keith, Joanna Rock and Christine Scalia The Effects of Varying Conceptualizations of Job Performance on Adverse Impact, Minority Hiring and Predicted Performance, *J. of Applied Psychology* 82(5) 1997, 656-664, 660, (noting that, because general intelligence is more strongly related to job performance than any other trait, screens that de-emphasize ability can reduce disparate impact while still maintaining validity only in unselective situations and for jobs that greatly stress "citizenship" over "task performance").

[59] See Hattrup, Keith, Joanna Rock and Christine Scalia, The Effects of Varying Conceptualizations of Job Performance on Adverse Impact, Minority Hiring and Predicted Performance, *J. of Applied Psychology* 82(5) 1997, 656-664 (claiming to find a method that maintains validity while reducing adverse impact below the level that violates the 4/5 rule, but only for specialized jobs (such as sales) that are many times more dependent on citizenship-related factors than g-related factors, and only at very low selection ratios, i.e., more than 80% of applicants selected).

selection ratios– that is, only if most or all applicants are hired.[60]

In sum, it is difficult, if not impossible, to reduce reliance on intelligence-based screens, or to reduce the weight assigned to g-related measures, without compromising the ability to predict who is likely to succeed on the job. Likewise, minimizing adverse impact almost always dilutes the reliability and predictive power of the personnel process. Although cognitive ability is not the only factor that determines job success, it is the most important factor. It is also easy to measure, and the instruments used to gauge it are well-developed, reliable, and precise. Therefore, omitting or downplaying g-related measures in a competitive business environment almost always results in a less productive workforce. In short, efforts to develop new personnel practices that circumvent the validity diversity trade-off have largely failed. For most commonly encountered situations, the tradeoff is virtually unavoidable.

Although the research overwhelmingly supports the conclusion that adverse impacts are unavoidable, the case law shows that resistance to the insights. The argument is repeatedly made that written civil service exams, or existing protocols that heavily weight such exams for hiring and promoting police and firefighters, are unacceptable because those methods are not the best available. Rather, there exist equally valid alternatives with less adverse impact. According to disparate impact doctrine, those alternatives are legally required. The premise operating here is that the diversity-validity tradeoff can be overcome. In the context of litigation over civil service jobs such as firefighting, the focus in on the assertion that selection devices that de-emphasize or fail to measure key constructs – such as "command presence" in the case of firefighter supervisors – are inferior predictors of job success. Alternatively, the use of written tests rather than simulations of actual task performance are presumptively less valid and less able to identify

---

[60] See, e.g., Paul R. Sackett, Wilfried De Corte, and Filip Lievens. Decision Aids for Addressing the Validity-Adverse Impact Trade-Off. pp. 453-472, 455 in Outtz, Adverse Impact; Sackett, Paul R. and Jill E. Ellingson. (1997). The Effects of Forming Multi-Predictor Composites on Group Differences and Adverse Impact. *Personnel Psychology 50*(3), 707-721, 712. See also Schmitt, Clause, and Pulakos, supra (reviewing results for alternative job screening methods).

the most effective workers. For example, plaintiffs and their experts in *Ricci* argued that equally effective but less discriminatory promotional protocols were available to the city, including procedures that placed more emphasis on oral assessments of leadership skills and "command presence," and that made use of job simulations and assessment centers.[61] Nonetheless, no actual empirical support for this proposition was offered in the *Ricci* brief filed by IOP . In particular, no specific studies or data were cited for the assertion that the proposed alternative selection methods were as good or better predictors of firefighter captain performance than the civil service tests at issue in the *Ricci* case.

The notion that the written test in *Ricci* was inadequate for failure to measure key attributes of the firefighter's job is based on a misguided "construct fallacy": the idea that measures of "constructs" or traits peculiarly associated with specific jobs are the best predictors of success in those jobs. Although this notion has powerful intuitive appeal and is often asserted in discussions of disparate impact,[62] it is fallacious. Voluminous evidence, accumulated over a long period, indicates that almost all jobs rely more heavily on general cognitive ability than on other skills. Cognitive ability is the most effective predictor of job success for a broad range of jobs from most to least demanding. Moreover, the psychometric data indicate that this skill is best measured through written tests of analysis and learning. Accordingly, the notion that the best candidates for a fire captain position cannot be identified without gauging attributes like "command presence" or leadership is a product of wishful thinking unsupported by hard data or empirical research. No known studies demonstrate that various measures of leadership or "command presence" are superior to g-loaded tests or screens in predicting success in firefighter supervisory positions, and much evidence suggests the contrary. Nor is there reason to believe

---

[61] See Harris and West-Faulcon, supra, at 155. See also Industrial and Organizational Psychologists Brief in Ricci v. DeStefano. See also Helen Norton, Supra, 52 William and Mary L. Rev. At 220; Ricci v. DeStefano (Ginsburg J., dissenting) at 2072.

[62] See, e.g., Lani Guinier and Susan Sturm, Op-Ed, Trial by Firefighters, N.Y. Times, July 11, 2009, at A19 (stating that paper and pencil tests are not good predictors of later performance in public emergency service type jobs).

that adding measures of leadership to the mix or giving appropriate weight to those skills will decrease adverse impact to acceptable levels while increasing, or at least not sacrificing, validity. The assertion that selection protocols that place substantial weight on measures of leadership or command presence predict fire captain performance better with less adverse impact finds no support in empirical studies and indeed stands in stark opposition to the voluminous literature on the diversity-validity tradeoff. That literature reports on the near impossibility of devising alternatives to conventional written measures of cognitive ability, analytic skill, or knowledge that better predict job success under real world conditions. Indeed, the literature reveals that assessment center batteries, which are repeatedly touted as a superior alternative to written civil service exams and have been widely adopted for screening police and firefighter, generally correlate with performance at the level of about .24-.39, which compares unfavorably with the .5-.6 validities associated with heavily g-loaded screens.[63] And although performance on assessment center exercises shows less adverse impact on minorities, group disparities are still significant.[64]

---

[63] See, e.g., *Ricci v. DeStefano*, 129 S. Ct. 2658, 2703 (Ginsburg, J., dissenting)(noting nearly two-thirds of surveyed municipalities used assessment centers ('simulations of the real world of work') as part of their promotion processes). See also Harris and West-Faulcon, at 155 n. 290.

[64] See discussion above and, e.g., Gaugler, Barbara B., Douglas B. Rosenthal, George C. Thornton III, and Cynthia Bentson. (1987). Meta-Analysis of Assessment Center Validity. *Journal of Applied Psychology*, *72*(3), 493-511; Arthur, Winifred, Jr., Eric Anthony Day, Theresa L. McNelly, and Pamela S. Edens. (2003). A Meta-Analysis of the Criterion-Related Validity of Assessment Center Dimensions. *Personnel Psychology*, *56*(1), 125-153. See also Dean, Michelle A., Philip L. Roth, and Philip Bobko. (2008). Ethnic and Gender Subgroup Differences in Assessment Center Ratings: A Meta-Analysis. *Journal of Applied Psychology*, *93*(3), 685-691, 686 (noting d values in their sample for assessment center performance of .52 for blacks, and .28 for Hispanics compared to whites, which are noted to be smaller than cognitive ability disparities, but sufficient to "trigger disparate impact scrutiny in many cases").

The amicus brief filed by Industrial and Organizational Psychologists in *Ricci* repeats the assertion that, because New Haven firefighters test did not purport directly to measure "command presence" and leadership skills, the city failed in its duty to make use of a more valid but less discriminatory alternative. Although the brief quotes experts opining on this point, it cites no actual studies supporting the alleged superiority of the proposed alternatives to the type

IV. *Squaring the Circle: complying with the requirements of disparate impact*

These observations bode ill for employers' efforts to comply with the disparate impact rule. The present reality of group differences, and the intransigence of the diversity-validity tradeoff, mean that even modestly useful personnel selection criteria will screen out too many minorities most of the time. Achieving racial imbalance will prove difficult or impossible for jobs that are moderately selective, and imposing even mildly demanding skill-based hurdles will often produce a workforce that leaves employers vulnerable to a disparate impact challenge. Of course, being sued does not necessarily translate into liability. But it does force an employer to defend its practices as job-related or consistent with business necessity. The practical difficulties and uncertainties inherent in establishing this defense impose a costly burden on employers– one which, under present social conditions, they should not have to bear.

This IOP literature yields crucial insights into the constraints under which businesses operate in managing their workforce while simultaneously striving to comply with the strictures of the disparate impact rule. Given the magnitude of current group differences in performance on common selection criteria and in underlying proficiency, a broad range of valid and common personnel practices will routinely produce violations of disparate impact standards. Indeed, the IOP research makes clear that all screening methods except the least selective and most weakly predictive will fall short of satisfying the 4/5 rule. Although the 4/5 rule appears reasonable on the assumption of equal job readiness, it is far too stringent in light of the actual distribution of human capital. Performance gaps are currently so large that only a substantial narrowing of existing disparities would alleviate this situation. As stated by leading researchers in the field, it is "informative, although perhaps disheartening, to note that even d [group performance difference] values commonly viewed as small (e.g., .2) can produce violations of the 4/5 rule at a

---

of paper and pencil civil service tests administered by the city in that case, and adduces no evidence for the proposition that such alternatives select more effective supervisors. Rather, the experts and plaintiff rely heavily on the fact that many other cities make use of these alternatives. This does not, however, demonstrate superior validity.

variety of commonly occurring [job] selection ratios." [65]

These observations can be demonstrated through simple calculations from data available in the IOP literature. Paul Sackett and his colleagues have performed such calculations and assembled them in a table that summarizes the relationships between relevant parameters.[66] See Figure 1A. The table sets out the expected ratio of hiring from a minority group (for example, blacks) relative to a majority group (whites) as a function of the selectivity of a job (percentage hired relative to applicants) and the performance of the blacks compared to whites on a job screening test. The goal is to select persons who are equally qualified, in the sense of exceeding a threshold level on a job screen, regardless of group identity. When blacks and whites differ in the distribution of screening scores, the hires from each group, and the corresponding hiring ratios, can be determined from the pattern of two overlapping curves. Those numbers will in turn be a function of the positions available. Figure 2 presents an example of a possible distribution, with the distance between curves reflecting the d value, or SD difference between groups on a specific screen, and the numbers hired (above the cutoff) reflecting positions available versus applicants. (Smaller d values will show curves with more overlap; larger d values will show curves farther apart).

Corresponding to these ratios depicted, the numbers in the body of the table represent the fraction of hires and the minority/majority hiring ratio (the numbers in bold type)) expected for a designated minority group as a function of (1) the observed value of d, or the average standardized group difference in performance on a job screen (left side of the table); and (2) the

---

[65] Sackett, Paul R. and Jill E. Ellingson. (1997). The Effects of Forming Multi-Predictor Composites on Group Differences and Adverse Impact. *Personnel Psychology 50*(3), 707-721, 709.

[66] See, e.g., Paul R. Sackett, Wilfried De Corte, and Filip Lievens. Decision Aids for Addressing the Validity-Adverse Impact Trade-Off. pp. 453-472, 455 in Outtz, Adverse Impact; See also Sackett, Paul R. and Jill E. Ellingson. (1997). The Effects of Forming Multi-Predictor Composites on Group Differences and Adverse Impact. *Personnel Psychology 50*(3), 707-721, 712.

selectivity of the job, as measured by the percentage of majority group job applicants who are hired or promoted through the use of a selection device (top of the table). In sum, the table identifies the minority/minority hiring ratios that would be expected in situations of varying selectivity or competitiveness as a function of observed group differences in performance on job screening tests if the employer seeks workers with similar minimum qualifications. Although the table does not incorporate data on the predictive value of particular job screens, the relationships calculated hold good for all devices, regardless of validity. (As discussed above, data on the validity of conventional screens is available in the literature; alternatively, correlations can be established through studies conducted in particular workplaces.)

An examination of this table yields some important insights. First, for any degree of disparity (d) in performance on a job test between groups, the expected ratio of minority group (black) to majority group (white) hires is a function of the selectivity of the job (as reflected in the percentage of job seekers hired from the majority group), which in turn depends on the overall number of positions available relative to applicants. As competitiveness increases (with fewer slots relative to job seekers), the expected ratio of black to white hires declines, and adverse impact increases.[67] Second, the ratio of blacks to whites hired varies with d, or the gap between whites and blacks in the pertinent qualification, such as the score on a job screening test. The greater the discrepancy – that is, lower the average performance for blacks compared to whites -- the lower the black-white hiring ratio and the greater the adverse impact.

This analysis reveals an important defect in the 4/5 rule as currently applied: the existing benchmark does not vary with job selectivity – that is, the ratio of those applying to those hired. A fixed ratio makes sense under the assumption implicit in *Griggs* and its progeny, which is that

---

[67] It is well-recognized that placing significant weight on g-related predictors, especially for competitive jobs where a relatively small percentage of applicants are hired, will severely limit or even eliminate minority hires. See, e.g., Neal Schmitt and Abigail Quinn, Reductions in Measured Subgroup Mean Differences: What Is Possible? pp. 425-451, 426, in Outtz, Adverse Impact (noting that "selecting the highest-scoring individuals and a small proportion of applicants . . . will virtually eliminate members of lower-scoring subgroups.")

there are no discrepancies in expected job performance between groups. If that is the case, job competitiveness will have no effect on the relative number of hires from each category. (If all groups are equally well qualified on average and will perform equally well in most available positions, there is no need to vary the expected ratio with selectivity). As noted, the empirical evidence reveals that the assumption of equal job readiness is unrealistic. When one group lags behind another in actual performance, selectivity can significantly influence the expected hiring ratios from each group. Thus, a fixed ratio is an inappropriate benchmark for creating a presumption of unlawful (that is, arbitrary, non-job-related) disparate impact, because the probability of such adverse impact will vary depending on real-world disparities in performance.

The chart also shows that the 4/5 rule will routinely be violated for valid – and thus presumptively legitimate -- job screens in common use. The zigzag line superimposed on the chart separates the combinations that satisfy the 4/5 rule (above and to the right) from the more numerous ones that fall short (below and to the left). Except for the least selective positions, the group performance disparities reported in the IOP literature for many job screens will generate hiring ratios below 4/5. At the extremes – for example, using a pure test of cognitive ability generating a typical black-white disparity (d) of 1 standard deviation – only positions for which more than 95% of majority applicants (that is, virtually *all* applicants) are hired would achieve the target ratio. When fewer applicants are hired, the ratio of black to white hires expected would always fall below the .8 fraction dictated by the 4/5 rule. Thus, hiring from the top down based on such a test would virtually always expose an employer to a possible disparate impact challenge.[68]

As noted above, many selection devices show smaller group disparities than cognitive ability tests. Nonetheless, the data reveal that the adverse impact for other commonplace

---

[68] See also Schmitt, Neal, Catherine Clause and Elaine Pulakos, Subgroup Differences Associated with Different Measures of Some Common Job-Relevant Constructs, 11 International Rev. Of Ind. And Org. Psychology (1996) 115-139, 116 (noting that "with [job] selection ratios of .1, .5, and .9, a subgroup standardized predictor difference of 1.00 [d=1 SD], the proportion of a lower scoring group hired would be .013, .159, and .61 respectively.")

methods is still large enough to run afoul of the 4/5 rule on a regular basis. For instance, black-white d values as low as .25 have been reported for structured interviews.[69] For a difference of this magnitude, the 4/5 ratio is achieved only if enough positions are available to hire more than 50% of majority job candidates. A d value reported for educational credential ("bio-data") screens is about .33.[70] To meet the 4/5 criterion, 3/4 of job seekers must be hired. Finally, one meta-study reports an average black-white difference on job sample simulations of about .38 standard deviations.[71] Satisfying the 4/5 rule would require hiring more than 75% of majority job candidates. These numbers indicate that compliance with disparate impact targets is possible only for relatively unselective positions. As summarized by a review article on disparate impact compliance, "clearly optimal use of valid selection tests that demonstrate subgroup differences cannot occur without significant impact on the realization of equal representation of different groups."[72]

Of course, triggering a presumptive disparate impact violation is not equivalent to being held liable. The defense of job-relatedness is still available to employers who are convinced that their job selection methods are valid and important to the conduct of their business. As discussed in greater detail below, however, that possibility is cold comfort. Operating in the shadow of potential disparate impact liability generates considerable unfairness, imposes ongoing burdens, and encourages evasions and distortions of the underlying purposes of the doctrine.

---

[69] See Huffcutt and Roth, J. Of Applied Psychology 1998 at ---.

[70] Bobko, Philip, Philip L. Roth, and Denise Potosky. (1999). Derivation and Implications of a Meta-Analytic Matrix Incorporating Cognitive Ability, Alternate Predictors, and Job Performance. *Personnel Psychology*, *52*(3), 561-589. See also Schmitt and Quinn, Reductions in Measured Subgroup Mean Differences: What Is Possible? pp. 425-451, 437-438 in Outtz, Adverse Impact.

[71] See Schmitt, Clause and Pulakos, supra at 119.

[72] Schmitt, Clause, and Pulakos, at 116.

V. *Reforming disparate impact: Alter or Abolish It:*

The IOP literature indicates that the present law of disparate impact is flawed. Although legal scholars have noted the uncertainties created by the doctrine's ambiguities, its disappointing record as an instrument for vindicating employee rights,[73] and the onerous requirements it imposes on employers seeking to defend their personnel practices, they have paid little attention to the nuts and bolts of personnel selection, job performance, and group differences. Attention to those aspects reveals that the empirical assumptions underlying *Griggs* and its progeny are inaccurate. In light of these shortcomings, this section proposes a significant reform in the doctrine as applied or, in the alternative, a wholesale repeal.

The most important defect in the current disparate impact regime is that it incorporates labor market assumptions that fail to square with reality. As noted, the doctrine as currently constructed operates on the implicit premise that persons from each racial and ethnic group are similar in their skills, aptitude, or productivity. This is not the case at present. The reality of racial gaps virtually guarantees that employers seeking to maximize job productivity will generate a workforce that falls short of the stringent requirements of the 4/5 rule or that shows a statistically significant imbalance in workforce composition by race. Those employers' practices will be vulnerable to challenge and to the requirement of demonstrating business necessity or job-relatedness.

Under current doctrine, disparate impact litigation is expensive, risky, and complex. Because commonly used personnel methods routinely produce adverse impacts, many employers' procedures are subject to challenge. The need to establish the defense of job-relatedness places businesses at the mercy of the ambiguous rules and unsettled standards. Formal validation is costly and often practically impossible, and the extent to which courts will demand such validation is unpredictable. Efforts to establish content validation frequently

---

[73] See, e.g., Selmi, Michael. (2006). Was the Disparate Impact Theory a Mistake? *UCLA Law Review, 53*(3), 701-782.

depend on expert testimony, which is expensive and resource-intensive. The outcome of any validation inquiry is uncertain. These burdens encourage businesses to engage in inefficient, counterproductive, and potentially illegal maneuvers designed to avoid disparate impact challenge. Reforming disparate impact to relax the criteria for presumptive liability, or abolishing disparate impact the doctrine altogether, will reduce wasteful litigation and lower incentives to engage in undesirable self-protective measures. Neither is likely to have much effect on workplace opportunities or the racial composition of jobs, since these are mainly the product of deeper social forces.

A. *Alter It: disparate impact realism*

The proposal advanced here is that disparate impact law should be revised to reflect group disparities currently measured in actual on-the-job performance. The data points the way toward a disparate impact standard that is more in line with current demographic facts as well as with the doctrine's stated purpose. As already discussed, the best evidence available estimates that blacks and whites differ, on average, by roughly .3 to a third of a standard deviation in performance outcomes for most jobs, with a reported range between about .24 and .39 SD. That observation suggests, first, that screening practices currently in use are generating a workforce in which blacks lag somewhat behind whites, but with gaps of smaller size than for many valid personnel screening tests. It also indicates that businesses and employers have not succeeded in narrowing these discrepancies. Although, as discussed more extensively below, the reasons for this situation are speculative and complex, some implications emerge.

First, in light of the evidence of real underlying differences in the average productivity in blacks and whites, there is no reason to believe that businesses are discriminating against, and thus disproportionately excluding, more able minority workers. It is thus unlikely that the status quo is the product of invidious discrimination or of businesses erecting arbitrary barriers to minority employment.

Second, on the assumption employers are trying to operate as a competitive meritocracy geared towards maximizing productivity, it suggests that they are missing the mark. Current

levels of racial balance (whether disparate impact compliant or not) are being achieved through the employment of minorities who are, on average, somewhat less productive than whites. This situation is out of synch with the meritocratic ideal that lies at the heart of the disparate impact rule. A perfectly functioning meritocratic system should ideally show no racial gaps in job performance, since persons should be selected for jobs based on their capacities regardless of their group identity. But this is not what is happening today.

Third, the existence of gaps favoring the lower performing group indicates that a legal standard geared to an expectation of equal representation by race (or its close proxy, the 4/5 rule) is too stringent. Although legal expectations are probably not the only factor contributing to this "overshoot" in minority hiring, they might well contribute to it. And the status quo evidence shows that current disparate impact doctrine is not alleviating its effects.

These observations suggest that the disparate impact standard should be dialed back. They also suggest a plausible strategy for this retrenchment. The proposal here is to adopt a regime of "disparate impact realism." Under this reform, the dominant standard for triggering disparate impact liability under Title VII would be relaxed to reflect real, concurrent group differences in *actual* job performance (as opposed to performance differences on job screens). The expectation is that the rule would move the pattern of job performance between groups closer to parity by allowing employers to be somewhat more selective in hiring or promoting minority workers. By gearing staffing to observed patterns of on-the-job success, realism attempts to erase racial disparities through an appropriate downward adjustment in expected hiring ratios in some cases. Ideally, the would cause race differences in job performance to narrow and or even converge. Whether or not this would happen (and, for reasons explained below, it might not), the rule would still be desirable because it would decrease employer's vulnerability to a disparate impact lawsuit.

How would realism work? The first step is to make the assumption that, if staffing is proportionate to the racial composition of the background population (or appropriate pool of job candidates), the workforce will display group (e.g., black/white) differences in actual job

performance that match current measured disparities. The second step is relax the assumption of racially proportionate workplace representation, and to calculate the group hiring ratios that would be *expected* under a meritocratic (valid and predictive) selection system as a function of current productivity differences and level of job selectivity. The key idea here is that employers are entitled to hire workers who will be equally productive, regardless of race – a concept central to the competitive meritocracy. Thus, the ultimate objective of the rule is to produce a cohort of job-holders who will be matched on their projected job productivity, regardless of race. In other words, the goal is to try to close the on the job performance gap among actual workers. Given current patterns, this will necessarily involve hiring fewer blacks than whites from the candidate pool. In many cases, these ratios may fall significantly short of the standard 4/5 ratio.

According to figure 1A, discussed above, the expected hiring ratios for different groups can be calculated as a function of two parameters: group differences in performance on a (presumptively) valid screening test and the selectivity of the position in question. A critical insight, however, is that these relationships between expected hiring ratios and job selectivity also hold good when the parameter of group differences (d) in *actual* job performance is substituted for group differences on job screening tests (which are merely predictive of performance).[74] Assuming that the d value represents the gap in *actual* expected job performance, rather than in performance on a job screen, the expected ratio of minority to majority hires for positions of varying selectivity (corresponding to the slots available) can be seen through the example depicted in Figure 2. On the (highly stylized) assumption that everyone over a threshold level is equally capable of doing a job, the goal is to select persons from each group who will perform the job equally well. When two groups differ in the distribution of their job performance, the hires from each group, and the hiring ratios, can be

---

[74] This application is adumbrated by Sackett and Wilk, at 930 and 934 (noting that, for a designated black-white difference in job performance, "we can determine the number of white and black applicants who would be hired under different scenarios . . . if one were able to select on the basis of actual job performance.")

determined from the pattern of two overlapping curves. Those numbers (above the vertical cutoff) will in turn be a function of the positions available.

The values corresponding to these numbers can be identified from the chart. See figures 1A and 1B. The tabulation in figure 1B reveals that, given a standardized group difference in actual productivity of about 1/3 of a standard deviation (roughly the black-white difference reported in the IOP literature), the ratio of minority to majority hires expected (as adjusted for each group's representation in the applicable candidate pool) would range from approximately .4 to .99, depending on the selectivity of the personnel process in question. For example, for a job in which 25% of the majority applicants were hired, a black-white gap of .3 in expected productivity would predict a ratio of blacks to whites hired of no more than .66. Although not negligible, this is significantly lower than the ratio (.8) that satisfies the 4/5 rule. Likewise, for a d value of .4 where 25% of white applicants are hired, the expected black/white ratio of persons hired (adjusted for applicant population) would be .57, which is also well below .8. For more competitive jobs, the requirements would be even less stringent. For example, if the position allowed the hiring of only 10% of majority candidates, the ratio of blacks to whites hired relative to the candidate population for each group, would be .between .46 and .57 – that is, one would expect that roughly half as many blacks as whites would occupy a given position. Under disparate impact realism, any process that achieved at least that result (that is, produced a ratio of blacks to white hires of this magnitude, or higher), would be immune from disparate impact challenge. See figure 1B (hiring ratios within the dotted line box would satisfy the rule.)

It is important to note that the implicit assumption behind disparate impact realism is that the actual distribution of each group's performance ability in the background job-eligible population corresponds to the distribution, and group differences, in job performance *actually* observed. This is a highly conservative assumption since, as explained below, job incumbents have a narrower range of abilities than the background population. There is, however, no ready

method (apart from random hiring and ex post screening)[75] to determine the performance profile of an unscreened population for particular jobs. Thus, the assumption that job incumbents match the background population on this measure is a conceit that is, if anything, highly favorable to under represented groups.

Despite these limitations, disparate impact realism possesses signal strengths. First, the approach abandons reliance on a fixed target and replaces it with a sliding scale of staffing ratios that better reflects the expected racial composition of the workforce under competitive conditions. Although the sliding scale threshold is somewhat more complicated than the 4/5 rule, it is not significantly harder to apply. The appropriate ratio of hires from each group can be determined as a function of data that IOP experts have collected, and could continue to collect. These data are available in the standard literature, and could be made more readily available in forms accessible to employers, prospective litigants, and judges. An employer faced with a challenge to particular staffing practices for specified positions could refer to the expected hiring ratios provided in the chart. Using this information, employers could determine if they were in compliance with applicable standards, and prospective plaintiffs could too. As with current law, an employer whose hiring patterns met the revised adverse impact targets would escape liability altogether. If he were sued, the case could be expeditiously resolved through summary judgment. Any additional complexity is more than compensated by the narrower scope of liability facing employers in some instances. Indeed, the most compelling reason to adopt this reform is to potentially reduce the number of workplace practices that are vulnerable to a disparate impact challenge and thus expose employers to potential liability. For cases that satisfy revised criteria but would previously have fallen short of the 4/5 rule, an employer's task is simplified. He can invoke evidence of group job performance disparities. He need not establish validity, which requires showing a correlation between job performance and job screens.

---

[75] See discussion below.

To be sure, disparate impact realism does not give employers a completely free hand – far from it. Hiring ratios that fall short of the revised standards are still subject to challenge. And, as figure 1B reveals, in many cases the standard will remain the same (although it is worth observing that d values greater than about .5-.6 are not observed for most commonly used job screens). And even when expected hiring ratios do fall significantly below those mandated by the 4/5 rule, they are still substantial. For example, for a job in which a quarter of white candidates are hired, the expected ratio of black to white hires relative to applicants is .66. If those ratios are not achieved, presumptive liability would apply as with current law, and the employer would have to defend its practices as job-related. Moreover, employers may still struggle to achieve target ratios. That is because, as noted, the most valid job selection methods in current use (such as job interviews, or reliance on educational credentials or skill proficiency) produce group disparities that can exceed the average gaps in productivity currently observed on the job. Thus, some hiring ratios generated using commonplace screens may still fall short of the patterns expected under disparate impact realism.

These observations suggest that disparate impact realism is not a radical change, and that its effects on law and practice cannot be precisely predicted.[76] Nonetheless, the hope is that it will broaden the safe harbor for at least some segments of the labor market and relieve some employers of the burden of demonstrating that their practices are job-related. This is all to the good, because there is no evidence (and indeed, given blacks' lower average job performance, evidence to the contrary) that minorities are being arbitrarily excluded from jobs. Although the proposed revisions appear modest overall, the new standard will make the most dramatic difference for the most selective jobs. The more candidates apply for a position, the fewer minorities will have to be hired. For example, if only 5% of white candidates are hired, the expected black-white hiring ratio relative to candidates is .52. For 1%, the ratio is .4. See figure 1B. These ratios, although not negligible, represent a significant dialing back of the 4/5

---

[76] See additional discussion of predicted effects, infra.

standard. While the most selective jobs are often the most prestigious and skilled, that is not always the case. Even low level jobs sometimes have many more applicants than slots. And competition for police and firefighter positions – which do not require an advanced education – is often fierce, with many seeking the positions available. As figures 1A and 1B indicate, if one group lags behind another in performance on a screening test, the adverse impact on that group will increase as the percentage of candidates who are hired declines. That is likewise true if one group lags behind another in actual job performance. Figure 1B. For the most competitive positions, the ratio of blacks to whites hired will be relatively small and black job-holders will be scarce relative to whites. The data shows that this is to be expected, so such patterns should not be actionable, let alone result in liability. Under disparate impact realism, employers would be immune from challenge for more outcomes that are in line with expected workplace patterns.

Another reason to adopt this approach is that the disparate impact liability doctrine now operates in the shadow of potential Constitutional difficulties. In *Ricci v. DeStefano*, the City of New Haven invalidated a promotional test for firefighters after learning that too few blacks passed the test. The City explained the action as an effort to avoid a disparate impact lawsuit. White firefighters challenged the invalidation as racially motivated, claiming violations of Title VII and the Fourteenth Amendment's Equal Protection guarantee. The Supreme Court ruled that the city's decision to throw out the test constituted unlawful race-based disparate treatment because the city lacked a "strong basis in evidence" to believe that the promotional test would violate the disparate impact doctrine. In suggesting (without deciding) that the Equal Protection ban on intentional discrimination might limit the race-conscious steps employers (whether public or private) can take to avoid violating the disparate impact rule, the *Ricci* decision potentially casts a shadow over disparate impact doctrine, at least in some circumstances, and adds to employers' uncertainty.[77] The specter of unconstitutional action is a compelling reason to

---

[77] See *Ricci v. DeStefano*, 127 S.Ct. 2658 (2009) (Scalia, J., concurring)(noting that "the war between disparate impact and equal protection will be waged sooner or later, and it behooves us to begin thinking about how - and on what terms - to make peace between them").

narrow the doctrine's ambit and reduce the number of situations vulnerable to disparate impact challenge.

Third, compared to the current rule, disparate impact realism better vindicates the core purposes of the doctrine, is more efficient, and is fairer to employers. Current data shows that employers are not arbitrarily screening out minorities from jobs for which they are competitive. Although, as discussed below, there are several possible explanations for existing job productivity gaps by race, one is that employers may, in response to legal mandates (or other social influences), be overshooting the mark by putting a thumb on the scale of diversity, or altering selection practices to achieve greater racial balance. But, as noted, the disparate impact rule does not require bending over backwards for some groups, and this outcome is actually at odds with its stated purpose. In fact, as noted, the situation that bests comports with the goals of disparate impact is a workforce that is equally productive regardless of race. That is what realism aims to achieve.

Second, given the demographic and workplace reality, virtually every employer today operates in the shadow of a potential disparate impact liability. Personnel practices that produce an adverse impact are pervasive. Plaintiffs can easily establish a prima facie case and employers are widely vulnerable to suit. Virtually no aspect of the doctrine surrounding the business necessity defense is well-established, and uncertainties abound. Regardless of the standard applied, it is cumbersome, difficult, and expensive to establish business necessity or job-relatedness. Thus employers with a racially lopsided workforce always face a significant risk of liability. Those seeking to avoid the expense of being sued and the uncertainties that surround litigation are put to unattractive choices. They can abandon the most valid hiring and promotion practices, lower standards across the board, or engage in covert affirmative action, or "race-norming."

See also Primus, Richard. (2010) The Future of Disparate Impact. *Michigan Law Review*, 108(8), 1342-1387 (discussing the constitutional implications of Ricci).

If a court demands formal validation of methods with adverse impact, the task may be insurmountable. First, the proper method for demonstrating predictive validity is far from straightforward, and establishing the necessary correlations requires statistical sophistication and an awareness of methodological pitfalls and controversies. Second, to meet accepted standards, employers must accumulate a large body of data. They must collect and document the scores, ratings, or credentials of job candidates or the appropriate pool of the work-eligible population. The performance of workers who are actually hired must be measured using reasonably reliable methods (or methods that are reasonably resistant to challenge), and those ratings compared to screening outcomes. In addition, correlations for majority and minority group workers must be separately compiled and analyzed, with proper corrections and adjustments for range restriction[78] and other study design limitations. Ideally, this analysis will be conducted separately for distinct jobs. Even if feasible, this process is, at best, cumbersome, expensive, and riddled with the potential for error. Employers are often forced to engage expensive experts to advise them on their practices and to assist them in satisfying data-intensive legal standards.

In addition, outcomes are uncertain. The courts have left many questions surrounding the business necessity defense unaddressed or unresolved. Little guidance is provided on when different standards of job-relatedness are appropriate, and judicial practice in this area is erratic and unpredictable. Indeed, the spate of challenges to civil service exams for police and firefighters – which are designed to reflect the specifics of the job and thus are good candidates for content-validation -- demonstrate that there is no safe haven. Some courts have accepted content validation for civil service exams, whereas others have demanded a more rigorous showing.[79] In any event, content validation, which often rests on the testimony of experts, is

---

[78] See discussion of range restriction in [text and note 92] below.

[79] Compare *Ricci v. DeStefano* (where the Supreme Court accepted the content validity of the firefighters' test at issue in that case) with *Vulcan Society v. City of New York* (see MacDonald article, supra) where the district court judge rejected content validity and faulted the absence of formal validation. See note – supra.

cumbersome and costly. As the recent litigation in *Ricci v. Destefano* reveals, proving that a job selection test is facially valid can be a protracted, complex, and expensive process that consumes considerable private and judicial resources. In addition, the 1991 Civil Rights Act has been interpreted to suggest that even a highly valid job screen may fail to pass muster unless it can be shown that no other equally valid method produces less disparate impact.[80] This is a very tough standard to meet, if only because it is difficult to anticipate the alternative methods opponents might propose. In sum, the ambiguities surrounding the meaning of the business necessity defense, and the uncertainties as to which standards the courts will apply, expose employers to significant burdens and risks. For this reason, companies and businesses are loathe to become embroiled in disparate impact disputes.

This reluctance creates a strong incentive for employers to sacrifice validity to diversity by relaxing standards or by reducing the stringency of personnel selection across the board. Because this can result in a decrement in worker quality and a less effective workforce, firms would prefer to avoid this strategy. But the costs of litigation are so high that employers may nevertheless choose this option. Alternatively, firms can switch to more ad hoc or haphazard methods of selection that will serve as a cover for informal affirmative action or "race-norming" – that is, applying different standards, criteria, or cut-offs across racial groups.[81] Such race-conscious practices are dubious under Title VII, and, for public entities at least, arguably run

---

[80] See discussion supra, and notes --.

[81] See, e.g., Rutherglen, supra, 2009 Supreme Court Review, at 107 (noting that "[t]he heavier the burden of justifying practices with adverse impact, the more likely an employer is to respond to the threat of liability by eliminating the adverse impact, and the easiest way to do this is by engaging in affirmative action"). See also Samuel Issacharoff and Erin Scharff, Antidiscrimination in Employment: The Simple, the Complex, and the Paradoxical, New York University School of Law, Center for Law, Economics, and Organization, Working Paper No 10-10 (paper on file with author) at 6 (noting that, "as disparate impact took hold," employers "turned increasingly to affirmative action to buttress the representation of historically excluded groups").

afoul of the Equal Protection guarantee.[82] Further, in the wake of the Supreme Court's decision in *Ricci v. DeStefano*, public employers have limited leeway to take race-conscious steps to avoid liability for workforce racial imbalances, and those restrictions may extend to private employers also.[83] Nonetheless, informal affirmative action or other race-conscious action is hard to detect and prove. This means that employers have some leeway to engage in various strategies to align minority hires with disparate impact targets. By scaling back the scope of unjustified potential liability, disparate impact realism will lower employers' incentives to engage in such potentially inefficient or unauthorized strategies.[84]

---

[82] As noted, Congress amended Title VII of the Civil Rights Act in 1991 to outlaw race-norming or racial adjustments in scores on job tests. See Civil Rights Act of 1991, Sections 106, 107(a) and note [12] supra.

[83] For a discussion of these issues, see Primus, Richard. (2010) The Future of Disparate Impact. *Michigan Law Review*, 108(8), 1342-1387. See also discussion below.

[84] [? **Move this note up**] It may be claimed that a proper application of the disparate impact rule will ameliorate this entire dilemma. Defining and adjusting the pool of potential job candidates to contain only "qualified" individuals will dramatically narrow the situations in which adverse impact even occurs, thus reducing or obviating the need to prove business necessity. If the baseline candidate pool is appropriately narrowed as a function of factors like geography, local demographics, and standard job credentials (such as years of education and experience, or specialized training, or licensing requirements in a particular field), and the baseline benchmark adjusted accordingly, the 4/5 rule will prove much easier to meet. The assumption is that expected group performance disparities can be erased by controlling for common threshold job requirements, because, regardless of group membership, applicants with similar education and background should not differ significantly in their ability to do a job. Thus, using such threshold requirements should control for group disparities.

There are several problems with this line of reasoning. The main sticking points, as already discussed, is that the law is unclear on this composition of the relevant pool against which disparate impact is assessed. The courts have set no clear standard for identifying the baseline population for measuring adverse impact on minorities, and the lower courts vary in their approach. The baseline is yet another source of uncertainty, rather than of safe haven. In addition, the very threshold requirements that supposedly level the playing field are themselves subject to challenge on disparate impact grounds (as they often disproportionately screen out minorities). Finally, the data establishes that persons with similar years of education, training, or experience are not necessarily equivalent in their degree of learning and skill. As noted above, blacks and whites show average differences in academic achievement and proficiency at every level of education. Given this picture, applicants or job-eligible persons from different groups who are seemingly "qualified" for a job – in that they satisfy some threshold requirements or

Perhaps the most important virtue of disparate impact realism is that it functions as an information forcing device by enhancing employers' incentives to investigate possible ways to reduce the validity-diversity tradeoff. As discussed more fully below, the disparate impact rule is popular among those who insist that existing job selection practices exclude too many minorities who could actually succeed on the job. The contention is that refinements in personnel selection technology could reduce the diversity-validity tradeoff. Although, as already discussed, the search for such methods has so far proven disappointing, some researchers continue to believe that more intensive efforts are needed and might bear fruit.[85] To the extent that better methods can be developed, however, they are best tried and evaluated, at least in the first instance, by the very employers who stand to benefit from them. Thus, rewarding businesses for searching out and employing new methods that might better reconcile diversity and validity is certainly desirable. But employers currently have little incentive to try to devise or identify effective predictors that reduce or minimize disparate impact. Under the stringent and highly unrealistic standards for DI liability in force today, the great majority of valid personnel practice can be expected to generate enough racial disparity to trigger liability under a broad range of real-world conditions. Thus employers will rarely avoid triggering disparate impact liability despite best efforts to alleviate the validity-diversity tradeoff. Because efforts to find screens that substantially predict productivity while reducing disparate impact are mostly doomed to failure, employers have no reason to search for and devise ways to meet the disparate impact rule's threshold requirements.

In contrast, under the somewhat lower expectations of disparate impact realism, employers

---

meet basic criteria– will not necessarily perform equally well. Thus, adjusting the baseline for such criteria will not necessarily erase group differences in job success. In fact, as discussed, this is the situation that currently prevails.

[85] See, e.g., Schmitt, Neal, Catherine Clause and Elaine Pulakos, Subgroup Differences Associated with Different Measures of Some Common Job-Relevant Constructs, 11 International Rev. Of Ind. And Org. Psychology (1996) 115-139, 116-117 (suggesting that more research is needed on measurement methods and approaches to "minimiz[e] subgroup differences").

may find it easier to identify methods that, although still substantially predictive of productivity, generate a workforce that is diverse enough to meet target ratios. Since employers can thereby avoid being sued or summarily escape liability, they will have an incentive to search out those methods. In sum, a more relaxed DI standard operates as an information forcing device by rewarding employers for devising practices that achieve greater diversity while still maintaining efficiency. For this reason, realism is preferable to the standard in force today.

Although realism has many virtues, it is also subject to challenges. The most serious issues relate to its reliance on job performance ratings and on reported group disparities in worker productivity as the linchpin of the new standard. This feature generates a series of potential questions. Is the status quo of existing racial job performance gaps an appropriate benchmark for a revised disparate impact rule, and are the existing data on these gaps trustworthy? Should the expected ratios of minority hires be geared to differences currently observed among job incumbents in the ability to do the job, or will they be tied to updated data? If not subject to frequent updates, doesn't disparate impact realism run the risk of freezing current racial inequalities in place to the detriment of protected minorities, with no leeway for hoped-for narrowing of gaps? If real-time data is the rule, how would disparate impact realism play out in the face of a potentially dynamic situation? All of these are important questions, and addressing them requires subjecting the evidence concerning race gaps in job performance to a more searching analysis.

The proper application of disparate impact realism requires trustworthy measures of job performance and accurate estimates of group differences. A large market now exists for research by IOP experts who support plaintiffs and businesses in discrimination litigation and advise employers on how to design personnel policies to best avoid liability.[86] If disparate impact realism became the standard, experts could put less effort into the complicated business of

---

[86] See, e.g., Baker, M. R., Hughes, H. D., Mitchell, P.G.,& Tetlock, P. E. (in press). Proactive Responses to second-generation risks in labor and employment cases. Employment Relations Law Journal

validating job screens (that is, establishing their ability to predict job success). Instead, they could focus their attention simply on measuring job performance (which is already part of the validation process) and documenting group differences for different jobs. In fact, although data on the validity of employment selection methods is plentiful, there is somewhat less evidence on the adverse impact of personnel screens and on group differences in actual on-the job performance.[87] The studies to date are not as refined and up-to-date as they could be.[88] It would clearly be desirable to have better and more extensive data on patterns of job performance for specific types of work. The hope is that IOP experts would generate more and better data on job performance, including group differences, in response to the proposed disparate impact reform. Although the need for high quality, reliable evidence imposes an initial burden on employers, the hope is that this would abate as large standard databases for particular types of jobs are developed and become available. A sound empirical foundation for claims of group differences in job performance would hopefully deter many lawsuits and should be deemed acceptable at the initial stages of litigation, as on motions to dismiss and for summary judgment.[89]

---

[87] See Schmitt, Neal, Catherine Clause and Elaine Pulakos, Subgroup Differences Associated with Different Measures of Some Common Job-Relevant Constructs, 11 International Rev. Of Ind. And Org. Psychology (1996) 115-139, 120 (noting the relative paucity of adverse impact data relative to evidence on validity, and speculating that this is due to "concern regarding legal implications of presenting data on subgroup differences").

[88] See, e.g., McKay and McDaniel, 539 (noting the problem of combining older data with newer empirical studies in estimating job performance differences). See also Schmitt, Clause and Pulakos, supra (noting methodological problems with data on group disparities in performance on job screens and in the workplace). Methodological problems are commonplace in this area, with published studies reporting a range of values for group differences in performance that depend on methods of job assessment, the size and composition of the workforce samples at issue, the actual jobs examined, the selection screens employed, and statistical techniques and corrections applied. There is clearly room for improvement.

[89] Before general ability testing went out of fashion in the wake of disparate impact challenges, some courts would accept the large body of evidence documenting racial gaps on general ability tests as well the correlation between general ability and job performance across the board, in lieu of individualized validation studies assessing the employer defendants' specific procedures. See Kelman, General Ability Testing, 1215, note 160 ("A small but not insignificant

One oft-heard claim is that assessments of job performance are inherently unreliable because vulnerable to unconscious biases and subjective distortions.[90] A large literature has been devoted to assessing the accuracy, reliability, and reproducibility of job performance assessments with particular attention to whether subjective ratings are vulnerable to bias. Although some research claims to detect race effects in worker ratings, the expert consensus is that subjective supervisor ratings are reliable and a fairly accurate reflection of work performance overall.[91] In particular, studies reveal a high concordance between subjective appraisals and objective, quantifiable measures, such as absenteeism, tardiness, work errors,

---

number of courts already accept the validity generalization hypothesis [with respect to general ability tests]: defendants in such courts need not perform a local validation study"). That practice should be revived under disparate impact realism.

[90] Subjective hiring and promotion processes have been challenged as vulnerable to race and gender bias on theories of both disparate impact and disparate treatment. See, e.g., Amy L. Wax, The Discriminating Mind: Define It, Prove It, 40 Conn. Law Rev. 979 (Winter 2008)(noting the contention that subjective job evaluations are vulnerable to distortion by inadvertent or unconscious anti-minority bias). See also John Monahan, Laurens Walker, and Greg Mitchell, Contextual Evidence of Gender Discrimination: The Ascendance of 'Social Frameworks', 94 Virginia Law Rev. – (2008) (describing challenges to personnel policies at Home Depot and Walmart).

[91] See, e.g., James L. Outtz and Daniel A. Newman. A Theory of Adverse Impact. pp. 53-94, 76-78, in Outtz, Adverse Impact (concluding that racial bias in subjective job assessments is marginal and has little effect on rating accuracy). Even when the data suggests that subjective assessments are influenced by the racial identity of supervisors and workers, it has not been established whether this results from discrimination against opposite race or in favor of same-race workers, or whether black or white supervisors (or both) are biased. That is because there is no independent standard against which to measure the accuracy of subjective ratings. See e.g., Ford, J. Kevin, Kurt Kraiger, and Susan L. Schechtman. (1986). Study of Race Effects in Objective Indices and Subjective Evaluations of Performance: A Meta-Analysis of Performance Criteria. *Psychological Bulletin*, *99*(3), 330-337, 334 (summarizing and discussing data on rater and ratee race effects); Kurt Kraiger and J. Kevin Ford, A Meta-Analysis of Ratee Race Effects in Performance Ratings, 70 J. Of Applied Psychology (1985) 56-65, 58 (noting that lack of a benchmark for actual performance means that "a meta-analysis of race effects cannot separate the relative contributions of ratee performance and rater bias to ratings differences").

tasks completed, and unit output.[92] When detected, race effects are minimal. The modest size and lack of consistency in observed race effects, and the substantial correlations between objective and subjective measures of performance for all racial groups, have produced a cautious consensus among experts: most of the observed disparities in worker ratings by race are due to real differences in performance and cannot be attributed to supervisor bias.[93]

In any event, the argument that disparate impact realism is unworkable because job performance ratings are unreliable and tainted by bias applies with equal force to the disparate impact rule as currently applied. Central to its operation is the employer's assertion of the business necessity or job-related character of a selection practice. But that defense rests on

---

[92] See, e.g., Mark Kelman, General Ability Testing, Harv. L. Rev, supra, at 1211 (noting concerns about bias in subjective workers ratings, but discussing consensus among IOP experts that subjective and objective measures tend to align). As with other measures in the IOP field, the main challenges are methodological. Job assessment protocols are diverse and the elements of performance that employers value vary across contexts. IOP experts have been working to improve the quality and reliability of job ratings, finding that subjective assessments tend to be more reliable when procedural checks and safeguards against inconsistencies and arbitrariness are introduced. See, e.g., Phillip Tetlock, Proactive Responses to Second-Generation Risks in Labor and Employment Cases, unpublished ms on filed with author (2010), at 13 (noting that supervisor ratings can be made more reliable and consistent when managers are held accountable and forced to justify or explain judgments, when appraisal methods are highly structured and include objective instruments or assessments, and when multiple and diverse raters are employed.). For a review of job performance assessment methods, see, e.g., Patrick F. McKay. Perspectives on Adverse Impact in Work Performance: What We Know and What We Could Learn More About. pp. 249-270, in Outtz, Adverse Impact, at 251-252.

[93] See Outtz and Newman, A Theory of Adverse Impact, in Outtz, *Adverse Impact*, at 77 ("Altogether it would appear that black-white differences in job performance ratings are attributable, on average, to actual differences in job performance rather than to rater bias"). See also Ford, J. Kevin, Kurt Kraiger, and Susan L. Schechtman. (1986). Study of Race Effects in Objective Indices and Subjective Evaluations of Performance: A Meta-Analysis of Performance Criteria. *Psychological Bulletin, 99*(3), 330-337, 335 (noting that "consistent effects across . . . criterion types [for performance assessment] suggest that there are race differences in job performance in the organizations sampled"); Phillip Tetlock, Proactive Responses to Second-Generation Risks in Labor and Employment Cases, unpublished ms on filed with author (2010), at 23 ("Experts who claim that subjective assessments are inherently biased base their claim on lab studies from social psychology that do not replicate in the field, and those experts ignore the fields studies by [IOP] psychologists to the contrary").

demonstrating a correlation between a job selection method and actual on-the-job performance, which requires rating job performance. If such measures are distorted, whether through bias or otherwise, then the entire structure of disparate impact liability is fatally flawed. In short, the concept of validation, which is a key feature of the disparate impact rule, assumes that the criterion for validity – the appraisal of job success – is accurate and sound. If it is not, rejecting disparate impact realism does not eliminate the problem.

Even if employee ratings are fairly reliable, the question remains as to why existing racial job performance gaps are an appropriate benchmark for a revised disparate impact rule. Why does it make sense to gear the expected ratios of minority hires to observed group differences in job success? As already noted, realism indulges the assumption that group performance differences on the job reflect differences in ability to do the job in the pool of eligible workers. Does that assumption make any sense?

Addressing this question requires understanding racial differences in job performance in relation to the spectrum of abilities among job candidates and the winnowing function of job selection measures. In light of employers' interests in selecting candidates with the best predicted future performance, employers will seek to use selection devices that match applicants to the job. As already noted, job screening protocols will ideally choose people with similar ability to do a job and thus minimize or eliminate group differences present in the background population. Although such devices are imperfect, they will tend to screen out people who perform less well, regardless of their racial group. Because job incumbents, regardless of group identity, are deliberately chosen to possess similar skills, they are characterized by so-called range restriction – they are generally *not* representative of the background population as a whole or even of the pool of candidates for that job.[94] Range-restriction tends to be greatest for jobs

---

[94] See, e.g., Hunter, John E., Frank L. Schmidt, and Michael K. Judiesch. (1990). Individual Differences in Output Variability as a Function of Job Complexity. *Journal of Applied Psychology*, *75*(1), 28-42, 33 ("the standard deviation for incumbent workers is subject to restriction in range caused by selective hiring, selective promotion of better workers, and selective termination of poorer workers").

that turn away the most applicants and are the most competitive. The most competitive jobs often, although not always, require higher levels of skill, but even less skilled workers are not hired at random. Although the range varies widely, some degree of selectivity (and thus range restriction), however informal, operates at all levels of the job market.

Range restriction helps explain why racial disparities in job performance tend to be smaller than those measured for many standard job selection criteria. For example, black-white differences in mean scores on general mental ability tests are typically two to three times larger than differences commonly observed in job performance, and race gaps on standard job selection criteria, which tend to be smaller than for ability tests, can also exceed black-white performance gaps seen on the job, although by less.[95] Second, as noted, IOP experts have observed that performance ratings for jobs at every level depend in part on personal attributes that are not well-captured by conventional selection methods.[96] These attributes are de-emphasized in present job

---

Range restriction is an important methodological problem in the IOP field, affecting estimates of all parameters relevant to personnel management and their correlations. The IOP literature on potential inaccuracies in the data due to range-restriction, see e.g., Sackett, Paul R., Matthew J. Borneman, and Brian S. Connelly. (2008). High-Stakes Testing in Higher Education and Employment. *American Psychologist*, *63*(4), 215-227 ("Failure to take range restriction into account can dramatically distort research findings [on personnel selection].").

[95] As noted, the average difference in cognitive ability for blacks and whites hovers around 1 SD, whereas most measured differentials in job performance range between about .24 and .39 SD. See, e.g., James Outtz and Daniel Newman, A Theory of Adverse Impact, in Outtz, *Adverse Impact*, at 85 ("[S]ubgroup differences on cognitive ability tests are far larger than subgroup differences on actual job performance"). See also Kevin R. Murphy. How a Broader Definition of the Criterion Domain Changes Our Thinking About Adverse Impact. pp. 137-160, 138 in Outtz, *Adverse Impact* ("The adverse impact of cognitive tests is particularly egregious because test score differences are known to be substantially larger than differences in job performance, academic achievement, and other criteria typically used to evaluate the success of selection decisions"). For comparisons of job performance ratings with scores on selection tests, see, e.g., Patrick McKay, Perspectives on Adverse Impact in Work Performance, etc, pp. 249-270, in Outtz, Adverse Impact.

[96] See discussion of "task performance" and "contextual performance,"or job citizenship, supra.

selection relative to cognitive ability because they do not correlate as strongly with performance and cannot be measured as precisely or accurately. To the extent that commonly used personnel screens fail to fully capture the full range of worker behavior, scores on selection criteria will not perfectly predict on-the-job success.

These observations help explain why racial job performance gaps are smaller than on standard selection criteria. The also help account for why racial disparities in on-the-job ratings exist and persist despite some degree of screening. As noted, a perfectly functioning meritocratic system with omnisciently predictive job selection should produce no racial gaps in job performance. Regardless of the distribution of skills in the background populations, job incumbents should be equally capable of doing the job, and racial gaps among job-holders should disappear.

This is not the pattern observed. Although job holders tend to be more similar than job candidates (or the general population) in ability to perform a job, racial disparities in job success persist. Several factors probably contribute to this. First, as described, existing personnel devices are inherently imperfect, and thus fail fully to control for all background differences or attributes that bear on performance. Skill disparities in the background population will thus tend to carry over into the workplace.[97] Second, employers operating in the shadow of current anti-discrimination law may shy away from screens that produce too much adverse impact. Unfortunately, these tend to be the devices that best predict job success. This trend reduces employers' ability to match candidates to jobs, and tends to preserve background disparities in the population hired. A commonly relied-on job credential, for example, is years of education or

---

[97] Thus, the target of performance parity is an oversimplification and will likely never be achieved. Employers typically hire people who are best qualified first but then proceed to fill slots from the top down, or alternatively hire candidates who possess minimal qualifications but who are not all equally able. Either method can produce a range of measured productivity among those hired, because not everyone will perform exactly the same. This can in itself generate some racial differences (because of the distribution of skills even among those hired). Obviously the reality deviates somewhat from the ideal.

specialized training ("biodata"). However, hiring people with similar years of schooling can preserve group disparities because years of education are an imperfect proxy for actual skill.[98] Thus, matching workers for years of schooling will not necessarily eliminate racial gaps in ability to do the job.[99] In sum, although workers in particular jobs should be equally able regardless of race, they are not. Group differences persist in measured job success.

How do these observations bear on disparate impact realism? Realism ties hiring ratios to actual group differences in productivity currently measured in the real world. The strength of this standard is that it is geared to actual patterns that workplaces have achieved through the various personnel methods and screening devices in common use. The data suggest that minority job-holders are, if anything, lagging behind the white majority in actual performance on the job. This means that existing job selection practices are rather less stringent in screening out minorities than whites, given actual performance measures. The reasons given above for this "overshoot" are speculative, and the relative contribution of the possible factors cannot be known for sure. The inherent technical limitations in current personnel practices are probably

---

[98] See Robert E. Ployhart, The Diversity-Validity Dilemma: Strategies for Reducing Racioethnic and Sex Subgroup Differences and Adverse Impact in Selection, *Personnel Psychology* 61 (2008) 153-172 , 167(noting that years of education are not a good proxy for actual ability when used without gauges of achievement); See, e.g., Roth, Philip L. and Philip Bobko. (2000). College Grade Point Average as a Personnel Selection Device: Ethnic Group Differences and Potential Adverse Impact. *Journal of Applied Psychology*, *85*(3), 399-406 (noting .78 standard deviation difference between blacks and whites on college grade point average). See also Thomas Espenshade and Alexandria Radford, No Longer Separate, Not Yet Equal: Race and Class in Elite College Admissions and Campus Life.

[99] The effects of job screening on worker performance and range restriction would appear to predict that measured job performance differences by race should be relatively greater for unselective jobs than for highly competitive positions. If almost all candidates for a particular position are hired, then persons hired from each group will be more representative of their background population. Since, as noted, g-loaded criteria are the most predictive of job success, and the black-white difference on such measures approaches 1 standard deviation, the gap in on-the-job performance should be greater for workers who are not stringently screened for ability. Of course, whether this will be observed depends critically on the profile of the job candidate pool. The literature does not appear to be focused on analyzing performance differences in jobs based on selectivity.

crucial, and the fact that employers must hire from a range of abilities to fill jobs will also play a role.[100] However, some mild degree of affirmative action in staffing – whether in the service of diversity goals or in the shadow of the law – cannot be ruled out.

This discussion reveals that the answer to question of whether measured patterns of on-the-job performance match background group profiles in ability to perform is almost certainly no. It also reveals why it doesn't really matter. Realism is a quick and dirty method for moving the situation closer to the disparate impact ideal or, barring that, at least narrowing the scope of potential liability. As noted, affirmative action is controversial and is arguably at odds with the stated purpose of the disparate impact doctrine. By allowing employers to be more selective than under the 4/5 rule, realism is designed to advance the objective of equal standards, and equal job performance, by race . As explained, although employer's target ratios are predicted on the basis of actual job performance gaps, the realism rule is not designed to preserve them. Rather, the hiring ratios expected (see Figures 1A and 1B) are calculated to allow employers to winnow out minority job candidates to the degree necessary to generate a cadre of similarly capable individuals from all groups (just as, *ceteris paribus*, the hiring ratios expected given scores on job screens are designed for the same purpose). Thus, the smaller minority ratios permitted by disparate impact realism would ideally bring the measured job performance ratings of blacks and whites closer together.

The very fact that disparate impact realism is supposed to narrow racial differences in job success generates a serious problem for the implementation of the rule. If realism does shrink or eliminate performance gaps, that will potentially alter the magnitude of the very benchmark (racial differences in on the job performance) that determines whether an employer violates the rule in the first place. More importantly, it will do so in a perverse direction: as the black-white productivity gap is seen to become smaller, the rule will require employers to increase the ratio of blacks to whites hired. But that makes no sense under the terms of the rule

---

[100] See note [86] supra.

itself.

The problem is that, once disparate impact realism is implemented, an observed racial convergence in measured job performance could have two possible causes: it could result from employers becoming more selective, as the realism rule itself allows, or it might be the product of underlying improvements in minority skill levels. Employers should not be required to change their standards in response to a convergence that results from the operation of disparate impact reform. After all, a relaxation of the disparate impact rule does not in itself change the distribution of human capital among workers. On the other hand, if blacks upgrade their skills relative to whites over time, it is fair to expect employers to hire more of them. These possibilities create the dilemma of distinguishing changes that result from the operation of disparate impact realism from those that are independent of the rule. The latter would justify increasing pressure on employers to achieve more racial balance, but the former would not.

This is likely a non-problem, however. That a narrowing of group differences in actual job performance is one of realism's potential effects does not mean it will actually occur. As noted, there is good reason to believe that many employers are currently routinely violating the 4/5 rule despite the danger of incurring liability. Indeed, given the size of racial skill gaps and the disparate impact of many routinely employed job filters, many workplaces can be expected to fall short of the diversity required by disparate impact realism. And informal observation suggests that racial imbalances are commonplace across a spectrum of jobs. If workplace diversity does not currently differ significantly from what disparate impact realism would allow, reforming the law would not have much effect on minority representation or patterns of on-the-job performance. That is no argument against it, however. The hope is that disparate impact realism will reduce employers' potential legal exposure, which, as already argued, is desirable in itself.[101]

---

[101] It must be admitted, however, that this effect is far from a foregone conclusion. Many common job screens – such as educational credentials and structured interviews – show black white differences greater than .3 SD, which place them outside the realism box, and thus beyond

Beyond that, the situation may depend on how realism is implemented. One key question is whether employers will be required to present real-time data on productivity. If recent data is required, then any narrowing of incumbent performance from disparate impact realism would increase the expected hiring ratios under the rule, exemplifying the perverse spiral already described. If employers can present baseline data from the time of the rule's implementation, however, then this problem will go away. The main drawback of the latter rule is that it runs the risk of making realism insensitive to any actual upgrades in minority human capital that might occur. This could freeze in place an unequal status quo to the detriment of protected groups.

Although it is obviously desirable to monitor job performance patterns over time, it should not follow that employers need to offer updated data. Rather, evidence on racial discrepancies prior to and around the time of implementation, suitably reinforced by additional evidence from this time frame, should suffice. Race gaps in skill have been fairly stable in recent decades. Although those differences might significantly narrow in the future with improvements in education, experience, or training, this has yet to occur. Such a development, or other evidence pointing to significant upgrades in the distribution of human capital, would force a re-evaluation of the rule and the benchmark ratios used in applying it. Short of that, however, numbers geared to implementation should serve as a rough and ready guide to expected staffing patterns for the foreseeable future.

That disparate impact realism may end up preserving the status quo by permitting employers to do pretty much what they are now doing is not a reason to oppose reform. It is important to emphasize that employers currently are *not* screening out minorities who could do the job as well or better than whites. They are not discriminating by selectively or

---

the reach of its modest safe harbor. If employers are already using these screens despite their adverse impact and the danger of legal challenge, realism will not only fail to alter actual hiring ratios but, but will not appreciably expand the safe-harbor for liability either. (See figure 1B). Whether this proposal will actually expand the safe harbor available to employers is an empirical question.

disproportionately excluding *able* minorities. Rather, the evidence suggests just the opposite.[102] To the extent they continue with current practices under a relaxed rule, minorities will not suffer harm.

In short, allowing employers to target lower ratios may advance the goals of the disparate impact doctrine – which is to achieve a competitive meritocracy. And even if realism does not close racial performance gaps, employers will still be better off. The data shows that employers are tolerating existing group disparities on the job. They are in fact living with them. And even if employment patterns don't change much, employers may still benefit from a somewhat narrower scope of potential liability under the realism rule.

At the end of the day, disparate impact realism is rooted in practical compromise. It responds to the concerns that motivated the Supreme Court's articulation of the doctrine in the first place – which is that employers might exclude minorities for reasons unrelated to job performance– while recognizing that prevailing practices do not in fact pose this danger. It

---

[102] A practice with disparate impact that could generate the observed pattern of somewhat lower average productivity for blacks relative to whites would be one that disproportionately (and selectively) screened out more *capable* blacks in favor of the less capable. This explanation is farfetched in light of what is known about commonly used job selection devices, which is that they are generally "unbiased" – that is, they predict performance equally well for all groups, with more capable candidates tending to score higher than persons who perform worse on the job. See supra. Thus, although these devices are not perfectly predictive, there is no indication that they differentially screen out more capable individuals from some groups relative to others, and thus no evidence that employers are differentially rejecting more capable blacks.

Moreover, even if a manager was attempting to produce a "whiter" workforce, it would make more sense to reject black applicants generally, rather than differentially to exclude the most able ones. It is possible to imagine that a particularly racist employer might want to keep blacks from attaining success, but rejecting the most capable job candidates is a particularly self-defeating way to accomplish this. The more likely strategy is just to exclude blacks from a particular job category altogether.

It is worth noting, finally, that the fact that blacks and whites in a given position are observed to have roughly *equal* productivity does not rule out that an employer is using selection devices with disparate impact. The employer could still be hiring too few blacks relative to qualified persons available. However, that is not a worry under current conditions, because the pool of qualified persons for many jobs now differs significantly by race.

honors the competitive meritocracy while recognizing the underlying reality – which is that, under present conditions, a properly functioning system will generate significant racial disparities in many job categories. And it holds employers to diversity requirements that are, if anything, overly generous to minority job-seekers in light of the background distribution of skill between groups.[103]

B. Abolish It

Disparate impact realism is a modest proposal. Although it relaxes previous

---

[103] Another limitation of the disparate realism reform proposed here is that it is mainly directed at cases involving employment, and especially those in which liability is based on a racially adverse impact. The disparate impact rule has a broader reach, encompassing claims of discrimination by gender in employment or in other areas such as housing, consumer credit and mortgage lending. See, e.g., Sara Aronchick Solow, Racial Justice at Home: The Case for Opportunity-Housing Vouchers, 28 Yale Law and Policy Review 481, 488 (2010)(noting that the federal Fair Housing law "outlaws disparate impact in housing just as Title VII does in employment"); Schwemm, Robert G. And Jeffrey L. Tarne, Discretionary Pricing, Mortgage Discrimination, and the Fair Housing Act, 45 Harvard Civ. Rts and Civ. Lib. Law Rev. 375, 416-417 (Summer 2010)(describing lawsuits alleging unlawful disparate impact in mortgage lending practices); see also Ian Ayres, Testing for Discrimination and the Problem of Included Variable Bias (draft on file with author – Penn Law and Economics Seminar series, October 6, 2010) (describing lawsuits under the Equal Credit Opportunity Act against car dealers, car loan underwriters, and mortgage lenders for practices that have an adverse impact on minorities). See also id. at 26, note 47 (quoting commentaries on ECOA regulation stating that "The act and regulation may prohibit a creditor practice that is discriminatory in effect because it has a disproportionately negative impact on a prohibited basis, even though the creditor has no intent to discriminate and the practice appears neutral on its face, unless the creditor practice meets a legitimate business need that cannot reasonably be achieved as well by means that are less disparate in their impact." citing Official Staff Interpretations, Regulation B (Equal Credit Opportunity Act), 12 C.F.R. §202.6(a)- (2009)).

Claims involving gender sometimes target requirements for physical strength, appearance, or fitness, on which the genders do differ significantly. However, differences in cognitively related abilities between men and women tend to be small or non-existent, although there are exceptions (for example in fields like math or engineering, and especially for jobs requiring very high ability). Thus, adjustment of the 4/5 rule would generally not be indicated. See Selmi, Michael. (2006). Was the Disparate Impact Theory a Mistake? *UCLA Law Review*, *53*(3), 701-782, 746 (discussing lawsuits claiming disparate impact by gender, including class actions against Walmart and Home Depot, and actions challenging strength, fitness, hairstyle, and appearance requirements that differentially affect women). On disparate impacts generally, see Harris and West-Faulcon, supra, at 112 n. 147 (noting contexts in which the courts have "declined to intervene to ameliorate racial disparities" generated by facially neutral laws).

requirements in some circumstances, it preserves them in many. See figure 1B. Significant restrictions still apply, and presumptive liability will be triggered in many cases. In addition, as described above, realism potentially (and ideally) sets in motion a convergence in patterns of performance that contains the seeds of its own destruction. This argues for viewing realism as a temporary adjustment – a stopgap substitute for the present rule and a way station towards a new equilibrium.

That disparate impact realism is a modest proposal may serve as a mark in its favor, but also opens it to the criticism that it does not go nearly far enough. A more radical approach is to abolish the disparate impact doctrine altogether. This would be equivalent to overruling *Griggs v. Duke Power* and amending Title VII to remove racially disparate impact as a basis for employment discrimination liability.

The principal argument for repealing disparate impact is that, under present social conditions, racial imbalances in employment are exceedingly weak evidence of discrimination, either in the form of race-based disparate treatment or through unlawful disparate impact. The IOP data indicate that differences in the distribution of skill and human capital, not race-based exclusion or arbitrary barriers to employment, are the principal factors behind racial imbalances on the job. In light of these realities, the disparate impact rule is fatally overbroad and ensnares far too much conduct in its net. Under current social conditions, the vast majority of commonly employed selection procedures are valid and job related, and thus do not actually violate the disparate impact rule. Yet most personnel practices will routinely show enough adverse impact to create a prima facie case of discrimination, thereby shifting the burden of justification to employers. Given the legal uncertainties and practical difficulties surrounding disparate impact claims, employers run a significant risk of being found liable regardless of whether their defense is valid, and even though they are not actually violating the rule. To be sure, relatively few disparate impact cases are filed relative to the cases that could potentially give rise to challenges,

and employers almost always prevail.[104] But because virtually no aspect of the business necessity defense is settled law, employers still face the prospect of protracted, expensive, uncertain, and resource intensive litigation to defend their practices, which encourages them to engage in perverse, inefficient, and evasive tactics. In sum, the overbreadth of the criterion for presumed disparate impact liability is not only inefficient, but is also fundamentally unfair.

There are other good arguments in favor of repealing the disparate impact doctrine. First, most claims brought under Title VII allege unlawful disparate treatment.[105] Thus, abolishing disparate impact would therefore have little overall effect on the vindication of worker rights under Title VII. To be sure, disparate impact remains an important avenue for challenging hiring and promotions into sought-after government positions, most notably as police and firefighters.[106] But the evidence suggests that the costs imposed by these cases are not worth the alleged benefits. These lawsuits impose a considerable burden on local governments, consuming enormous time, attention, and resources that could be devoted to other purposes. Although these challenges may marginally increase diversity in select instances, the objective could be accomplished more simply by others means (including dropping civil service exam

---

[104] See, e.g., Selmi, Selmi, Michael. (2006). Was the Disparate Impact Theory a Mistake? *UCLA Law Review*, *53*(3), 701-782; see also, e.g., Wendy Parker (2006). Lessons in Losing: Race Discrimination in Employment. *Notre Dame Law Review*, *81*(3), 889-954, 899 (noting that disparate impact employment actions are relatively uncommon, comprising about 15% of claims represented in published cases, and that impact claims are almost always coupled with disparate treatment allegations). See also David Sherwyn and Michael Heise, The *Gross* Beast of Burden of Proof: Experimental Evidence on How the Burden of Proof Influences Employment Discrimination Case Outcomes, 42 Arizona St. L. J. 901, 906 (Fall 2010)(noting that "the vast majority of discrimination cases are disparate treatment or intentional discrimination cases").

[105] See Selmi; see Parker, ibid.

[106] See, e.g., *Ricci v. de Stefano*; Heather MacDonald, Fighting Fire with Quotas, Manhattan Institute website, http://www.city-journal.org/2010/eon1024hm.html (describing litigation initiated by the Justice Department challenging the civil service exam for New York City firefighters and the district court ruling finding that the city's test had an unlawful disparate impact on black job candidates.). See also Helen Norton, supra, William and Mary Law Rev. At 254, notes 233 & 234 (detailing cases presenting challenges to civil service exams).

requirements, or selecting randomly from eligible pools). In any event, diversity for its own sake is not what disparate impact commands. The evidence in fact suggests that qualified black candidates are *not* being arbitrarily screened out or disproportionately denied jobs as police or firefighters. Although evidence on performance patterns of police and firefighters by race is scarce, the data indicate that civil service exams are good predictors of success in these types of jobs.[107] The familiar validity-diversity trade-off applies as much to firefighter or police positions, and civil service jobs generally, as to other positions available in the economy.[108] Given racial gaps in developed skills and abilities, racial imbalances in these jobs will likely persist within any kind of stringently meritocratic system.

Yet another compelling argument for abolishing disparate impact liability is that, although the potential for liability is widespread, enforcement is selective, arbitrary, and erratic. Adverse impact is everywhere, and the world is full of disparate impact lawsuits waiting to happen. Racial imbalance is pervasive in business, the professions, technological fields, academia, and finance.[109] This pattern is not confined to elite and lucrative positions. The U.S.

---

[107] See, e.g., Gerald V. Barrett, Michael D. Polomsky, and Michael A. McDaniel, Selection Tests for Firefighters: A Comprehensive Review and Meta-Analysis, 13 J. of Business and Psychology, 507-513 (finding that commonly used civil service tests predict training expertise and supervisor ratings of firefighters). See generally Michael G. Aamodt, Research in Law Enforcement Selection (2004), at 34 (noting that cognitive ability, and to a lesser extent civil service exams, predict evaluations of police performance).

[108] Of course, disparate impact plaintiffs challenge or simply ignore this evidence. See Heather MacDonald http://www.city-journal.org/2010/eon1024hm.html (reporting on U.S. federal district court Judge Garaufis's grant of summary judgment for plaintiffs and his dismissal of evidence of racial differences in human capital as an explanation for the adverse impact of a firefighter's qualifying exam given by New York City.). Plaintiffs in these cases frequently contend that alternative selection methods are available that can preserve or boost productivity with less disparate impact, and courts sometimes buy this argument. As already discussed, there is no reason to believe such options are currently available – for civil service positions or any other job – and good reasons to believe they are not.

[109] Under *Wards Cove Packing v. San Antonio*, 490 U.S. 642 (1989), and section 703(k)(1) of the 1991 Civil Rights Act, plaintiffs must ordinarily specify the practice or procedure that is the source of the observed disparate impact. However, the credentialing and

government has long used tests of cognitive ability, or that draw heavily on such ability, to determine admission to the military and assignments within it. The military's entrance exams have a pronounced disparate impact by race.[110] Yet such practices persist without serious challenge. It may be argued that plaintiffs don't bother to sue because they believe judges will defer to the military or to professional standards. Although that may be true, it is arguably unfair to employers who end up being sued for similar practices and must bear the burden of litigation regardless of whether they ultimately prevail.

A more critical question is whether, in the absence of disparate impact liability, employers would adopt or revert to selection methods that arbitrarily exclude minorities – the very fear that underwrites the doctrine in the first place. The argument that firms operating in a competitive environment have no interest in screening out good workers from any group is unlikely to convince proponents of strong anti-discrimination laws, who are generally suspicious of market forces. In fact, the simple answer is that there is no airtight guarantee against employers adopting overly exclusionary practices, whether intentionally or not. Rather, the case

---

training requirements for every white collar and professional job – such as nurse, lawyer, law clerk, accountant, physician, pilot, engineer, computer programmer, college professor, teacher, administrator, etc. – produce disparate impacts at every stage, and stand as important hurdles to the entry of minorities into remunerative positions. See, e.g., Elizabeth Bartholet, Application of Title VII to Jobs in High Places, 95 Harvard L. Rev. 945 (1982). See also [DATA on JOB credentialing exams]. This author is unaware of any cases featuring disparate impact challenges to professional credentialing requirements.

[110] According to a recent report, while 16% of otherwise qualified whites applying for admission to the military (i.e., men and women with a high school degree and no serious criminal record) scored below the minimum required on the Armed Forces Qualifying Test (AFQT), 39% of African American applicants scored below the cutoff. Among those achieving the minimum score for admission, over 43% of white test-takers, but fewer than 18% of African American test-takers, scored high enough (in the top two categories) to qualify for special technical training and placement in elite service jobs. See, e.g., "Shut Out of the Military: Today's High School Education Doesn't Mean You're Ready for Today's Army," *Report of the Education Trust* (December 2010), at http://www.edtrust.org/dc/press-room/press-release/shut-out-of-the-military-more-than-one-in-five-recent-high-school-gradua

for repeal rests on a clear-eyed assessment of the main forces producing racial imbalance in the workplace today.

Most personnel selection devices presently in use have a pronounced disparate impact, and violate the 4/5 rule. Nonetheless, most are valid: they really help employers match candidates to jobs, and they do predict, albeit with varying accuracy, subsequent job performance. Moreover, the evidence suggests that very few employers are actually violating the disparate impact rule. Although their hiring ratios may flunk the 4/5 test, the requirements they impose are job related. And they are not disproportionately screening out capable minorities.

All this evidence strongly suggests that the under-representation of minorities in large segments of the job market is overwhelmingly the result of real human capital disparities rather than employer indifference to unjustified racial impacts. Indeed, measured patterns are far more consistent with *de facto* affirmative action than with unlawful disparate impact. And even if employers moved to significantly *less* diversity in the wake of disparate impact repeal, that could still be consistent with meritocratic ideals. Indeed, the repeal of the disparate impact rule would allow employers to adopt more g-loaded – and predictive – screens, which might generate greater racially adverse impact than current practices. Strictly speaking, this would not even be inconsistent with current doctrine: the courts have never barred employers from adopting the strongest and most valid predictors of job performance, regardless of adverse impact, and presumably such a practice would be consistent with the job-relatedness requirement. If the disparate impact doctrine were repealed, however, employers would bear the burden of proving the validity of their screens or worry that their job-relatedness defense would not be accepted.

In sum, existing differentials are more than accounted for by supply side differences in job preparation or other cognitive or non-cognitive group-based factors. Workforce imbalance is likely to persist without significant changes in the distribution of skill and human capital. Inequalities in job qualifications are not of employers' making, and they are ill-equipped to address them. Likewise, supply side disparities are no business of disparate impact law, and the

doctrine is not designed to correct them. As noted by a recent commentator, "[T]he theory of disparate impact . . . comes too late in an individual's career to compensate for a variety of inequalities earlier in life – in upbringing, education, or health care. . . .[Employers] do not have to redress the cumulative disadvantages that individuals face from discrimination elsewhere in society."[111]

Those who view a racially balanced workforce as desirable in itself may feel little compunction about forcing employers to "redress the cumulative advantages" that minorities suffer. On this view, a disparate impact rule that results in more racial balance is all to the good. The problem with the approach is that the proper application of the disparate impact rule will not lead to more racial balance in the current climate, because groups are not currently equally qualified for most jobs. Rather, racial balance will only be achieved by encouraging businesses to engage in self-protective affirmative action. But it is perverse to use disparate impact to accomplish a result that is at odds with the doctrine's stated goals, which is to enforce a race-blind meritocracy. If racial balance and greater diversity are the goals, it makes no sense to achieve them indirectly via the disparate impact doctrine, in contravention of that doctrine's avowed purpose. Rather, those priorities should be implemented forthrightly, with a clear articulation of expectations and requirements.

There is currently no general legal or Constitutional requirement that employers operate as a meritocracy. Employers need not hire and promote on the basis of workers' ability to do the job so long as they don't rely on forbidden criteria.[112] It is a paradox of disparate impact that employers are restricted to some form of meritocratic selection– that is, selection on criteria related to the job – only if the workplace shows a racial imbalance. But that requirement is fundamentally at odds with a blanket imposition of diversity goals or with the routine

---

[111] Rutherglen, supra, at 110

[112] See, e.g., David Sherwyn and Michael Heise, supra, at 910 (noting Supreme Court's statement, in *Texas Department of Community Affairs v. Burdine*, that there is "no obligation to hire the best candidate for a job. Instead, the employer simply could not discriminate").

consideration of race to achieve demographic balance.

What Title VII does not do is mandate a departure from meritocratic (or any other) criteria in the service of more racial balance. If the courts are indeed circumventing Title VII and constructing a de facto mandatory affirmative action regime through the application of disparate impact liability, this represents a judicial imposition of a substantive norm through the guise of statutory interpretation. The indirect imposition of racial balance through disparate impact enforcement amounts to an illegitimate judicial usurpation of the lawmaking function. By achieving a result at odds with stated statutory principles, this outcome effectively circumvents ordinary political channels. What cannot be done above-board and forthrightly should not be accomplished indirectly through subterfuge and by judicial fiat. If race-based selection is good policy, it should be mandated directly and defended on its own merits, or left to spontaneous private initiatives. Alternatively, as discussed more below, the solution lies in getting rid of competency or performance-based screens altogether, and adopting other approaches – such as lotteries or random selection procedures – that ensure maximal workforce diversity without perverting the law.

One frequently voiced argument in favor of retaining the disparate impact doctrine is that disparate impact liability is needed to "smoke out" forms of subtle race or sex-based discrimination that might be “cloaked in race-neutral selection processes.” In other words, disparate impact challenges provide “a way of finding the stealth disparate treatment case” in which a worker suffer discrimination “because of” a protected characteristic.[113] One problem

---

[113] See Harris and West-Faulcon, supra, at 114; see also Primus, Equal Protection and Disparate Impact: Round Three, 117 Harv. L. Rev. 493, 498-499 (2003); Ricci v. DeStefano (Justice Scalia, concurring)(noting the “smoking out” rationale). See also David Sherwyn and Michael Heise, The *Gross* Beast of Burden of Proof: Experimental Evidence on How the Burden of Proof Influences Employment Discrimination Case Outcomes, 42 Arizona St. L. J. 901, 906 (Fall 2010)(noting that disparate treatment cases “are particularly difficult to assess because fact finders must ascribe motivation to the actions of the employer,” and “most employers are now sophisticated enough to avoid creating the proverbial smoking gun that would easily establish unlawful intent”).

with this rationale, as noted, is that the amount of racial imbalance that triggers a prima facie case of disparate impact provides poor support for unlawful disparate treatment. In fact, under present social conditions, racial disparities alone are exceedingly weak evidence for forbidden conduct under either theory. Once again, that is because supply side factors can be expected to generate pronounced disparities even in the absence of unlawful discrimination.[114]

Nonetheless, the "smoking out" argument embodies the perception that the disparate impact framework is more effective tool than disparate treatment for targeting some forms of discriminatory conduct. For example, a private employer might intentionally adopt a neutral policy for the purpose of excluding blacks. Or a neutral policy -- such as subjective worker assessments-- might permit unconscious race-based biases to contaminate outcomes, either by skewing evaluations or by altering the weight given to evaluative factors.[115]

The disparate impact rule is not needed to get at these scenarios, as they are – or should be -- actionable as forbidden disparate treatment. A neutral policy adopted for the purpose of exclude minorities would clearly fall within the ambit of forbidden disparate treatment under Title VII, since the policy was adopted "because of race." And the contention that a disparate impact rule is necessary because the disparate treatment doctrine covers only "intentional" – i.e., conscious or deliberate – discrimination – is not supported by the statute's language, which

---

[114] See, e.g., Amy Wax, The Discriminating Mind: Define It, Prove, 40 Connecticut Law Review 979, — (2008).

[115] See Watson v. Fort Worth Bank, supra (subjective ratings); Carle, Susan, A Social Movement History of Title VII Disparate Impact Analysis, 63 Florida L. Rev. 251, 258 (2011)(noting that "supporters of disparate impact analysis also advance arguments based on the difficulty of proving hidden prejudice" and "the problems of subtle and subconscious bias"); Elaine W. Shoben, Disparate Impact Theory in Employment Discrimination: What's Griggs Still Good for? What not?, 42 Brandeis L.J. 597, 607-613 (2004)(asserting that a disparate impact theory might succeed in imposing liability in meritorious cases where disparate treatment allegations would fail); Shin, Patrick S., Liability for Unconscious Discrimination? A Thought Experiment in the Theory of Employment Discrimination Law, 62 Hastings L. J. 67, 75-83 (November 2010) (hypothesizing that an employer might inadvertently rate "work experience" as more important when white candidates possess comparatively more work experience than blacks).

forbids adverse treatment "because of race." This language is not restricted to discrimination based on "intentional" or conscious motives, because race can sway a person's decisions without that person's awareness. Although the court are somewhat confused on the question of whether Title VII covers unconscious as well as conscious disparate treatment, the statutory language fits best with that broad interpretation.[116] Thus, the disparate impact rule is not needed to ensure that inadvertent conduct described is actionable under Title VII. Although inadvertent disparate treatment may be difficult to prove, that problem exists regardless of the theory of discrimination that is advanced, and the burdens and complications of prosecuting a disparate impact claim detract from any advantages of using that rule. In sum, virtually all racially disparate treatment can be tackled by alleging disparate treatment. Little or no advantages would be lost by dropping the disparate impact rule.

One remaining concern is the fate of allegations of discrimination directed against state actors under the Equal Protection Clause. It is often stated that this clause only protects against discrimination of bias that is "intentional."[117] But that limitation is properly understood to differentiate between the governmental disparate treatment based on race (which is forbidden by the Constitution) and the government's adoption of neutral policies with a racially disparate impact – which are not considered actionable under the Equal Protection Clause. However, the Supreme Court has never ruled that adverse treatment by government actors that is inadvertently or unconsciously (and in that sense "unintentionally") motivated by race is permissible under the Equal Protection Clause. Adverse decisions taken "because of" race, whether inadvertent or deliberate, should be regarded as violating the Equal Protection guarantee. Likewise, neutral policies taken "because of" their racially adverse effects should also be regarded as running afoul

---

[116] Shin, Patrick S., Liability for Unconscious Discrimination? A Thought Experiment in the Theory of Employment Discrimination Law, 62 Hastings L. J. 67 (November 2010). See also Amy L. Wax, Discrimination as Accident, 47 Indiana Law J. 1129 (Fall 1999)

[117] See, e.g., Washington v. Davis, 426 U.S. 229 (1976); see also Amy L. Wax, Discrimination as Accident, 47 Indiana Law J. 1129 (Fall 1999)

of Constitutional protections. Once again, the disparate impact doctrine should not be needed to get at situation where unconscious racial biases play a causal role in real-world outcomes.

C. *The Problem of False Negatives*

Perhaps the most serious argument for retaining a strict disparate impact rule, or even going beyond it to adopt a more pro-diversity regime, on a critique of many competitive job selection methods as fundamentally unfair to poorly performing groups, and especially to blacks. Because strict enforcement of the DI doctrine will reduce or discourage practices with the most adverse impact, those alleging such unfairness would favor preserving the status quo..

Although initially directed at pure tests of cognitive ability, this accusation has been leveled at all competitive job selection methods that show significant differences by race. The thrust of the critique is that such devices generate too many false negatives among lower performing minority groups. That is, they exclude too many minorities who could adequately perform the job in question.

The problem of false negatives can be traced to limitations inherent in all personnel screens, which are imperfect predictors of eventual job success. Even tests of cognitive ability, which are the most powerful known pre-screening devices, have a correlation of around .5 with measured job performance. If combined with the common practice of hiring from the top down or above a given cut-off score, these tests will generate a significant amount of error in the form of false negative and false positive results.

Moreover, it can be demonstrated numerically that, for any valid selection device on which one group performs better than another, the absolute number of false negatives (that is, people who can perform the job but are not hired) relative to applicants in the higher performing group is smaller than in the lower performing group, which, in the case of cognitively loaded tests, is blacks. That is because a relatively larger number of persons from the lower performing

group will fall below the job selection cutoff.[118] Likewise, the absolute number of false positives – those who are hired but end up not succeeding on the job – will be greater for the higher performing group (whites).[119] These effects are not specific to the racial context. Rather, they are observed for any screening device with imperfect validity (that is predictive power) on which two identifiable groups differ in average performance and/or the distribution of performance.[120]

On the question of excessive numbers of false negatives, the IOP community has engaged in a complex debate that cannot be exhaustively reviewed here. In general, psychometricians have taken a number of tacks. First, some have challenged the notion that fairness equates with racial equality in the number of false negatives, regardless of average

---

[118] For a test on which blacks score lower than whites, more blacks will fail. The false negative rate at any given score is applied against many more negatives (e.g., failures), generating a higher number of persons who could do the job but aren't hired. See, e.g., Sackett, Paul R. and Steffanie L. Wilk. (1994). Within-Group Norming and Other Forms of Score Adjustments in Preemployment Testing. *American Psychologist*, *49*(11), 929-954 at 935; see also Gottfredson, Linda S. 1994. The Science and Politics of Race-Norming. *American Psychologist*, *49*(11), 955-963, 956 (noting the critique that "minority workers have less chance of being selected at the same level of job performance and thus are burdened with higher level false-rejection [false negative] rates"). See also Mark Kelman, Concepts of Discrimination in 'General Ability' Job Testing, 104 Harvard L. Rev. at 1223-1227 (noting that a larger number of blacks will be in the false negative category for a test on which blacks as a group perform worse, and demonstrating this by numerical example); id. at 1230 (Noting that "in any ex ante probabilistic screening system, the existence of false positives and negatives ensures that factual equals will not be treated equally.").

[119] See Kelman at 1226 (noting that "whites who will in fact turn out to be poor workers get hired considerably more frequently than do blacks who will turn out to be poor workers.") In sum, a comparison of group profiles reveals that more people from the lower performing group will fail to "pass the test" relative to the higher performing group – including more people who could have done the job but nonetheless don't make the cutoff score. Likewise, relatively more people from the higher performing group will pass the test, including more people who ultimately cannot do the job (false positives).

[120] See Sackett, Paul R. and Steffanie L. Wilk. (1994). Within-Group Norming and Other Forms of Score Adjustments in Preemployment Testing. *American Psychologist*, *49*(11), 929-954, 933 (noting that group differentials in false negatives and positives "are inevitable when a test that predicts performance with less than perfect accuracy and on which group differences exist is used in a top-down fashion.")

group performance. Rather, many experts embrace the notion of a fairly meritocractic test as one that is both valid and unbiased – in being equally predictive of productivity for persons from all groups.[121] In fact, there is extensive evidence that commonly used employment screens are not biased against minorities on this metric. Second, critics have noted that proposals for equalizing false negatives come at considerable cost or have other undesirable consequences, including generating an excessive number of false positives (or persons who are hired, but fail at the job) from lower performing groups.[122]

One proposal advanced by the IOP community for dealing with false negatives as well as performance disparities generally is subgroup (race) norming.[123] By applying different standards or cut-offs for candidates from different groups, race-norming can reduce or eliminate adverse

---

[121] This is the widely used "Cleary" model of test fairness, which looks at whether the test has the same predictive validity for different social and racial groups, regardless of their average level of performance. Most job selection criteria have been demonstrated to be fair on this model. See, e.g., Sackett, Paul R., Matthew J. Borneman, and Brian S. Connelly. (2008). High-Stakes Testing in Higher Education and Employment. *American Psychologist, 63*(4), 215-227, 223 (noting that, for most commonly used employment screening devices, "the regression lines relating test scores to criterion performance" are similar for blacks and whites, and may even over-predict minority performance). See also Paul R. Sackett, Wilfried De Corte, and Filip Lievens. Decision Aids for Addressing the Validity-Adverse Impact Trade-Off. pp. 453-472, 468 (discussing the Cleary criterion of fairness); Kelman, Mark (1991). Concepts of Discrimination in "General Ability" Job Testing. *Harvard Law Review*, 104(6), 1157-1247, 1223 (referring to Cleary's assertion that "[a] test is unbiased so long as it predicts minority performance on the job as well as it predicts nonminority performance."); Newman, Daniel A., Paul J. Hanges, and James L. Outtz. (2007). Racial Groups and Test Fairness, Considering History and Construct Validity. *American Psychologist, 62*(9), 1082-1083 (discussing the Cleary criterion of bias in testing).

[122] See, e.g., Linda Gottfredson, The Science and Politics of Race-Norming, Nov. 1994 American Psychologist, 955-963, 961 (discussing the unavoidable trade-off between false negatives and false positives). See also note – infra.

[123] On subgroup norming in the race context (race-norming), see, e.g., Dianne C. Brown, Subgroup Norming: Legitimate Testing Practice or Reverse Discrimination, Nov. 1994 American Psychologist, 927-928. See also Pyburn, Keith M Jr., Robert E. Ployhart, David A. Kravitz, The Diversity-Validity Dilemma: Overview and Legal Context, Personnel Psychology 61 (2008) 143-151, —; Sackett and Wilk, supra, at 929.

impact through the adjustment of the number of candidates selected. One common method is to employ "dual lists" and to select the best candidates from each group through a top-down selection process. Another is to establish distinct score or performance cut-offs for members of each group, or otherwise to relax qualifications for one group relative to another.[124]

Race-norming has long been popular among IOP experts and psychometricians. A consensus has developed that it represents the most efficient method for reducing disparate impact in the employment arena.[125] Indeed, in 1989 the National Academy of Sciences issued a report recommending race-conscious score adjustments on the General Aptitude Test Battery (GATB), a test that was widely used by the U.S. Employment Service of the Department of Labor to screen potential government employees.[126] The Academy endorsed the GATB exam as a valid, unbiased predictor of job success across the board, and acknowledged that the tests's racially disparate impact "is not due to [the test's] imperfections, but to substantial racial

---

[124] See, e.g., Mark Kelman, Harvard L. Rev., at 1241 (discussing race-norming, and describing how "employers might systematically and openly add points to black applicants' test scores or hire a higher proportion of black applicants with lower test scores").

[125] See e.g. Wayne F. Cascio, Rick Jacobs, and Jay Silva. Validity, Utility, and Adverse Impact: Practical Implications From 30 Years of Data. pp. 271-288, 282, in Outtz, Adverse Impact (noting that "race-norming" is "the single best way to maximize validity and utility simultaneously, while minimizing adverse impact."); Sackett, Paul R. and Steffanie L. Wilk. (1994). Within-Group Norming and Other Forms of Score Adjustments in Preemployment Testing. *American Psychologist*, *49*(11), 929-954, 931 (noting that race norming best reconciles the "competing goals" of "achieving productivity gains through the use of [a] selection device, but at the same time wanting to reduce or eliminate adverse impact against members of any group."); Sackett, Paul and Lawrence Roth, (1996), Multi-stage Selection Strategies: A Monte Carlo Investigation of Effects on Performance and Minority Hiring, *Personnel Psychology 49*(3), 549-572, 566 (noting that race-norming achieves greater diversity with less sacrifice in validity than alternative race-neutral adjustments, and commenting that none of the proposed alternatives "come[s] remotely close to a minority hiring rate consistent with minority representation in the applicant pool").

[126] See Hartigan, J.A., and Wigdor, A.K. eds, (1989), Fairness in employment testing: Validity generalization, minority issues, and the General Aptitude Test Battery (Washington, D.C., National Academy Press); See also Linda Gottfredson, The Science and Politics of Race-Norming, Nov. 1994 American Psychologist, 955-963.

differences in the job-related skills, abilities, and knowledge [the test] reveal[s]."[127] In justifying race-norming the GATB, the report explained that race-conscious selection best minimizes the costs of achieving diversity because it relaxes standards only for the minority population and allows a more competitive process to be retained for others. It thus maintains overall employee quality better than the generalized lowering of requirements that would otherwise be necessary to avoid an adverse impact.[128]

In further defending its recommendation to race-norm the GATB, the NAS report acknowledged the objection that a single race-blind cutoff excludes too many minorities who could do the job. Race-norming does indeed mitigate this exclusion by mandating the hiring or promotion of more minorities who would have been rejected. This reduces the number of black false negatives relative to whites, because relatively more blacks who would have been rejected are now hired. As with relaxing standards more generally, lowering the cutoff score for minorities is not without costs. Although previously false negatives are now pushed into the positive category, this produces more minority hires with lower scores. This increases the number of false positives (that is, failed hires) from the minority, lower scoring group. If the

---

[127] See Gottfredson, at 955, citing Wigdor, A.K & Garner, W.R. (Eds) (1982) Ability testing: uses, consequences, and controversies: Part 1: Report of the committee (Washington D.C., National Academy Press).

[128] This argument is similar to that made by Jeff Rosen for allowing affirmative action in education. He asserts that a bar on achieving diversity by lowering admission requirements selectively for lower- performing minority groups will lead to relaxing them across the board, thus compromising the academic quality of institutions generally. See Rosen, Jeff, How I Learned to Love Quotas, *New York Times Magazine*, June 1, 2003 (noting that "selective universities can't achieve colorblindness, diversity, and high admission standards at the same time," and expressing a preference for the "relatively modest concession represented by affirmative action" over the lowering of academic standards that will inevitably accompany rejecting that practice.) See also Lott, John R., Jr. (2000). Does a Helping Hand Put Others at Risk? Affirmative Action, Police Departments, and Crime. *Economic Inquiry*, *38*(2), 239-277, 249 (noting that "changing tests to employ a greater percentage of blacks can make it more difficult to screen out lower quality candidates generally.")

race-norming is pronounced, most of the people observed to fail on the job will be minorities.[129] That effect is exacerbated by the fact that the false positive rate is not independent of a candidate's test performance. The probability of failure increases as a candidate's score declines. Under race-norming, most of the lowest scoring individuals will be from the minority group, and relatively more of them will fail to meet performance standards.

By imposing a less exacting requirement for a lower performing groups while maintaining a higher and different standard for others, race-norming is a form of affirmative action that incorporates a race-conscious double standard. This places the practice at odds with the meritocratic underpinnings of disparate impact. It is also expressly illegal under the 1991 Civil Rights Act. Thus, although race-norming is an efficient way to achieve more diversity, its formal adoption is not presently feasible without a change in the law.[130]

Like the problem of adverse impact generally, the problem of false negatives has generated continuing concern, fueling repeated attempts to devise new personnel methods that effectively predict job performance while mitigating or even eliminating racially disparate impact. The hope is that this problem can be solved by developing more sensitive and accurate job screens. Unfortunately, efforts to modify selection methods to generate fewer false negatives

---

[129] On this point see, e.g., Linda Gottfredson, The Science and Politics of Race-Norming, Nov. 1994 American Psychologist, 955-963, 961, who provides a numerical example based on hypothetical scores on a typical job screening test, to show that "the same score adjustments that reduce the rate of false negatives" can also "increase the rate of false positives among blacks, because they bump many true negatives into the false positive category." The example of race-norming provided shows that, while the number of false negatives "falls significantly," "the rate of false positives increases from 17% to 59%," and "two thirds of all poor workers [among hires] would be black, despite blacks composing less than one third of all workers hired."

[130] As noted, Congress amended Title VII of the Civil Rights Act in 1991 to outlaw race-norming or racial adjustments in scores on job tests. See [note 84] supra. Further, in the wake of the Supreme Court's decision in *Ricci v. deStefano*, public employers have limited leeway to take race-conscious steps to avoid liability for the racially disparate impact of employment practices under the Constitutional Equal Protection guarantee, and those restrictions may extend to private employers also.

have not borne fruit, and there is also no guarantee that devising more "sensitive" methods – that is, those that are better able to identify successful workers – will reduce racially adverse impact. In any event, there are currently inherent limits to the accuracy of even the most predictive screens. As one IOP expert has noted, existing limits on prediction and measurement mean "there is clearly nothing an employer can do to design a better selection system." Multiple factors affect individual performance, but they are not all "knowable before hire," and thus are not amenable to "accurate testing ahead of time." Thus, even conceding that current methods exclude many blacks (and other candidates) who could do the job, "the optimal selection system" remains one with substantial disparate impact.[131] In sum, given the limits inherent in identifying good workers, attempting to produce more diversity by fiddling with job criteria or creating more predictive screens is unlikely to work. In any event, the real problem lies in the distribution of human capital, not in the instruments used to measure it. As long as there is a relative shortage of skilled minority workers, adverse impact is likely to persist.

Finally, the proposal to repeal the disparate impact rule must be compared to an even more radical alternative, which is to abolish *ex ante* requirements altogether and shift to a system of probationary hiring. Mark Kelman has proposed that employers abandon all pre-hire screening in favor of random selection. Managers would assess on-the job performance directly and discharge individuals whose work falls short.[132] This proposal has serious drawbacks. Where positions are scarce relative to job-seekers, employers will miss out on applicants who are better matched to jobs than those likely to be chosen, with significant costs to efficiency. Second, even if the number of minority false negatives could be somewhat reduced, this would not necessarily solve the problem of adverse impact. Although racially proportionate hiring

---

[131] Sackett and Wilk, supra, at 934.

[132] See discussion at – supra; Kelman, Harv. L. Rev. At 1226. See also Sackett and Wilk, at 994 ("If one gave all applicants a job tryout and kept the highest performers, one would retain more blacks than would be hired" using screens with greater black-white differences than actual performance differences).

could be accomplished *ex ante*, imbalances will almost surely emerge *ex post*. Screening after hiring rather than before does nothing to alter the background distribution in job-related skills. More minorities will likely be fired following the probationary period, thus reintroducing the problem of adverse impact on the back end.[133] This result would be achieved at the cost of lowering productivity and imposing a weighty burden on employers to supervise, evaluate, and deal with probationary hires with a wide range of proficiency. Considerable investments in on-the-job training would be lost, and the high risk of being fired would deter employees from developing job-related skills.[134]

Nonetheless, Kelman's proposal suggests an alternative to current practice and a possible way out of the dilemma it poses. As noted above, a proper understanding and application of disparate impact rule is unlikely to achieve greater workforce racial balance, and indeed might move the situation in the opposite direction. If diversity is the goal, however, there are other ways to achieve it. Directly enacting race-based affirmative action in the workplace might prove difficult politically, and may run afoul of the Equal Protection Clause of the Constitution. However, as noted above, there is no general requirement to adopt a system of competitive meritocratic job selection. A more diverse government workforce – including fire and police departments – can be achieved by relaxing job requirements or abandoning meritocratic criteria. Disparate impact rules would appear to allow employers to use a screen with less adverse impact, even if it were less predictive of job success. So civil service exams could be selected or

---

[133] This will reintroduce familiar challenges to the fairness of on-the-job assessments and raise suspicions concerning the "discretion inherent in ex post screening systems." See Kelman, Harv. L. Rev. at 1233.

[134] See, e.g., Douglas O. Staiger and Jonah E. Rockoff, Searching for Effective Teachers with Imperfect Information, 24 J of Economic Perspectives 97-118 , 98, 115 (2010)(recommending a new approach to hiring teachers that combines "an easy entry policy" for all college graduates with "an aggressive dismissal policy" that "identif[ies] large differences between teachers by observing the first few years of teaching performance and retaining only the highest performing teachers," but expressing the reservation that potential teachers "might be uneasy about investing time and effort in the difficult first years of teaching" in the face of a significant probability of being fired.)

re-designed to reduce adverse impact. To be sure, there are unresolved questions in the wake of the *Ricci* decision about the circumstances in which employers are permitted to resolve the diversity-validity tradeoff in favor of more diversity. Although the Court invalidated a decision to discard an existing test that was motivated by a desire to avoid a racially disparate impact, that case involved identifiable victims with reliance interests in the results. The opinion does not necessarily cast aspersions on the full range of choices that employers might take to achieve greater diversity.

The diversity-validity tradeoff indicates that any test that significantly reduces adverse impact will also seriously compromise validity. The question is then: why bother with a test at all? Diversity might be better achieved, at little additional cost to efficiency, by dropping civil service exams altogether in favor of minimal threshold requirements coupled with lotteries or methods of random selection. Because many civil service laws and labor contracts mandate competitive exams, a shift to this system would might require some political will.[135] This method might ultimately prove more efficient, and certainly would be much simpler, than elaborate litigation based on the chimerical, elusive quest for civil service screens with less adverse impact. Adopting this method would require facing up to tradeoffs and giving up on we the fiction that the diversity-validity tradeoff can be overcome. We may have to accept that a more diverse civil service may come at the cost of a less capable one.

VI. *Conclusion*

Under present job market conditions, the diversity-validity tradeoff prevails. The more predictive a job selection device, the greater its adverse impact. To date there is no known way around this dilemma. The situation is not just an artifact of limitations in our assessment measures (although they do have limitations). Rather, it reflects real underlying differences in developed abilities, and the real impact these abilities have on job performance. These gaps in measured human capital are in turn traceable to complex social circumstances with myriad

---

[135] See Rutherglen, 2009 Supreme Court Review, at 113.

present and historical roots.

The argument for abolishing disparate impact doctrine is that its rules do not currently comport with reality and do little to change it. The rules goals and assumptions are at odds with what we know about educational disparities, productivity, job performance, and human capital, The 4/5 rule and other elements of DI doctrine embody this mismatch. Above all, disparate impact litigation represents a costly, misplaced effort that distracts from the root causes of workforce imbalance and draws resources away from the measures needed to address it.[136] In light of this reality, disparate impact should be altered or abolished.

---

[136] See, e.g., Russell K. Nieli, Competitive Colleges: Addressing Minority Performance Gaps, 23 Academic Questions 358, 355-356 (Fall 2010)(noting that "[i]mproving the academic performance of under-represented minority students constitutes the only viable long-run strategy for preserving meaningful minority representation" in higher education and in demanding jobs, and asserting that "[we] need to make closing the racial achievement gap a high social priority and to move aggressively with the greatest determination to make it happen"). Relaxing the standard for disparate impact liability, or abolishing it altogether, could have other salutary effects. The college admissions process and college degree requirement that employers impose currently serves as a form of screening for ability. Eliminating liability for disparate impact would free employers to test for ability directly without worrying about racial effects, thus making the acquisition of expensive and prestigious educational credentials less important. This change would also encourage potential employees to hone the skills that employers are seeking. See, e.g., Jonathan V. Last, America's One-Child Policy, The Weekly Standard, Sept. 27, 2010, 22-30, 29 (suggesting that "[i]f *Griggs* were rolled back, it would upend the college system at a stroke" because it would free employers to test directly for general ability, thus decreasing the importance of educational credentials that employers rely on for screening).

Paul R. Sackett, Wilfried De Corte, and Filip Lievens from Adverse Impact by James L. Outtz

Figure 1A

Minority Group Selection Ratios and Four-Fifths Ratios When the Majority Group Selection Ratio Is 1%, 5%, 10%, 25%, 50%, 75%, 90%, 95%, or 99%

Standardized group difference (*d*)	Majority group selection ratio[a]								
	1%	5%	10%	25%	50%	75%	90%	95%	99%
0.0	.010	.050	.100	.250	.500	.750	.900	.950	.990
	1.00	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
0.1	.008	.041	.084	.221	.460	.716	.881	.938	.987
	.80	**.82**	**.84**	**.88**	**.92**	**.95**	**.98**	**.99**	**.99**
0.2	.006	.033	.069	.192	.421	.681	.860	.925	.983
	.60	**.66**	**.69**	**.77**	**.84**	**.91**	**.96**	**.97**	**.99**
0.3	.004	.026	.057	.166	.382	.644	.837	.910	.978
	.40	**.52**	**.57**	**.66**	**.76**	**.86**	**.93**	**.96**	**.99**
0.4	.003	.021	.046	.142	.345	.606	.811	.893	.973
	.30	**.42**	**.46**	**.57**	**.69**	**.81**	**.90**	**.94**	**.98**
0.5	.002	.016	.038	.121	.309	.568	.782	.873	.966
	.20	**.32**	**.38**	**.48**	**.62**	**.76**	**.87**	**.92**	**.98**
0.6	.002	.013	.030	.102	.274	.528	.752	.851	.957
	.20	**.26**	**.30**	**.41**	**.55**	**.70**	**.84**	**.90**	**.97**
0.7	.001	.010	.024	.085	.242	.488	.719	.826	.947
	.10	**.20**	**.24**	**.34**	**.48**	**.65**	**.80**	**.87**	**.96**
0.8	.001	.007	.019	.071	.212	.448	.684	.800	.936
	.10	**.14**	**.19**	**.28**	**.42**	**.60**	**.76**	**.84**	**.95**
0.9	.001	.006	.015	.058	.184	.409	.648	.770	.922
	.10	**.12**	**.15**	**.23**	**.37**	**.54**	**.72**	**.81**	**.93**
1.0	.000	.004	.011	.047	.159	.371	.610	.739	.907
	.00	**.08**	**.11**	**.19**	**.32**	**.49**	**.68**	**.78**	**.92**
1.1	.000	.003	.009	.038	.136	.334	.571	.705	.889
	.00	**.06**	**.09**	**.15**	**.27**	**.45**	**.63**	**.74**	**.90**
1.2	.000	.002	.007	.031	.115	.298	.532	.670	.869
	.00	**.04**	**.07**	**.12**	**.23**	**.40**	**.59**	**.71**	**.88**
1.3	.000	.002	.005	.024	.097	.264	.492	.633	.846
	.00	**.04**	**.05**	**.10**	**.19**	**.35**	**.55**	**.67**	**.85**
1.4	.000	.001	.004	.019	.081	.233	.452	.595	.821
	.00	**.02**	**.04**	**.08**	**.16**	**.31**	**.50**	**.63**	**.83**
1.5	.000	.001	.003	.015	.067	.203	.413	.556	.794
	.00	**.02**	**.03**	**.06**	**.13**	**.27**	**.46**	**.59**	**.80**

[a] Selection ratio = number of applicants hired/number of applicants applied. Per cell, two values are given. The first value refers to the minority group selection ratio. The second value in bold represents the four-fifths ratio (i.e., the minority group selection ratio/

Paul R. Sackett, Wilfried De Corte, and Filip Lievens from Adverse Impact by James L. Outtz

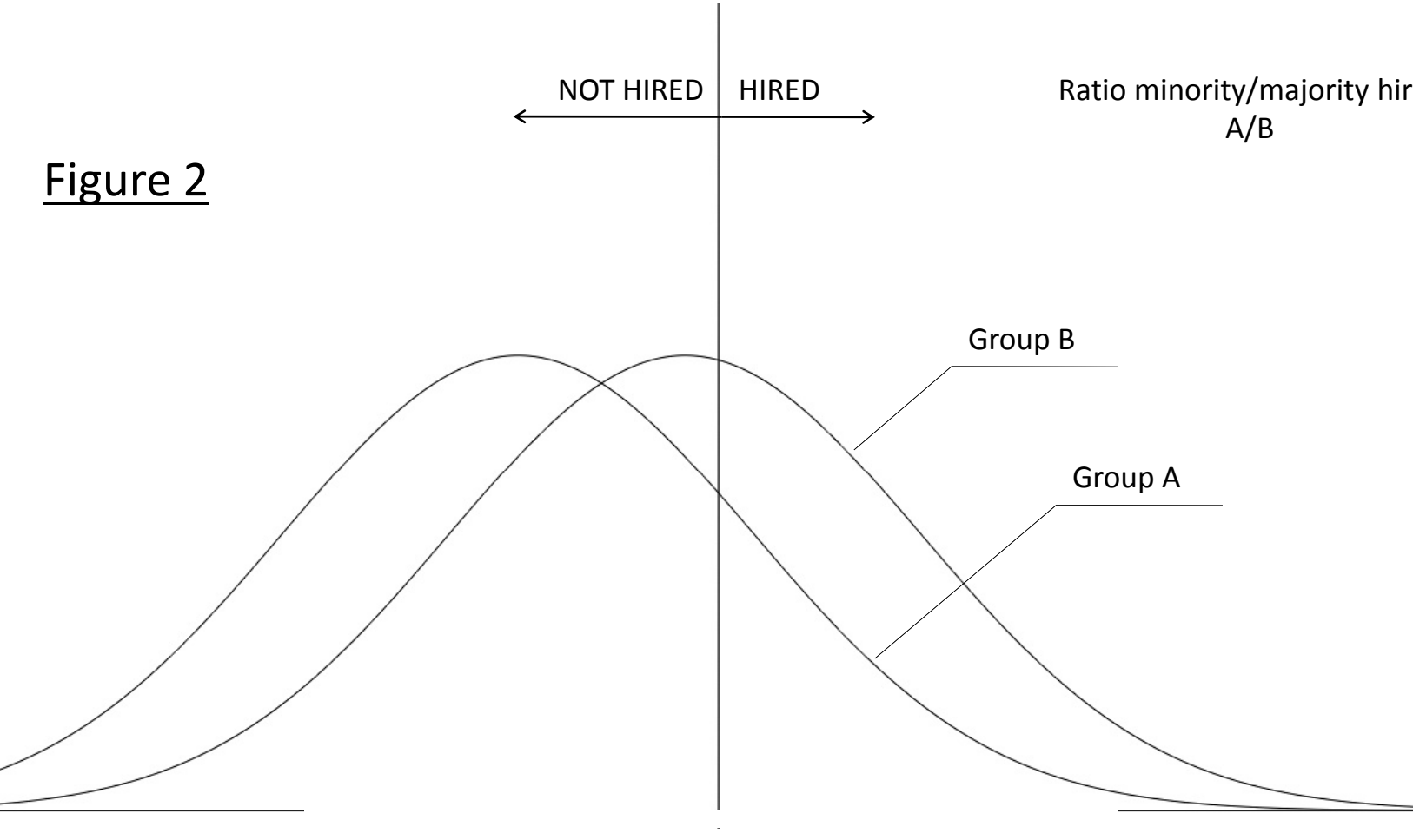
Figure 1B

Minority Group Selection Ratios and Four-Fifths Ratios When the Majority Group Selection Ratio Is 1%, 5%, 10%, 25%, 50%, 75%, 90%, 95%, or 99%

Standardized group difference (*d*)	Majority group selection ratio[a]								
	1%	5%	10%	25%	50%	75%	90%	95%	99%
0.0	.010	.050	.100	.250	.500	.750	.900	.950	.990
	1.00	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
0.1	.008	.041	.084	.221	.460	.716	.881	.938	.987
	.80	**.82**	**.84**	**.88**	**.92**	**.95**	**.98**	**.99**	**.99**
0.2	.006	.033	.069	.192	.421	.681	.860	.925	.983
	.60	**.66**	**.69**	**.77**	**.84**	**.91**	**.96**	**.97**	**.99**
0.3	.004	.026	.057	.166	.382	.644	.837	.910	.978
	.40	**.52**	**.57**	**.66**	**.76**	**.86**	**.93**	**.96**	**.99**
0.4	.003	.021	.046	.142	.345	.606	.811	.893	.973
	.30	**.42**	**.46**	**.57**	**.69**	**.81**	**.90**	**.94**	**.98**
0.5	.002	.016	.038	.121	.309	.568	.782	.873	.966
	.20	**.32**	**.38**	**.48**	**.62**	**.76**	**.87**	**.92**	**.98**
0.6	.002	.013	.030	.102	.274	.528	.752	.851	.957
	.20	**.26**	**.30**	**.41**	**.55**	**.70**	**.84**	**.90**	**.97**
0.7	.001	.010	.024	.085	.242	.488	.719	.826	.947
	.10	**.20**	**.24**	**.34**	**.48**	**.65**	**.80**	**.87**	**.96**
0.8	.001	.007	.019	.071	.212	.448	.684	.800	.936
	.10	**.14**	**.19**	**.28**	**.42**	**.60**	**.76**	**.84**	**.95**
0.9	.001	.006	.015	.058	.184	.409	.648	.770	.922
	.10	**.12**	**.15**	**.23**	**.37**	**.54**	**.72**	**.81**	**.93**
1.0	.000	.004	.011	.047	.159	.371	.610	.739	.907
	.00	**.08**	**.11**	**.19**	**.32**	**.49**	**.68**	**.78**	**.92**
1.1	.000	.003	.009	.038	.136	.334	.571	.705	.889
	.00	**.06**	**.09**	**.15**	**.27**	**.45**	**.63**	**.74**	**.90**
1.2	.000	.002	.007	.031	.115	.298	.532	.670	.869
	.00	**.04**	**.07**	**.12**	**.23**	**.40**	**.59**	**.71**	**.88**
1.3	.000	.002	.005	.024	.097	.264	.492	.633	.846
	.00	**.04**	**.05**	**.10**	**.19**	**.35**	**.55**	**.67**	**.85**
1.4	.000	.001	.004	.019	.081	.233	.452	.595	.821
	.00	**.02**	**.04**	**.08**	**.16**	**.31**	**.50**	**.63**	**.83**
1.5	.000	.001	.003	.015	.067	.203	.413	.556	.794
	.00	**.02**	**.03**	**.06**	**.13**	**.27**	**.46**	**.59**	**.80**

[a] Selection ratio = number of applicants hired/number of applicants applied. Per cell, two values are given. The first value refers to the minority group selection ratio. The second value in bold represents the four-fifths ratio (i.e., the minority group selection ratio/

Figure 2


NOT HIRED
HIRED
Ratio minority/majority hir
A/B
Group B
Group A
Distribution of Performance
(screening criterion; or on-the-job)