

University of Pennsylvania Carey Law School

Penn Carey Law: Legal Scholarship Repository

All Faculty Scholarship

Faculty Works

Fall 10-1-2011

Disparate Impact Realism

Amy L. Wax

University of Pennsylvania Carey Law School

Author ORCID Identifier:

 Amy Wax 0000-0002-8200-5555

Follow this and additional works at: https://scholarship.law.upenn.edu/faculty_scholarship



Part of the [Business Law, Public Responsibility, and Ethics Commons](#), [Civil Rights and Discrimination Commons](#), [Labor and Employment Law Commons](#), [Law and Economics Commons](#), [Policy Design, Analysis, and Evaluation Commons](#), [Public Law and Legal Theory Commons](#), and the [Race and Ethnicity Commons](#)

Repository Citation

Wax, Amy L., "Disparate Impact Realism" (2011). *All Faculty Scholarship*. 352.
https://scholarship.law.upenn.edu/faculty_scholarship/352

This Article is brought to you for free and open access by the Faculty Works at Penn Carey Law: Legal Scholarship Repository. It has been accepted for inclusion in All Faculty Scholarship by an authorized administrator of Penn Carey Law: Legal Scholarship Repository. For more information, please contact PennlawIR@law.upenn.edu.

DISPARATE IMPACT REALISM

AMY L. WAX*

TABLE OF CONTENTS

| | |
|---|-----|
| INTRODUCTION AND SUMMARY | 622 |
| I. DISPARATE IMPACT EMPLOYMENT DISCRIMINATION: THE DOCTRINE AND ITS UNCERTAINTIES | 626 |
| II. EVIDENCE ON DISPARATE IMPACT: INDUSTRIAL AND ORGANIZATIONAL PSYCHOLOGY (IOP) RESEARCH | 638 |
| <i>A. Job Screening Methods and Predictions of Job Performance</i> | 640 |
| <i>B. Job Screening Methods and Adverse Impact: The Validity-Diversity Tradeoff</i> | 644 |
| III. SQUARING THE CIRCLE: COMPLYING WITH THE REQUIREMENTS OF DISPARATE IMPACT | 657 |
| IV. REFORMING DISPARATE IMPACT: ALTER OR ABOLISH IT ... | 664 |
| <i>A. Alter It: Disparate Impact Realism</i> | 666 |
| <i>B. Abolish It</i> | 693 |
| <i>C. The Problem of False Negatives</i> | 702 |
| CONCLUSION | 711 |

* Robert Mundheim Professor of Law, University of Pennsylvania Law School.

INTRODUCTION AND SUMMARY

In *Ricci v. DeStefano*,¹ the Supreme Court recently reaffirmed the doctrine, first articulated by the Court in *Griggs v. Duke Power Co.*,² that employers can be held liable under Title VII of the Civil Rights Act of 1964 (Title VII) for neutral policies with a disparate impact on minority workers.³ The Court has further held that employers can escape liability by showing that a policy is job-related or consistent with business necessity.⁴

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.⁵ 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.⁶ The so-called four-fifths rule, under which an employer is presumptively liable if the rate of hiring for minority workers is less than 80 percent of the rate for the majority white population, reflects this assumption.⁷ Second, the Court in *Griggs* noted the absence of evidence that the screening criteria in that case—a high school diploma and scores on a “professionally prepared aptitude test[.]”—were related to subsequent performance of the service jobs at issue, and expressed doubt about the existence of such a relationship.⁸

1. 129 S. Ct. 2658 (2009).

2. 401 U.S. 424 (1971).

3. See *Ricci*, 129 S. Ct. at 2672-73, 2676; *Griggs*, 401 U.S. at 429-32.

4. *Griggs*, 401 U.S. at 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.”).

5. See *infra* Part III.

6. See *Griggs*, 401 U.S. at 429-30.

7. See Uniform Guidelines on Employee Selection Procedures, 29 C.F.R. § 1607.4(D) (2010); *infra* notes 48-54 and accompanying text.

8. *Griggs*, 401 U.S. at 427-28 (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 *id.* at 431-32 (noting the absence of a demonstrable connection between the tests at issue and job performance and observing that

Social science research casts doubt on both of these aspects of *Griggs*. First, research in industrial and organizational 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) remain the best predictors of performance for jobs at all levels of complexity.⁹ 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.¹⁰ These gaps are reflected in lower scores on the types of g-loaded job screens that best predict job success.¹¹ 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.¹² Indeed, the evidence indicates that most valid screening devices will have a significant adverse impact on blacks and will also violate the four-fifths rule under the law of disparate impact.¹³

In sum, the IOP literature demonstrates that the empirical and demographic premises behind the disparate impact rule do not match reality and have turned out to be myths.¹⁴ As a consequence, most *legitimate* job selection practices, including those that predict productivity better than alternatives, will routinely trigger liability under the current rule.¹⁵ Although the Supreme Court in *Griggs* and subsequent cases has repeatedly stated that disparate impact doctrine is consistent with a rigorously competitive meritocracy,¹⁶

workers at the company who had not graduated from high school or taken the test were performing in a satisfactory manner).

9. See *infra* Part II.B. The tests at issue in *Griggs* were the Wonderlic Personnel Test, a standard type of intelligence test, and the Bennett Mechanical Comprehension Test. *Griggs*, 401 U.S. at 428.

10. See *infra* Part II.B.

11. See *infra* Part II.A.

12. See *infra* Part II.B.

13. See *infra* Part II.B.

14. See *infra* Part II.

15. See *infra* Part III.

16. See Kenneth R. Davis, *Wheel of Fortune: A Critique of the “Manifest Imbalance” Requirement for Race-Conscious Affirmative Action Under Title VII*, 43 GA. L. REV. 993, 1037-39 (2009) (citing *Griggs v. Duke Power Co.*, 401 U.S. 424, 431 (1971)).

employers seeking to maintain such a meritocracy among a diverse population will run a high risk of being sued for violations of the rule.¹⁷ Such lawsuits will put employers to the onerous, uncertain, and sometimes impossible task of justifying their job selection practices. This may result in unwarranted liability or induce undesirable, self-protective strategies.¹⁸ Even in the absence of those consequences, a proper application of the doctrine is unlikely to change the racial composition of the workplace or to increase demographic diversity. The best explanation for current workforce imbalances is the existence of real average group differences in knowledge, skills, and abilities. These human capital disparities, and not the use of non-merit-related selection or the erection of arbitrary barriers, best explain observed employment patterns. And given the present magnitude of skill differences and the shortage of qualified minority workers, the correct application of the disparate impact rule will not increase workforce diversity and could well make some jobs less diverse.

In light of these observations, this Article proposes to modify the doctrine of disparate impact to adopt a new standard of “disparate impact realism.”¹⁹ The disparate impact rule should be revised by making two changes in the standard that triggers potential liability. First, the target four-fifths 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 nature of the threshold ratio should be abandoned in favor of a sliding scale relationship, documented in the IOP literature,²⁰ that pegs expected group staffing patterns to measured disparities in group performance and the selectivity of particular positions.

Although it does not altogether relieve 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

17. *See infra* Part IV.B.

18. *See infra* Part IV.B.

19. *See infra* Part IV.A.

20. *See infra* Part IV.A.

potential constitutional difficulties generated by the Supreme Court's recent decision in *Ricci v. DeStefano*²¹ make it desirable to cut down on the number of situations that can generate disparate impact claims.²² Finally, disparate impact realism functions as an information-forcing device.²³ By making it easier for employers to satisfy the rule, and by 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 abolishing the disparate impact rule altogether.²⁴ The principal argument for repeal 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 the distribution of skill and human capital best explains observed patterns of adverse impact. The gaps by race in developed abilities, not race-based exclusion or arbitrary barriers to employment, are the principal factors behind racial imbalances on the job.²⁵ 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 used selection procedures are valid and job-related and thus do not actually violate the disparate impact rule.²⁶ 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.²⁷ 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 defenses are valid and even though they are not actually violating the rule.²⁸ Because virtually no aspect of the business necessity defense is settled law, employers face the

21. 129 S. Ct. 2658 (2009).

22. See *infra* note 195 and accompanying text.

23. See *infra* Part IV.A.

24. See *infra* Part IV.B.

25. See *infra* Part III.

26. See *infra* Part II.B.

27. See *infra* Part IV.A.

28. See *infra* Part IV.A.

prospect of protracted, expensive, uncertain, and resource-intensive litigation to defend their practices.²⁹ This encourages them to engage in perverse, inefficient, and evasive tactics, including de facto affirmative action.³⁰ And even if employers avoid this response or successfully resist a disparate impact challenge, they may be required to expend significant time, attention, and resources on defending their personnel selection methods. In sum, the overbreadth of the disparate impact rule is both inefficient and fundamentally unfair. If properly applied, the rule will prove costly but do little or nothing to increase minority representation in the workforce. And the racial preferences that may result are directly at odds with the meritocratic goals of disparate impact principles.

The data currently reveal that most jobs are more diverse than disparate impact doctrine actually requires.³¹ Indeed, blacks lag behind whites in performance on the job in many categories.³² This indicates that employers are not arbitrarily excluding minorities from the workforce but are rather bending over backwards to include them. In addition, disparate impact litigation does nothing to correct the underlying skill deficits reflected in these on-the-job gaps and drains resources from that task.³³ The doctrine represents a costly, misplaced effort that could better be directed at addressing the root causes of workforce racial imbalance.

I. DISPARATE IMPACT EMPLOYMENT DISCRIMINATION: THE DOCTRINE AND ITS UNCERTAINTIES

In *Griggs v. Duke Power Co.*,³⁴ 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 of 1964.³⁵ The Court further held that an employer could escape

29. See *infra* Part IV.A.

30. See *infra* Part IV.B.

31. See *infra* Part IV.A.

32. See *infra* notes 114-19 and accompanying text.

33. See *infra* Part IV.B.

34. 401 U.S. 424 (1971).

35. *Id.* at 430. 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

liability if the job criteria at issue were shown to have “a manifest relationship to the employment in question.”³⁶ As the Court stated, “[t]he touchstone is business necessity.”³⁷ The Court ruled in *Griggs* that the job requirements in that case—a high school diploma and a minimum score on an IQ test—were impermissible because they screened out too many black applicants and were not shown to be “related to job performance” for the positions in question.³⁸

In *Griggs* and subsequent cases expanding on the disparate impact doctrine, the Supreme Court has repeatedly stated that disparate impact rules do not mandate a particular racial balance or ethnic makeup in the workplace.³⁹ Rather, the objective is to “achieve equality of employment opportunities and remove barriers that have operated in the past to favor an identifiable group.”⁴⁰ 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.⁴¹ On this view, employers are entitled to hold all job seekers to uniform requirements, as long as they are work-related.⁴² The goal of the disparate impact doctrine is therefore at odds with race-conscious double standards or forms of affirmative

a protected characteristic, such as race or sex. See David Sherwyn & Michael Heise, *The Gross Beast of Burden of Proof: Experimental Evidence on How the Burden of Proof Influences Employment Discrimination Case Outcomes*, 42 ARIZ. ST. L.J. 901, 905 (2010) (“Most 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.” (footnote omitted)); see also *McDonnell Douglas Corp. v. Green*, 411 U.S. 792, 802 (1973) (finding unlawful disparate treatment in a case of adverse treatment motivated by race).

36. *Griggs*, 401 U.S. at 432.

37. *Id.* at 431.

38. See *id.* at 427-28, 431-32.

39. See, e.g., *id.* at 429-31; *Watson v. Fort Worth Bank & Trust*, 487 U.S. 977, 992-93 (1988); see also Civil Rights Act of 1991 § 105(a), 42 U.S.C. § 2000e-2(j) (2006).

40. *Griggs*, 401 U.S. at 429-30.

41. See Davis, *supra* note 16, at 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.” (footnote omitted)); see also Cheryl I. Harris & Kimberly West-Faulcon, *Reading Ricci: Whitening Discrimination, Racial Test Fairness*, 58 UCLA L. REV. 73, 156-57 (2010) (“Title VII promotes selection that is more merit-based and thus is not a mechanism to enact racial preferences.”).

42. See Davis, *supra* note 16, at 1037-38.

action that selectively maintain less exacting criteria for some social or racial groups.⁴³

In the wake of *Griggs*, the Equal Employment Opportunity Commission (EEOC) and the courts have been charged with crafting rules consistent with the doctrine's stated goals and creating guidelines for their application.⁴⁴ The Supreme Court has addressed adverse impact in only a few cases, and its decisions are vague or silent on key questions that routinely arise in litigation.⁴⁵ 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.⁴⁶ Courts have also been asked to clarify the standards for job-relatedness and to determine the scope of the business necessity defense that employers can invoke to escape liability once a prima facie is demonstrated.⁴⁷ 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, or 80-

43. That the goals of disparate impact enforcement are inconsistent with race-based affirmative action has been widely and consistently recognized. *See, e.g.*, Susan O. Carle, *A Social Movement History of Title VII Disparate Impact Analysis*, 63 FLA. L. REV. 251, 258-59 (2011) ("Popular perception sometimes conflates disparate impact analysis with affirmative action, but the two antidiscrimination concepts are actually quite different." (footnote omitted)); *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"); Harris & West-Faulcon, *supra* note 41, at 156-57. *See generally* George Rutherglen, *Ricci v. DeStefano: Affirmative Action and the Lessons of Adversity*, 2009 SUP. CT. REV. 83, 83-84. Attempts to use disparate impact doctrine to engage in 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 Civil Rights Act of 1991. *See* Civil Rights Act of 1991 §§ 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* Part IV.A.

44. *See* Civil Rights Act of 1991 § 713.

45. *See* Sherwyn & Heise, *supra* note 35, at 905 ("[T]he Supreme Court has addressed adverse impact cases on a few occasions.").

46. *See infra* Part IV.A.

47. *See infra* Part IV.A.

percent rule, as a touchstone for determining adverse impact.⁴⁸ Hiring or promoting minorities at less than 80 percent of the rate for the majority group gives rise to a potential violation and suffices to establish a prima facie case of disparate impact discrimination.⁴⁹ Partly in response to critiques of the four-fifths rule as insufficiently nuanced and statistically naive,⁵⁰ the courts have not relied exclusively on this metric, with some scrutinizing workplace diversity using commonplace tests of statistical significance.⁵¹ Although the courts and the EEOC permit recourse to alternative

48. See Uniform Guidelines on Employee Selection Procedures, 29 C.F.R. § 1607.4(D) (2010) (“[A] selection rate for any race, sex, or ethnic group which is less than four-fifths (4/5) (or eighty percent) of the rate for the group with the highest rate will generally be regarded ... as evidence of adverse impact.”). Courts are not bound by the Guidelines, but the Supreme Court has said that they should receive “great deference.” *Griggs v. Duke Power Co.*, 401 U.S. 424, 433-34 (1971). The Supreme Court embraced the four-fifths rule in *Connecticut v. Teal*, 457 U.S. 440, 443 n.4 (1982), and subsequent decisions. See, e.g., *Ricci v. DeStefano*, 129 S. Ct. 2658, 2678 (2009); *Watson v. Fort Worth Bank & Trust*, 487 U.S. 977, 995 n.3 (1988). Lower courts have repeatedly applied the doctrine in disparate impact cases. See, e.g., *Bradley v. City of Lynn*, 433 F. Supp. 2d 157, 159 (D. Mass. 2006); *Thompson v. Miss. State Personnel Bd.*, 674 F. Supp. 198, 206 (N.D. Miss. 1987). For recent reviews of the standard for a prima facie case, see Harris & West-Faulcon, *supra* note 41, at 135-36; Rutherglen, *supra* note 43, at 105-06.

49. See 29 C.F.R. § 1607.4(D).

50. See, e.g., Nancy T. Tippins, *Adverse Impact in Employee Selection Procedures from the Perspective of an Organizational Consultant*, in ADVERSE IMPACT: IMPLICATIONS FOR ORGANIZATIONAL STAFFING AND HIGH STAKES SELECTION 201, 203 (James L. Outtz ed., 2010) [hereinafter ADVERSE IMPACT]; Sheldon Zedeck, *Adverse Impact: History and Evolution*, in ADVERSE IMPACT, *supra*, at 3, 14-15 (suggesting that, under some circumstances, a four-fifths rule violation could easily occur by chance).

51. See, e.g., *Clady v. Cnty. of L.A.*, 770 F.2d 1421, 1428-29 (9th Cir. 1985).

methods,⁵² the four-fifths rule remains an important benchmark for establishing a presumptive disparate impact violation.

In applying the four-fifths 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 four-fifths target should be gauged against a broader cohort, such as the adult work-eligible population, or defined more narrowly to include, for example, actual job applicants or work-eligible persons possessing threshold qualifications.⁵³ All of these 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.⁵⁴ Threshold requirements of any kind can end up screening out minority applicants.

52. See 29 C.F.R. § 1607.4(D) (stating that smaller differences in selection rate than dictated by the four-fifths rule may constitute adverse impact differences where these are “significant in both statistical and practical terms”); Jennifer L. Peresie, *Toward a Coherent Test for Disparate Impact Discrimination*, 84 IND. L.J. 773, 777 (2009) (“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 *Hazelwood Sch. 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” (internal quotation marks omitted)). See generally Philip Bobko & Philip L. Roth, *An Analysis of Two Methods for Assessing and Indexing Adverse Impact: A Disconnect Between the Academic Literature and Some Practice*, in *ADVERSE IMPACT*, *supra* note 50, at 29-34 (noting the use of the four-fifths rule as well as various tests of statistical significance in disparate impact cases and observing that the EEOC endorses the four-fifths rule but also permits the use of recognized tests of statistical significance in conjunction with or in place of that criterion in appropriate cases).

53. See Peresie, *supra* note 52, at 778 (“Plaintiffs must establish disparate impact with respect to the pool of qualified persons in the relevant labor market for the given 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.” (footnotes omitted)).

54. See, e.g., Scott Baker, Comment, *Defining “Otherwise Qualified Applicants”: Applying an Antitrust Relevant-Market Analysis to Disparate Impact Cases*, 67 U. CHI. L. REV. 725, 732 (2000) (“The obvious flaw [in using only actual applicants 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.”).

Thus using such requirements to define potential candidates is itself vulnerable to challenge under the disparate impact rule.⁵⁵ 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.⁵⁶ This aspect of disparate impact doctrine is in serious disarray.

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.⁵⁷ 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, construct, and criterion validation.⁵⁸ Both construct and criterion validation require demonstrating

55. 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, are 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. See *Griggs v. Duke Power Co.*, 401 U.S. 424, 429-31 (1971). The actual practice does not always adhere to this analysis. See Ian Ayres, Testing for Discrimination and the Problem of “Included Variable Bias” 54 (Oct. 6, 2010) (unpublished manuscript), available at <http://www.law.upenn.edu/currently/seminars/lawandeconomics/papers/AyresIncludedVariableBias.pdf>. On the problem of identifying the baseline population for purposes of disparate impact analysis, see *Wards Cove Packing Co. v. Atonio*, 490 U.S. 642, 650-55 (1989) (raising the question of how to define the eligible pool for disparate impact scrutiny). See generally Martha Chamallas, *Evolving Conceptions of Equality Under Title VII: Disparate Impact Theory and the Demise of the Bottom Line Principle*, 31 UCLA L. REV. 305 (1983).

56. See Baker, *supra* note 54, at 732 (“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.”); *id.* at 728-31 (discussing cases); Joseph L. Gastwirth, *Employment Discrimination: A Statistician’s Look at Analysis of Disparate Impact Claims*, 11 LAW & INEQ. 151, 157-58, 167 (1992) (discussing the uncertainty surrounding the identification of the majority and minority labor market pool available for particular jobs); see also, e.g., *NAACP v. Town of E. Haven*, 998 F. Supp. 176, 186-88 (D. Conn. 1998) (deeming applicant pool to consist of qualified teachers from defined geographical area); Peresie, *supra* note 52, at 778.

57. See Harris & West-Faulcon, *supra* note 41, at 144-45.

58. See Uniform Guidelines on Employee Selection Procedures, 29 C.F.R. § 1607.14 (2010). For a description of current EEOC guidelines, see Brief of Industrial-Organizational Psychologists as Amici Curiae in Support of Respondents at 5-7, *Ricci v. DeStefano*, 129 S. Ct. 2658 (2009) (Nos. 07-1428 & 08-328) [hereinafter IOP Brief]. The inference of validity

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).⁵⁹ As the most rigorous and demanding process, criterion validation is considered the gold standard and is the focus of considerable study by IOP experts.⁶⁰ In contrast, content or "facial" validation is regarded as less exacting.⁶¹ Content validation does not generally require a formal demonstration that a job criterion actually predicts superior job performance or productivity.⁶² Rather, it depends on showing a manifest relationship or plausible match between the abilities assessed and the tasks that must actually be performed on the job.⁶³

The intricacies of the validation rules compound the burden of defending personnel practices. The details of the Guidelines are complex and ambiguous, and courts' standards are erratic. Although the Guidelines recommend formal statistical validation, they also accept the use of content validation—the demonstration that selection criteria are closely geared to job tasks—without clarifying which methods are appropriate in which cases.⁶⁴ In the same vein, the Supreme Court made clear in *Watson v. Fort Worth Bank &*

rests on:

evidence ... that a test identifies those who are qualified to do the job. 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. 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.

Id. at 7 n.2 (citation omitted); see also Robert Belton, *The Unfinished Agenda of the Civil Rights Act of 1991*, 45 RUTGERS L. REV. 921, 939-42 (1993); Harris & West-Faulcon, *supra* note 41, at 144-47 (reviewing validation standards); Rutherglen, *supra* note 43, at 106-07 (reviewing validation standards).

59. See *supra* note 58.

60. See, e.g., *supra* note 58.

61. See, e.g., *supra* note 58.

62. See IOP Brief, *supra* note 58, at 7 n.2.

63. See Mark Kelman, *Concepts of Discrimination in "General Ability" Job Testing*, 104 HARV. L. REV. 1157, 1171 & n.35 (1991) (discussing forms of validation); Philip L. Roth, Philip Bobko & Lynn A. McFarland, *A Meta-Analysis of Work Sample Test Validity: Updating and Integrating Some Classic Literature*, 58 PERSONNEL PSYCHOL. 1009, 1016 (2005) (same).

64. See Uniform Guidelines on Employee Selection Procedure, 29 C.F.R. § 1607.14 (2010).

*Trust*⁶⁵ that rigorous statistical validation of selection methods is not required in every instance, but did not elaborate further.⁶⁶ This lack of guidance means that the courts vary widely in the standards they apply and retain broad discretion in deciding what kind of evidence satisfies the business necessity defense.⁶⁷ Some courts show a deferential attitude towards defendants' practices, whereas 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.⁶⁸ Some courts accept relaxed forms of content validation, resting on appeals to common sense,⁶⁹ or to the finding of a "manifest relationship" between pre-employment criteria and

65. 487 U.S. 977 (1988).

66. *Id.* at 998 (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 Sujata S. Menjoge, Comment, *Testing the Limits of Anti-Discrimination Law: How Employers' Use of Pre-Employment Psychological and Personality Tests Can Circumvent Title VII and the ADA*, 82 N.C. L. REV. 326, 360 (2003) ("[D]espite the recommendation by the EEOC that defendants use validation studies to determine whether a test is job-related, defendants do not need to provide any formal validation study that psychological or personality criteria predict actual on-the-job performance." (citing 29 C.F.R. § 1607.5 (2010))).

67. See *infra* notes 69-70.

68. See *supra* note 8 and accompanying text.

69. Compare *Douglas v. Hampton*, 512 F.2d 976, 984-85 (D.C. Cir. 1975) (demanding demonstration of empirical validity in black college graduates' challenge to the Federal Service Entrance Examination (FSEE) by "identifying criteria that indicate successful job performance and then showing a correlation between test scores and those criteria"), with *Stender v. Lucky Stores, Inc.*, 803 F. Supp. 259, 321-22 (N.D. Cal. 1992) (interpreting Civil Rights Act of 1991 to require an employer to "show that its selection criteria bear 'a manifest relationship to the employment in question'" (quoting *Griggs v. Duke Power Co.*, 401 U.S. 424, 432 (1971))).

successful job performance,⁷⁰ whereas others demand rigorous statistical data.⁷¹

Another area of uncertainty is the strength of the relationship that must be demonstrated in the validation process. It is unclear, for example, how predictive a hiring or promotion criterion must be to survive scrutiny. The law does not specify a number, and the courts have provided no guidance on the required magnitude of the link—as expressed, for example, in a correlation coefficient—between a job screen and subsequent workplace performance. 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 method. Relatedly, the law is unclear on whether employers are free to create as competitive a process for staffing the workplace as the

70. See *Ass'n of Mex.-Am. 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, 1312, 1315 (10th Cir. 1999) (using manifest relationship test); *NAACP v. Town of E. Haven*, 70 F.3d 219, 225 (2d Cir. 1995) (mandating manifest relationship between employment and job screen); *Zamlen v. City of Cleveland*, 906 F.2d 209, 216-17 (6th Cir. 1990) (following standard that uses manifest relationship test); *Davis v. City of Dall.*, 777 F.2d 205, 211 (5th Cir. 1985) (applying manifest relationship standard); *Robinson v. Lorillard Corp.*, 444 F.2d 791, 798-99 (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 closer relationship between criterion and job, although stopping short of demanding actual statistical validation. See, e.g., *Bew v. City of Chi.*, 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, 35-36 (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 A. Sullivan, *Disparate Impact: Looking Past the Desert Palace Mirage*, 47 WM. & MARY L. REV. 911, 994 (2005) (citing Linda Hamilton Krieger, *The Content of Our Categories: A Cognitive Bias Approach to Discrimination and Equal Employment Opportunity*, 47 STAN. L. REV. 1161, 1237 (1995)); see also *id.* (recognizing that formal validation is often impracticable but also that formal validation is not required "across the spectrum of disparate impact cases").

71. *Douglas*, 512 F.2d at 984-85.

market will bear, or whether they must give equal consideration to candidates who have demonstrated a minimum level of competence.⁷² Finally, some courts have compounded the uncertainty by adopting a more exacting approach based on language in some Supreme Court opinions, as ambiguously codified in the Civil Rights Act of 1991, 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.⁷³ Because satisfying this requirement 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 gives the courts considerable leeway to decide how strictly to apply the job-relatedness or business necessity defense.⁷⁴ It also has generated consider-

72. See, e.g., Samuel Issacharoff & Erin Scharff, *Antidiscrimination in Employment: The Simple, the Complex, and the Paradoxical* 6 (NYU Ctr. for Law, Econ. & Org., Working Paper No. 10-10, 2010). These 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.” *Id.* However, there is no support in the statute or the case law for this “minimalist” view of disparate impact’s requirements. In fact, Title VII permits employers “to give and to 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. Carle, *supra* note 43, at 285 n.208 (describing the Title VII language permitting the use of ability tests).

73. See Civil Rights Act of 1991 § 105, 42 U.S.C. § 2000e-2(k)(1)(A) (2006) (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 Civil Rights Act of 1991 suggesting that the plaintiff must demonstrate the existence of the alternative selection method, the courts have been divided on who must make this showing. *Compare* Int’l Bhd. of Elec. Workers v. Miss. Power & Light Co., 442 F.3d 313, 317 (5th Cir. 2006) (placing the burden on plaintiff), with *Bradley v. Pizzaco of Neb.*, 7 F.3d 795, 798 (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” (citation omitted)). See generally Harris & West-Faulcon, *supra* note 41, at 160-61 (discussing the “less discriminatory alternative” requirement).

74. See Rutherglen, *supra* note 43, at 106-07 (noting ambiguities surrounding the business necessity defense and stating that “[t]he particular language adopted by Congress [in the Civil Rights Act of 1991] 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” addressing 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”).

able confusion among potential parties and the judiciary, which is reflected in the cases challenging job testing in general and civil service tests for public servants in particular.⁷⁵ In addition, the technical nature of the validation process requires parties to present testimony from specialists in the IOP field, which often issues in an expensive and protracted “battle of the experts.”⁷⁶ 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.⁷⁷ That assumption finds expression in the subsequent development and wide acceptance of the four-fifths rule for establishing a prima facie case of unlawful disparate impact. That rule effectively embodies the expectation

75. 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 for the firefighter supervisors’ test at issue. See Rutherglen, *supra* note 43, at 107 (“On the spectrum between heavier and lighter burdens of justification, the [*Ricci*] 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. This evidence indicated that the exam was carefully geared to assessing the knowledge and skills needed to perform a firefighter supervisor’s job. *Id.* at 87. The courts did not demand a statistical demonstration of the tests’ ability to predict superior performance. *Id.* at 89-91. In contrast, lower courts in other cases have faulted the paucity of statistically valid data concerning the predictive validity of firefighter tests. See, e.g., *United States v. City of New York*, 637 F. Supp. 77, 131-32 (E.D.N.Y. 2009); Heather MacDonald, *Fighting Fire with Quotas*, CITY J., Oct. 24, 2010, <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 *Lewis v. City of Chi.*, No. 98 C 5596, 2005 WL 693618 (N.D. Ill. Mar. 22, 2005) (rejecting firefighters’ exam on disparate impact grounds), *aff’d*, 130 S. Ct. 2191, 2196, 2200-01 (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 WM. & MARY L. REV. 197, 254 nn.233-34 (2010) (collecting cases).

76. See, e.g., *Ricci v. DeStefano*, 129 S. Ct. 2658, 2664, 2670 (2009) (discussing voluminous testimony for both sides concerning the content validity of the firefighters’ test in that case).

77. See Zedeck, *supra* note 50, at 11-12.

that the workforce will reflect the racial composition of the background working-age population.⁷⁸ 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.⁷⁹ And the Court's opinion expresses skepticism about whether those requirements would bear any relationship to performance in the relatively unskilled jobs at issue.⁸⁰ Although the case law is mixed, subsequent decisions, especially those involving less skilled positions or government jobs, have been skeptical of screening tests that measure or draw heavily on general cognitive ability.

As the discussion below shows, however, subsequent work in social science reveals that the assumptions underlying *Griggs* lack support. First, IOP research, as well as work in labor economics and educational demography, indicates that blacks, and to a lesser extent Hispanics, currently lag behind whites in the skill sets that predict proficiency in a broad range of jobs.⁸¹ Direct measures of job performance ratings also reveal that, on average, these groups perform less well than whites at work.⁸² 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—remain the best predictors of performance for jobs at all levels of complexity.⁸³ The *g*-dependency of job selection methods both contributes to their usefulness and accounts for their adverse impact. As the subsequent discussion explains, these realities are inconsistent with the key understandings articulated in *Griggs* and carried forward in subsequent cases. The tension between the evidence and the founding myths of *Griggs* creates practical problems for disparate impact litigation and argues for a substantial revision in the rules for disparate impact cases.

78. Bobko & Roth, *supra* note 52, at 30-32.

79. *Griggs v. Duke Power Co.*, 401 U.S. 424, 431-32 (1971).

80. *Id.* at 431-33.

81. *See infra* Part II.

82. *See infra* Part II.

83. *See infra* note 97 and accompanying text.

II. EVIDENCE ON DISPARATE IMPACT: INDUSTRIAL AND ORGANIZATIONAL PSYCHOLOGY (IOP) RESEARCH

In confronting the challenges of selecting and managing a workforce without running afoul of the disparate impact rule, employers have enlisted the assistance of experts trained in IOP.⁸⁴ In the years since *Griggs*, an explosion of scholarly research has sought to address the problems employers face in seeking to comply with the mandates of disparate impact, including meeting the four-fifths target and establishing a job-relatedness or business necessity defense in the event of legal challenge.⁸⁵ Accordingly, research in this field has become increasingly focused on achieving two simultaneous goals: first, developing effective personnel practices that minimize adverse impact on racial minorities; and second, identifying valid job selection devices and establishing their link to job productivity for the purpose of satisfying the courts' definition of job-relatedness.

Basic premises of the IOP research literature are that the workplace should ideally function as a competitive meritocracy, and that this situation is consistent with disparate impact's requirements.⁸⁶ Thus, researchers proceed from the understanding that competitive staffing practices that help employers maximize employee productivity are legitimate and lawful. More specifically, job selection metrics that are both valid and unbiased—in being equally predictive of productivity for persons from all groups—are assumed to best advance the meritocratic idea of equal opportunity embodied in the law of disparate impact.⁸⁷ In addition, experts recognize that employers have an interest in creating the most productive workforce possible by finding the best workers regardless of race or background. Finally, it is understood that staffing decisions are routinely made under conditions of scarcity. Because there are usually more job applicants than there are positions available, employers must be

84. See, e.g., Tippins, *supra* note 50, at 202-03.

85. See, e.g., Bobko & Roth, *supra* note 52, at 29-30; Zedeck, *supra* note 50, at 17.

86. See *supra* note 16 and accompanying text.

87. Keith Hattrup & Brandon G. Roberts, *What Are the Criteria for Adverse Impact?*, in ADVERSE IMPACT, *supra* note 50, at 161, 161-63.

selective and are in a position to pick and choose among available workers.⁸⁸

In helping businesses meet disparate impact requirements in light of these conventions, IOP experts and psychometricians have generated a large body of empirical research and statistical data concerning the validity and adverse impact of various personnel screening devices. In analyzing the implications of this research, IOP experts have generally taken the four-fifths rule as an important benchmark for triggering liability under disparate impact, although they also make use of conventional tests of statistical significance.⁸⁹ Much effort is devoted to correlating performance on job screens with supervisor assessments of workers in various positions, as reflected by different types of ratings and evaluations commonly performed in the workplace. Data has also accumulated on group differences in performance on a range of widely used screens and in various occupations.⁹⁰

In fact, psychometricians and demographers have long been interested in identifying methods for selecting good workers, and much research on predicting job performance 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 criteria.⁹¹ Establishing the predictive validity of selection devices thus rests on the ability to measure job performance accurately. Accordingly, IOP experts are preoccupied with investigating and developing sound methods for evaluating workers, with the ultimate aims of devising job screening

88. *Building Better Organizations: Industrial-Organizational Psychology in the Workplace*, SOC'Y FOR INDUS. & ORG. PSYCHOL., <http://www.siop.org/visibilitybrochure/memberbrochure.aspx> (last visited Oct. 31, 2011); see also Jerard F. Kehoe, *Cut Scores and Adverse Impact*, in *ADVERSE IMPACT*, *supra* note 50, at 289, 289-90.

89. See Bobko & Roth, *supra* note 52, at 41.

90. See *Maximizing Human Potential Within Organizations: Learning the Science Behind Talent Management*, SOC'Y FOR INDUS. & ORG. PSYCHOL., <http://www.siop.org/visibilitybrochure/visibility.aspx> (last visited Oct. 31, 2011).

91. See *Guidelines for Education and Training at the Masters Level in Industrial-Organizational Psychology*, SOC'Y FOR INDUS. & ORG. PSYCHOL., <http://www.siop.org/guidelines.aspx> (last visited Oct. 31, 2011).

instruments that best predict performance and quantifying the predictive validity of those methods.⁹²

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 positions. Depending on the nature and selectivity of the jobs at issue, employers rely on a range of procedures for recruiting applicants and choosing workers.⁹³ In developing a pool of eligible candidates, some managers use informal methods like word-of-mouth referrals and personal recommendations, whereas 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.⁹⁴ 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 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.⁹⁵ Recently,

92. *Id.*

93. For a review of job selection methods, see Walter C. Borman, Mary Ann Hanson & Jerry W. Hedge, *Personnel Selection*, 48 ANN. REV. PSYCHOL. 299, 299-301 (1997); John E. Hunter & Ronda F. Hunter, *Validity and Utility of Alternative Predictors of Job Performance*, 96 PSYCHOL. BULL. 72, 72-73, 96 (1984); Frank L. Schmidt & John E. Hunter, *The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 85 Years of Research Findings*, 124 PSYCHOL. BULL. 262, 262-63 (1998); see also Bobko & Roth, *supra* note 52, at 29-31.

94. See sources cited *supra* note 93.

95. On general ability, or intelligence, tests in personnel selection, see Kelman, *supra* note 63, at 1158-59; James L. Outtz, *The Role of Cognitive Ability Tests in Employment Selection*, 15 HUM. PERFORMANCE 161, 169-70 (2002). On personality measures, see Murray R. Barrick & Michael K. Mount, *The Big Five Personality Dimensions and Job Performance: A Meta-Analysis*, 44 PERSONNEL PSYCHOL. 1, 1-2 (1991); Murray R. Barrick, Michael K. Mount & 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. SELECTION & ASSESSMENT 9 (2001).

prompted partly by concerns about the adverse impact of conventional paper-and-pencil tests that draw 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 alternatives make use of task or job simulations, real-time problem-solving exercises, work sample or “in-box” evaluations, and exams administered orally or using audio-visual techniques.⁹⁶

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 job-specific parameters, a consensus has emerged that measures of general cognitive ability—also designated as *g* or IQ—are generally the best predictors of work performance for all types of positions, with correlations in the range of approximately .5 or more with measured outcomes.⁹⁷ This

On job screening interviews, see Jose M. Cortina, Nancy B. Goldstein, Stephanie C. Payne, H. Kristl Davison & Stephen W. Gilliland, *The Incremental Validity of Interview Scores over and Above Cognitive Ability and Conscientiousness Scores*, 53 PERSONNEL PSYCHOL. 325, 325-26 (2000); Allen I. Huffcut & Winfred Arthur, Jr., *Hunter and Hunter (1984) Revisited: Interview Validity for Entry-Level Jobs*, 79 J. APPLIED PSYCHOL. 184, 189 (1994). On job knowledge and job skills tests, see IOP Brief, *supra* note 58, at 3-5.

96. On methods of job screening, including the use of so-called “assessment center” protocols to gauge an applicant’s ability to deal with situations commonly encountered on the job, see Paul R. Sackett & Filip Lievens, *Personnel Selection*, 59 ANN. REV. PSYCHOL. 419, 425-50 (2008). On job performance simulations and problem-solving exercises, see Barbara B. Gaugler, Douglas B. Rosenthal, George C. Thornton III & Cynthia Bentson, *Meta-Analysis of Assessment Center Validity*, 72 J. APPLIED PSYCHOL. 493, 493-94, 505 (1987). On work samples, see Roth, Bobko & McFarland, *supra* note 63, at 1009-10. For a review of alternatives to paper-and-pencil tests of ability and job knowledge, see Neal Schmitt, Catherine Clause & Elaine Pulakos, *Subgroup Differences Associated with Different Measures of Some Common Job-Relevant Constructs*, 11 INT’L REV. INDUS. & ORG. PSYCHOL. 115, 116-17 (1996).

97. See, e.g., Nathan R. Kuncel & Sarah A. Hezlett, *Fact and Fiction in Cognitive Ability Testing for Admission and Hiring Decisions*, 19 CURRENT DIRECTIONS PSYCHOL. SCI. 339, 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 Harold W. Goldstein, Charles A. Scherbaum & Kenneth P. Yusko, *Revisiting g: Intelligence, Adverse Impact, and Personnel Selection*, in ADVERSE IMPACT, *supra* note 50, at 95, 100 (“An additional central point of the psychometric perspective is that *g* is the most important quality that determines success of all types, including at work.”); *id.* at 116 (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 tests are the single best predictor of

observation, which comes out of decades of research and thousands of studies, is now widely accepted among IOP experts and, indeed, is a basic, shared premise of the field. The predictive value of cognitive performance, although perhaps not intuitively obvious and in tension with some popular ideas,⁹⁸ is grounded in the character and variety of tasks workers are asked to perform within our modern, complex economy. General intelligence has been observed to correlate with a host 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.⁹⁹ Thus, to the extent that *Griggs v. Duke Power Co.* suggests that cognitive ability 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,

job performance and thus should be afforded special status in the area of personnel selection”); Hunter & Hunter, *supra* note 93, at 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”); John E. Hunter & Frank L. Schmidt, *Intelligence and Job Performance: Economic and Social Implications*, 2 PSYCHOL. PUB. POL’Y & L. 447, 447 (1996) (expounding on the unmatched predictive power of IQ); Kuncel & Hezlett, *supra*, at 341 (noting the substantial correlation between cognitive ability and performance for jobs of high, medium, and low complexity); Patrick F. McKay, *Perspectives on Adverse Impact in Work Performance: What We Know and What We Could Learn More About*, in ADVERSE IMPACT, *supra* note 50, at 249, 253 (“Cognitive ability is the strongest single predictor of job performance.”); Outtz, *supra* note 95, at 161 (noting correlation of about .5 between cognitive ability and job productivity); James L. Outtz & Daniel A. Newman, *A Theory of Adverse Impact*, in ADVERSE IMPACT, *supra* note 50, at 53, 68 (“A great deal of empirical evidence has been amassed to support the correlation between cognitive test scores and job performance measures.”); Malcolm James Ree & James A. Earles, *Intelligence Is the Best Predictor of Job Performance*, 1 CURRENT DIRECTIONS PSYCHOL. SCI. 86, 87-88 (1992) (discussing data showing that IQ is the best predictor of performance, based on datasets on many thousands of military recruits); Schmidt & Hunter, *supra* note 93, at 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] is stronger than for any other personnel measure”). See generally Sackett & Lievens, *supra* note 96, at 440-41 (reviewing developments in personnel selection methods as confirming the superior predictive power of IQ); Robert L. Thorndike, *The Central Role of General Ability in Prediction*, 20 MULTIVARIATE BEHAV. RES. 241, 253 (1985) (“In the context of practical prediction ‘g’ appears to be alive and well.”); *infra* note 133 and accompanying text (comparing intelligence tests with other job performance predictors).

98. See, e.g., *infra* note 139 and accompanying text (discussing the construct fallacy); see also discussion of misconceptions about alternative predictors, including efforts to find filters that match or surpass cognitive performance, *infra* Part II.B.

99. Goldstein, Scherbaum & Yusko, *supra* note 97, at 97.

because the number of years of education achieved is closely correlated with intelligence.¹⁰⁰ In fact, people with higher IQs (and also more education) will, on average, perform better in any job than people with lower measured intelligence or less schooling.

A related observation is that job screens that depend heavily on intellectual ability tend to predict subsequent job performance better than metrics that place less emphasis on IQ or give greater weight to other characteristics. Devices such as integrity tests or personality measures, which are relatively uncorrelated with *g*, are significantly less predictive of outcomes than cognitive ability tests. For example, measures of conscientiousness—the personality trait with the strongest link to job success—show a correlation with job performance in the range of .18 to .37, in contrast to the correlation of .5 or more for IQ.¹⁰¹ 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, is 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 to learn, and job knowledge is a function of learning through studying or through job experience. Likewise, scores on

100. See, e.g., David Rowe, Wendy Vesterdal & Joseph Rodgers, *Herrnstein's Syllogism: Genetic and Shared Environmental Influences on IQ, Education, and Income*, 26 INTELLIGENCE 405, 405 (1998) (noting a correlation of .63 between years of education and IQ in a large sample of subjects from the National Longitudinal Survey of Youth).

101. See Barrick & Mount, *supra* note 95, at 1 (discussing conscientiousness as a performance predictor); Barrick, Mount & Judge, *supra* note 95, at 9-11 (same); Cortina, Goldstein, Payne, Davison & Gilliland, *supra* note 95, at 340-43 (same); Timothy A. Judge, Chad A. Higgins, Carl J. Thoresen & Murray R. Barrick, *The Big Five Personality Traits, General Mental Ability, and Career Success Across the Life Span*, 52 PERSONNEL PSYCHOL. 621, 640 (1999) (acknowledging that conscientiousness is a less powerful predictor of job performance than cognitive measures of mental ability); Deniz S. Ones, Chockalingam Viswesvaran & Frank L. Schmidt, *Comprehensive Meta-Analysis of Integrity Test Validities: Findings and Implications for Personnel Selection and Theories of Job Performance*, 78 J. APPLIED PSYCHOL. 679, 696-97 (1993) (reporting on the predictive validity of integrity tests); Daniel J. Ozer & Veronica Benet-Martinez, *Personality and the Prediction of Consequential Outcomes*, 57 ANN. REV. PSYCHOL. 401, 411-12 (2006) (finding conscientiousness to be moderately predictive of work performance); Paul R. Sackett & James E. Wanek, *New Developments in the Use of Measures of Honesty, Integrity, Conscientiousness, Dependability, Trustworthiness, and Reliability for Personnel Selection*, 49 PERSONNEL PSYCHOL. 787, 787-90 (1996) (discussing relationship between conscientiousness and work performance).

interviews are sensitive to verbal skill and analytic acumen, which have an established correlation with general mental ability. Nonetheless, none of these methods has proven as reliable as pure measures of general intelligence, and none predicts job success as well as IQ.¹⁰²

*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 comply with disparate impact's requirements. Unfortunately, the goal of maximizing workforce productivity while reducing or eliminating disparate impact has proven elusive. Indeed, the literature consistently demonstrates that, given current realities, the most effective job selection methods available show a substantial adverse impact on non-Asian minorities, and in particular tend to screen out blacks. Moreover, the higher the predictive validity of the metric 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.¹⁰³ There is now a consensus in the field that, over a wide

102. For example, the correlation between structured job interviews and subsequent job performance ratings has been estimated as around .36. Michael A. Campion, James E. Campion & J. Peter Hudson, Jr., *Structured Interviewing: A Note on Incremental and Alternative Question Types*, 79 J. APPLIED PSYCHOL. 998, 1000 (1994). The predictive correlation of work performance with other methods, including situational assessments, tests of job knowledge, and performance of sample job tasks, is similar in magnitude. For a survey of methods and correlations, see Borman, Hanson & Hedge, *supra* note 93, at 314-17; Calvin C. Hoffman & George C. Thornton III, *Examining Selection Utility Where Competing Predictors Differ in Adverse Impact*, 50 PERSONNEL PSYCHOL. 455, 456-57 (1997); Sackett & Lievens, *supra* note 96, at 431; *see also* Schmitt, Clause & Pulakos, *supra* note 96, at 134 (discussing validities for various job selection methods and the need for further research).

103. *See, e.g.*, Wilfried De Corte, *Weighing Job Performance Predictors to Both Maximize the Quality of the Selected Workforce and Control the Level of Adverse Impact*, 84 J. APPLIED PSYCHOL. 695, 700 (1999) (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"); Hatrup & Roberts, *supra* note 87, at 161-65 (describing the validity-diversity tradeoff as a pervasive empirical feature of IOP research and practice); Keith M. Pyburn, Jr., Robert E. Ployhart & David A. Kravitz, *The Diversity-Validity Dilemma: Overview and Legal Context*, 61 PERSONNEL PSYCHOL. 143, 145 (2008) ("Unfortunately, many of the most predictive [knowledge, skill, and ability measures] and predictor methods (for

range of real-world conditions, this tradeoff is an unavoidable feature of personnel practice.¹⁰⁴

The difficulty of finding selection methods that predict job success while avoiding racially adverse impact is a product of two established social science facts. First, as already 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 selection methods have a significant correlation with cognitive ability.

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.¹⁰⁵ This gap has

example, assessment centers) produce varying degrees of mean *subgroup differences*, with racio-ethnic minority groups usually scoring lower than majority groups.”); *see also* Schmitt, Clause & Pulakos, *supra* note 96, at 134 (describing the pervasive problem in personnel practice of “[t]he conflict between equal representation of different groups and the achievement of maximum expected productivity”).

104. *See, e.g.*, Paul R. Sackett, Neal Schmitt, Jill E. Ellingson & Melissa B. Kabin, *High-Stakes Testing in Employment, Credentialing, and Higher Education: Prospects in a Post-Affirmative Action World*, 56 AM. PSYCHOLOGIST 302, 302 (2001) (summarizing conclusions from a review of the IOP literature and noting that “it is unreasonable to expect that one can maximize both the performance and ethnic diversity of selected individuals”); *id.* at 303 (“What quickly becomes clear [from the research] is that these two values—performance and diversity—come into conflict. Increasing emphasis on the use of tests in the interest of gaining enhanced performance has predictable negative consequences for the selection of Blacks and Hispanics.”); *id.* at 314 (noting the “tension between pursuing a validity-maximization strategy and a diversity-maximization strategy”); *see also id.* at 315 (noting that the observed “subgroup differences on cognitively loaded tests of knowledge, skill, ability, and achievement,” and the resulting validity-diversity tradeoff, “simply document persistent [background] inequities”).

105. *See, e.g.*, Leaetta M. Hough, Frederick L. Oswald & Robert E. Ployhart, *Determinants, Detection and Amelioration of Adverse Impact in Personnel Selection Procedures: Issues, Evidence and Lessons Learned*, 9 INT’L J. SELECTION & ASSESSMENT 152, 153 (2001) (“Regarding general intelligence, the commonly accepted mean difference between Blacks and Whites is about one standard deviation, with Blacks scoring lower than Whites.”); Kelman, *supra* note 63, at 1158 (“As a group, blacks score significantly lower on ‘general ability’ tests than do whites.”); Philip L. Roth, Craig A. Bevier, Philip Bobko, Fred S. Switzer III & Peggy Tyler, *Ethnic Group Differences in Cognitive Ability in Employment and Educational Settings: A Meta-Analysis*, 54 PERSONNEL PSYCHOL. 318 (2001) (describing data showing that Hispanics

important consequences for personnel practice and policy. The combination of 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.¹⁰⁶

In the wake of the *Griggs* decision, cognitive ability tests have been targeted for criticism and legal challenge based on their adverse impact on black job applicants. Although “pure” tests of ability survive in some quarters,¹⁰⁷ their use has diminished significantly.¹⁰⁸ Nonetheless, minimizing or eliminating reliance on general ability tests has not solved the adverse impact problem. Group disparities in other criteria, such as educational biodata, outcomes of job interviews, or scores on tests of job skills and

lag behind whites by about 0.72 standard deviations on standard cognitive tests, as compared to 1 standard deviation for blacks); Paul R. Sackett & Steffanie L. Wilk, *Within-Group Norming and Other Forms of Score Adjustments in Preemployment Testing*, 49 AM. PSYCHOLOGIST 929, 929 (1994) (“Black-White differences of approximately 1.0 standard deviation units and Hispanic-White differences of approximately 0.6-0.8 standard deviation units have been widely and consistently reported for measures of cognitive ability.”); Sackett, Schmitt, Ellingson & Kabin, *supra* note 104, at 302 (noting that “test score distributions consistently reveal significant mean differences” by race, and that “Blacks tend to score approximately one standard deviation lower than Whites, and Hispanics score approximately two thirds of a standard deviation lower than Whites”). For a discussion of black-white differences in IQ and standard deviation measurements, see Kelman, *supra* note 63, at 192 n.91. See also Goldstein, Scherbaum & Yusko, *supra* note 97, at 120 (discussing majority-minority differences in performance on tests of general intelligence).

106. See Sackett, Schmitt, Ellingson & Kabin, *supra* note 104, at 302 (noting that mean differences in test scores “can translate into [a] large adverse impact against protected groups when test scores are used in selection and credentialing decision-making”).

107. See WAYNE L. FOX, JOHN E. TAYLOR & JOHN S. CAYLOR, GEORGE WASHINGTON UNIV. HUMAN RES. RESEARCH OFFICE, APTITUDE LEVEL AND THE ACQUISITION OF SKILLS AND KNOWLEDGES IN A VARIETY OF MILITARY TRAINING TASKS vi (1969); Philip Crewson, *A Comparative Analysis of Public and Private Sector Entrant Quality*, 39 AM. J. POL. SCI. 628, 633 (1995) (describing the Armed Forces 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”); Malcolm James Ree & Thomas R. Carretta, *Central Role of g in Military Pilot Selection*, 6 INT’L J. AVIATION PSYCHOL. 111, 111-13 (1996). For more on armed forces screening tests, their reliance on measures of cognitive ability, and their racially disparate impact, see *infra* note 256.

108. See *Douglas v. Hampton*, 512 F.2d 976, 979 (D.C. Cir. 1975) (discussing black college graduates’ challenge to the general ability Federal Service Entrance Examination (FSEE)); Hunter & Schmidt, *supra* note 97, at 466 (1996) (acknowledging a decline in the use of pure intelligence tests for job screening); Kelman, *supra* note 63, at 1204-05 (discussing the General Ability Test Battery (GATB), used for many years by the U.S. Employment Service to screen and rank prospective federal government workers but now defunct).

knowledge, although somewhat smaller than for standard tests of intelligence or IQ, are still fairly substantial—and substantial enough routinely to run afoul of disparate impact’s threshold guidelines.

In describing observed patterns, the IOP literature refers to the standardized difference between two comparison groups (designated as “ d ”) associated with a given predictor of job performance.¹⁰⁹ Measured d values of commonly used 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 standard deviation commonly reported for tests of cognitive ability, to significantly lower to negligible racial differences for some kinds of personality tests. Most d values fall somewhere in between, with more heavily g -loaded, or ability-dependent, measures showing greater group differences.¹¹⁰

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, significant

109. Denise Potosky, Philip Bobko & Philip Roth, *Forming Composites of Cognitive Ability and Alternative Measures To Predict Job Performance and Reduce Adverse Impact: Corrected Estimates and Realistic Expectations*, 13 INT’L J. SELECTION & ASSESSMENT 304, 305 (2005) (“The d statistic is computed by subtracting the mean of the focal minority group from the mean of the majority group in the numerator. The denominator is the sample-weighted average standard deviation of the minority and majority groups. For example, a d of 0.5 indicates that the majority group scored, on average, one half of an averaged standard deviation higher than the minority group.”).

110. For a review of the racially adverse impact of commonly used screening methods, see Schmitt, Clause & Pulakos, *supra* note 96, at 118-28 (reporting a range of values for blacks and hispanics); *see also* Pyburn, Ployhart & Kravitz, *supra* note 103, at 143-47; Roth, Bevier, Bobko, Switzer & Tyler, *supra* note 105, at 297-99. On the disparate impact of structured interviews, see Allen I. Huffcutt & Philip L. Roth, *Racial Group Differences in Employment Interview Evaluations*, 83 J. APPLIED PSYCHOL. 179, 179-81 (1998); Philip L. Roth, Chad H. Van Iddekinge, Allen I. Huffcutt, Carl E. Eidson Jr. & Philip Bobko, *Corrections for Range Restriction in Structured Interview Ethnic Group Differences: The Values May Be Larger Than Researchers Thought*, 87 J. APPLIED PSYCHOL. 369, 369-72 (2002). On racial differences in problem solving exercises and job simulations, see Michelle A. Dean, Philip L. Roth & Philip Bobko, *Ethnic and Gender Subgroup Differences in Assessment Center Ratings: A Meta-Analysis*, 93 J. APPLIED PSYCHOL. 685, 685-86, 688-89 (2008) (from a meta-analysis of assessment center data, noting d values of 0.52 for blacks and 0.28 for Hispanics, as compared to whites, which are smaller than cognitive ability disparities but sufficient to trigger disparate impact scrutiny in many cases).

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 (NAEP), the average black twelfth-grade student reads at an eighth-grade level.¹¹¹ Blacks are consistently observed to enter elementary school, high school, college, graduate school, and professional education with lower test scores, grades, measures of learning and achievement, and proficiency levels than whites.¹¹² Thus, blacks and whites with an equivalent amount of education and training and similar years of schooling show marked performance disparities, with blacks on average lagging systematically behind.¹¹³

The differences in ability and achievement have concrete consequences for how well people perform in the workplace.¹¹⁴ IOP research reveals that the achievement gap does not stop at the schoolhouse door. Although there is disagreement on the precise

111. Eric Hanushek, *How Well Do We Understand Achievement Gaps?*, FOCUS, Winter 2010, at 5.

112. See *id.* at 6-7 (“[P]erformance appears roughly flat for almost four decades.... The analysis vividly underscored the huge difference in the achievement of students by race and background.... The magnitude of the gap is stunning.”); see also THOMAS J. ESPENSHADE & ALEXANDRIA WALTON RADFORD, NO LONGER SEPARATE, NOT YET EQUAL: RACE AND CLASS IN ELITE COLLEGE ADMISSIONS AND CAMPUS LIFE 226-29 (2009) (providing data on group differences in scores and academic achievement); AMY L. WAX, RACE, WRONGS AND REMEDIES: GROUP JUSTICE IN THE 21ST CENTURY 47-52 (2009) (discussing racial gaps in educational performance). For a review of scholarship on this issue, see Richard C. Hunter & RoSusan Bartee, *The Achievement Gap: Issues of Competition, Class, and Race*, 35 EDUC. & URB. SOC’Y 151, 154-56 (2003).

113. For example, a recent report on college readiness among high school graduates in New York State reveals that 51 percent of whites and 56 percent of Asian Americans completing high school statewide are “college ready,” whereas only 13 percent of black students and 15 percent of Hispanic students were deemed prepared for college level courses after graduation. See Sharon Otterman, *37 Percent of New York Graduates Are College-Ready, Data Show*, N.Y. TIMES, June 15, 2011, at A23. The psychologist Richard Nisbett reports on a study showing that, among a cohort of new high school graduates who eventually completed college, blacks scored 1 standard deviation below whites on an IQ test. RICHARD E. NISBETT, INTELLIGENCE AND HOW TO GET IT 147 (2009). When the students were tested again at the end of college, the measured gap was 0.4 standard deviation. *Id.*

114. See Hanushek, *supra* note 111, at 6 (“[S]tudies 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.” (footnote omitted)).

magnitude of the disparities, with variation based on methodology, type of position, job selectivity, and other parameters,¹¹⁵ 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 0.3, or a bit less than a third of a standard deviation across a spectrum of jobs,¹¹⁶ with one meta-analysis estimating the gap as between 0.24 and 0.39 standard deviations.¹¹⁷ This gap would appear larger than expected, given that most employers try to select workers with similar qualifications, including job-relevant background and level of education.¹¹⁸ Indeed, years of education and degrees completed have emerged as important and widely used personnel filters. Yet racial gaps in job performance persist. These results reflect documented differences in actual academic achievement, which are independent of years of schooling and degrees completed. Because many jobs draw on skills learned in school, it should not be surprising that measures of learning and academic mastery, such as grades and test scores, are predictive of job success.¹¹⁹ The harsh reality is that more capable

115. See Patrick F. McKay & Michael A. McDaniel, *A Reexamination of Black-White Mean Differences in Work Performance: More Data, More Moderators*, 91 J. APPLIED PSYCHOL. 538, 548-50 (2006); see also McKay, *supra* note 97, at 251-52.

116. See, e.g., Outtz & Newman, *supra* note 97, at 77 (noting “average race differences around $d=0.3$ for a variety of job performance measures”); see also Frank Landy, *Performance Ratings: Then and Now*, in ADVERSE IMPACT, *supra* note 50, at 227, 230-35 (summarizing range of studies of racial differences in job performance); Philip L. Roth, Allen I. Huffcutt & Philip Bobko, *Ethnic Group Differences in Measures of Job Performance: A New Meta-Analysis*, 88 J. APPLIED PSYCHOL. 694, 702 (2003) (noting that blacks lag behind whites in measured job performance ratings); Paul R. Sackett, Matthew J. Borneman & Brian S. Connelly, *High-Stakes Testing in Higher Education and Employment*, 63 AM. PSYCHOLOGIST 215, 223 (2008); Sackett & Lievens, *supra* note 96, at 435; Sackett & Wilk, *supra* note 105, at 934 (noting reported black-white differences in job performance as approximately 0.3-0.4 standard deviations, with values perhaps higher when corrected for measurement error). Most data focuses on black-white differences, although evidence on Hispanics has also been collected. See, e.g., Sackett & Wilk, *supra* note 105, at 943-44; Schmitt, Clause & Pulakos, *supra* note 96, at 121-22.

117. See McKay & McDaniel, *supra* note 115, at 538; see also McKay, *supra* note 97, at 251.

118. See *supra* Part II.A (discussing biodata as an important screening device).

119. See, e.g., Philip L. Roth & Philip Bobko, *College Grade Point Average as a Personnel Selection Device: Ethnic Group Differences and Potential Adverse Impact*, 85 J. APPLIED PSYCHOL. 399, 399 (2000) (discussing predictive power of grades and school achievement for job performance); Sackett, Borneman & Connelly, *supra* note 116, at 215 (noting that tests of developed intellectual ability are valid predictors of academic as well as occupational performance).

people tend to perform better at work, and that minorities—most notably blacks and to a lesser degree Hispanics—know less and possess fewer skills, on average, than whites and Asians across a number of domains that reflect modern workplace demands. Those shortfalls in turn influence ratings on criteria used to determine hiring and promotion. In other words, the reason that valid job predictors—which are factors correlated to some extent with observed job outcomes—tend to show a racially disparate impact is that there are real group differences in what is being predicted: actual ability to do the job.

An analysis of the sources of these patterns is beyond the scope of this Article. As one commentator notes, “there are many realistic disadvantages that distinguish racial subgroups, and these disadvantages logically have some implications for job performance.”¹²⁰ 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.¹²¹

In sum, the tendency of valid job selection methods to screen out minorities is traceable to underlying group differences that correlate with ability to function in the workplace. Regardless of the sources of these differences, the validity-diversity tradeoff reflects an unfortunate reality with predictable consequences for personnel selection—consequences that are hard or impossible to avoid. Although “[p]ersonnel psychologists ... have devoted considerable effort to identifying test and test presentation strategies that reduce

120. Outtz & Newman, *supra* note 97, at 84-85; *see also* Sackett, Schmitt, Ellingson & Kabin, *supra* note 104, at 315 (noting that subgroup differences in performance “simply document persisting inequities”).

121. *See, e.g.*, George Farkas, *Cognitive Skills and Noncognitive Traits and Behaviors in Stratification Processes*, 29 ANN. REV. SOC. 541, 544-47 (2003) (noting the contribution of differential family investments, economic circumstances, cultural style, and other factors to racial gaps in human capital development); Amy L. Wax, *The Discriminating Mind: Define It, Prove It*, 40 CONN. L. REV. 979, 998-1002 (2008) (noting socio-demographic differences between blacks and whites that could account for disparities in skill development and thus lead to average group differences in occupational success); *see also* Reva B. Siegel, *From Colorblindness to Antibalkanization: An Emerging Ground of Decision in Race Equality Cases*, 120 YALE L.J. 1278, 1320-21 & nn.128-29 (2011) (“[D]isparate impact’s critics are inclined to interpret the underrepresentation of minorities as evidence of racial group differences in taste or aptitude.”).

adverse impact,” they have found that “few strategies have eliminated adverse impact.”¹²²

That has not prevented IOP experts from trying, however. Efforts to circumvent the validity-diversity tradeoff have spawned a voluminous literature. In attempting to develop novel screens with a smaller disparate impact, researchers have pursued a number of approaches and relied on various observations in the field. The evidence suggests, for example, that supervisors value both “task performance,” which relates to carrying out core work requirements, and “contextual performance,” or “organizational citizenship,” which depends on cooperative and pro-social behaviors on the job.¹²³ 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 nonexistent.¹²⁴ In addition, minorities are known to lag significantly behind whites in scores on conventional paper-and-pencil tests that draw heavily on reading and written communication skills.¹²⁵ Based on these observations, IOP experts have worked to formulate sophisticated, multi-step “composite” assessments that place greater weight on personality attributes

122. Zedeck, *supra* note 50, at 22.

123. See, e.g., Philip Bobko, Philip L. Roth & Denise Potosky, *Derivation and Implications of a Meta-Analytic Matrix Incorporating Cognitive Ability, Alternate Predictors, and Job Performance*, 52 PERSONNEL PSYCHOL. 561, 562 (1999) (noting the role of “contextual performance factors” or “organizational citizenship behaviors” in worker assessments); Keith Hattrup, Joanna Rock & Christine Scalia, *The Effects of Varying Conceptualizations of Job Performance on Adverse Impact, Minority Hiring and Predicted Performance*, 82 J. APPLIED PSYCHOL. 656, 657 (1997) (discussing dimensions of measured job performance and distinguishing between “task performance”—going to the “technical core” of the job—and “contextual performance”—which reflects support of the organization’s “climate and culture” and the display of “helping, prosocial, and citizenship behaviors”); Kevin R. Murphy, *How a Broader Definition of the Criterion Domain Changes Our Thinking About Adverse Impact*, in ADVERSE IMPACT, *supra* note 50, at 137, 141 (“[T]he 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.”).

124. See McKay & McDaniel, *supra* note 115, at 540 (noting larger racial differences for task performance, which is more ability based, than for contextual performance, which is dependent on extra-role and prosocial behaviors); McKay, *supra* note 97, at 253-54 (“[C]riteria that are highly dependent on cognitive ability will exhibit larger black-white mean disparities than those more contingent on personality.”).

125. See *supra* note 105 and accompanying text.

like conscientiousness and integrity than on conventional measures of cognitive acumen. Alternatively, they have devised “tests that [are] more interactive, behaviorally-oriented, and orally- or aurally-oriented,”¹²⁶ such as simulations or real-time problem-solving exercises (as often employed at job assessment centers).¹²⁷

The goal of this research is to reduce disparate impact while maintaining the predictive power of screening instruments. Although the details of this quest are technical, the conclusions can be summarized succinctly. Despite extensive and strenuous efforts to circumvent the validity-diversity tradeoff, the project has failed. As a practical matter, adverse impact cannot readily be reduced without sacrificing either accuracy or predictive validity.¹²⁸ Novel methods that minimize reliance on written communication and attempt to simulate real work situations still show considerable

126. Schmitt, Clause & Pulakos, *supra* note 96, at 126.

127. For a comprehensive review of attempts to develop innovative job screening protocols that maintain validity while reducing adverse impact, see *id.* Job assessment centers have received a good deal of play as a possible 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., IOP Brief, *supra* note 58, at 28-33. The research shows that, although assessment center protocols can somewhat reduce race gaps, they do not eliminate adverse impact. See Winifred Arthur, Jr., Eric Anthony Day, Theresa L. McNelly & Pamela S. Edens, *A Meta-Analysis of the Criterion-Related Validity of Assessment Center Dimensions*, 56 PERSONNEL PSYCHOL. 125, 125 (2003); Kobi Dayan, Ronen Kasten & Shaul Fox, *Entry-Level Police Candidate Assessment Center: An Efficient Tool or a Hammer To Kill a Fly?*, 55 PERSONNEL PSYCHOL. 827, 827 (2002); Dean, Roth & Bobko, *supra* note 110, at 685; Gaugler, Rosenthal, Thornton & Bentson, *supra* note 96, at 503-04; Eran Hermlin, Filip Lievens & Ivan T. Robertson, *The Validity of Assessment Centres for the Prediction of Supervisory Performance Ratings: A Meta-Analysis*, 15 INT'L J. SELECTION & ASSESSMENT 405, 405 (2007); see also Tippins, *supra* note 50, at 218 (discussing the greater costs of using assessment centers, which are cumbersome and labor-intensive devices).

128. See, e.g., De Corte, *supra* note 103, at 700 (discussing the difficulty of developing new approaches to job selection that control the level of racially adverse impact without “neglect[ing] the goal of maximizing the quality of the selected workforce”); Paul R. Sackett & Lawrence Roth, *Multi-Stage Selection Strategies: A Monte Carlo Investigation of Effects on Performance and Minority Hiring*, 49 PERSONNEL PSYCHOL. 549, 567-71 (1996) (noting the potential for complex and multi-factorial selection practices to achieve small reductions in group differences without major sacrifices in validity, but only under rare and specialized conditions that are difficult to predict or identify systematically ahead of time); Neal Schmitt et al., *Adverse Impact and Predictive Efficiency of Various Predictor Combinations*, 82 J. APPLIED PSYCHOL. 719, 723 (1997) (analyzing alternative composite job predictors and 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”).

disparities by race.¹²⁹ Adding additional predictors or fiddling with their weight has not solved the dilemma, nor have job screens that de-emphasize cognitive skills and rely on traits like conscientiousness, agreeableness or integrity. Although those attributes correlate 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 instruments for measuring them are less reliable and precise.¹³⁰ Consequently, combining more cognitively loaded with less cognitively loaded metrics can modestly reduce disparate impact without compromising validity only in exceptional cases.¹³¹ More generally, a selection method that boosts minority representation will almost certainly reduce workplace productivity unless the

129. See, e.g., Sackett, Schmitt, Ellingson & Kabin, *supra* note 104, at 315 (noting that altering the mode of assessment, including adopting methods that “reduc[e] the verbal component (or reading level) of tests, may have a positive effect on subgroup differences, although *d* is often still large enough to produce adverse impact”).

130. See John M. Avis, Jeffrey D. Kudisch & Vincent J. Fortunato, *Examining the Incremental Validity and Adverse Impact of Cognitive Ability and Conscientiousness on Job Performance*, 17 J. BUS. & PSYCHOL. 87, 87 (2002) (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 significant adverse impact associated with that cognitive ability component); Hattrup, Rock & Scalia, *supra* note 123, at 657, 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”); Ones, Viswesvaran & Schmidt, *supra* note 101, at 694 (noting that integrity tests predict job performance with only “moderate” validity); Elaine D. Pulakos & Neal Schmitt, *An Evaluation of Two Strategies for Reducing Adverse Impact and Their Effects on Criterion-Related Validity*, 9 HUM. PERFORMANCE 241, 253, 255 (1996) (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 percent of applicants; also observing 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 & Filip Lievens, *Decision Aids for Addressing the Validity-Adverse Impact Trade-Off*, in ADVERSE IMPACT, *supra* note 50, at 453, 457-59 (noting the limited to null potential for composite and novel predictors combining low and high adverse impact instruments to reduce overall disparate impact and maintain predictive validity).

131. See Hattrup, Rock & Scalia, *supra* note 123, at 660, 662 (claiming to find a method that maintains validity while reducing adverse impact below the level that violates the four-fifths rule, but only for those rare specialized jobs, such as sales, that are significantly more dependent on citizenship-related factors than *g*-related factors, and only when selection is relatively uncompetitive, that is, when more than 80 percent of applicants are hired).

selection ratio is very high—that is, only if most or all applicants are hired.¹³²

In sum, it is difficult, if not impossible, to reduce reliance on intelligence-based screens or diminish 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 selection 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. And that workforce will be chosen at far greater expense, because proposed alternatives to *g*-loaded tests are notoriously cumbersome, time consuming, and expensive.¹³³ In sum, efforts to develop new personnel practices that circumvent the validity-diversity tradeoff have failed. For most commonly encountered situations, the tradeoff is unavoidable.

Although the research overwhelmingly supports this conclusion, the case law resists. 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 these methods are not the best available. Rather, there exist equally or more valid screens with less adverse impact that, according to the legal doctrine, employers must adopt in lieu of existing procedures.¹³⁴ The premise operating here is that the validity-diversity tradeoff can easily be overcome.

In the context of litigation over civil service jobs, the focus is often on the assertion that selection devices that de-emphasize or ignore

132. See, e.g., Sackett, De Corte & Lievens, *supra* note 130, at 455 tbl.17.1; Paul R. Sackett & Jill E. Ellingson, *The Effects of Forming Multi-Predictor Composites on Group Differences and Adverse Impact*, 50 PERSONNEL PSYCHOL. 707, 712 tbl.2 (1997); see also Schmitt, Clause & Pulakos, *supra* note 96, at 117 (reviewing results for alternative job screening methods and protocols).

133. See Sackett, Schmitt, Ellingson & Kabin, *supra* note 104, at 315 (“Complicating matters further, attempts to overcome issues associated with reliable measurement [of job-related skills] often result in a testing procedure that is cost-prohibitive when conducted on a large scale.”); see also Tippins, *supra* note 50, at 218 (noting that assessment center evaluations are cumbersome, costly, and labor-intensive).

134. See *supra* notes 73-74 and accompanying text.

key criteria—such as “command presence” in the case of firefighter supervisors—are inferior predictors of job success.¹³⁵ The use of written tests rather than simulations of actual task performance is thus seen as presumptively less valid and less able to identify the most effective workers. In this vein, respondents and their experts in *Ricci v. DeStefano* argued that equally effective but less discriminatory promotional methods were available to the city. These included job simulations and assessment center procedures that placed more emphasis on oral measures of leadership skills and “command presence.”¹³⁶ Nonetheless, no actual empirical support for this proposition was offered in any of the materials before the courts, including the Supreme Court amicus brief for the respondents in *Ricci* filed by five individual IOP specialists.¹³⁷ In particular, no specific studies or data were adduced 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.¹³⁸

The notion that the written test in *Ricci* was inadequate for its failure to measure key attributes of the firefighter’s job is based on a misguided “construct fallacy”: the notion that measures of “constructs” or traits peculiarly associated with specific types of work are the best predictors of occupational success. Although this notion has powerful intuitive appeal and is often asserted in discussions of disparate impact,¹³⁹ 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.¹⁴⁰ Cognitive ability is the most effective predictor of occupational success for a broad range of jobs, from most to least demanding.¹⁴¹ Moreover, the psychometric data indicate that this skill is most effectively measured through written tests of analysis and

135. See Harris & West-Faulcon, *supra* note 41, at 126.

136. See *id.* at 155; see also *Ricci v. DeStefano*, 129 S. Ct. 2658, 2695 (2009) (Ginsburg, J., dissenting); IOP Brief, *supra* note 58, at 28-33; Norton, *supra* note 75, at 252.

137. IOP Brief, *supra* note 58, at 28-33.

138. *Id.*

139. See, e.g., Lani Guinier & Susan Sturm, Op-Ed., *Trial by Firefighters*, N.Y. TIMES, July 11, 2009, at A19 (stating, without research citation, that paper-and-pencil tests are not good predictors of actual performance in public emergency service jobs).

140. See *supra* note 97 and accompanying text.

141. See *supra* notes 97, 102 and accompanying text.

learning.¹⁴² Accordingly, the notion that the best candidates for a fire captain position cannot be identified without gauging attributes like “command presence” or leadership skill is a product of wishful thinking unsupported by hard data or empirical research. No known studies demonstrate that evaluating leadership or “command presence” is 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 that adding leadership assessments to the mix or giving greater weight to noncognitive skills will decrease adverse impact to acceptable levels while increasing, or at least not sacrificing, validity. The assertion that 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 validity-diversity 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.¹⁴³ Indeed, existing research 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 firefighters,¹⁴⁴ generally correlate with performance at the level of about .25-.39,¹⁴⁵ which compares unfavorably with the .5-.6 validities associated with heavily *g*-loaded screens.¹⁴⁶ And although performance on assessment center exercises shows less adverse impact on minorities, group disparities are still pronounced.¹⁴⁷

142. See *supra* note 97 and accompanying text.

143. See *supra* note 97 and accompanying text; *supra* note 127.

144. See, e.g., *Ricci v. DeStefano*, 129 S. Ct. 2658, 2705 (2009) (Ginsburg, J., dissenting) (observing that 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 & West-Faulcon, *supra* note 41, at 155 n.290.

145. See Arthur, Day, McNelly & Edens, *supra* note 127, at 125.

146. See *supra* note 83 and accompanying text (discussing the validities of *g*-loaded screening mechanisms).

147. See *supra* note 127 and accompanying text; see also Dean, Roth & Bobko, *supra* note 110, at 688. The Industrial-Organizational Psychologists brief in *Ricci* filed by five individual experts asserts that, because New Haven’s firefighter promotional exam 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. See IOP Brief, *supra* note 58,

III. 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 validity-diversity tradeoff, mean that even modestly useful personnel selection criteria will screen out too many minorities most of the time. Achieving racial balance 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 employers to defend their 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, and which runs the risk of distorting their practices and incentives. Even absent that result, the pursuit of disparate impact claims will, at most, have only marginal effects on workplace diversity. If the courts do their job and apply the disparate impact doctrine correctly, almost all adverse racial impacts will be found to be justified and workplace demographics will remain virtually unchanged. The magnitude of existing group differences in job-related skills is large enough to account for the underrepresentation of minorities across the workforce. That reality, and not arbitrary employment hurdles, is the principal source of workplace imbalances and employers' failure to comply with the four-fifths rule.

The IOP literature yields crucial insights into the constraints under which businesses operate in managing their workforce while simultaneously striving to meet diversity targets. Given the magni-

at 28-33. 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 of paper-and-pencil civil service exam administered by the city in that case and adduces no evidence for the proposition that such alternatives select more effective supervisors. Rather, the experts as well as respondents, rely heavily on the fact that many other cities make use of the assessment center option. *Id.* at 30-31. It should be obvious that this fact alone fails to establish the superior validity of proposed alternatives.

tude 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 procedures except the least selective and most weakly predictive can be expected to fall short of satisfying the four-fifths rule.¹⁴⁸ Although the four-fifths standard 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 leading researchers in the field have observed, “[i]t is informative, though perhaps disheartening, to note that even d [group performance difference] values commonly viewed as small [for example 0.2] can produce violations of the four-fifths rule at a variety of commonly occurring [job] selection ratios.”¹⁴⁹ Because average d values frequently exceed 0.2, meritocratic staffing practices will routinely fall short of the four-fifths target.

This observation can be demonstrated through simple calculations from data readily available in the IOP literature. Paul Sackett and his colleagues have performed such calculations and assembled the results in a table that summarizes the relationship between relevant parameters.¹⁵⁰ The table in Figure 1A 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 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 of performance on a job screen, regardless of group identity.

148. See *supra* notes 127-28 and accompanying text; *supra* note 131.

149. Sackett & Ellingson, *supra* note 132, at 712.

150. See, e.g., Sackett, De Corte & Lievens, *supra* note 130, at 455 tbl.17.1; see also Sackett & Ellingson, *supra* note 132, at 712 tbl.2.

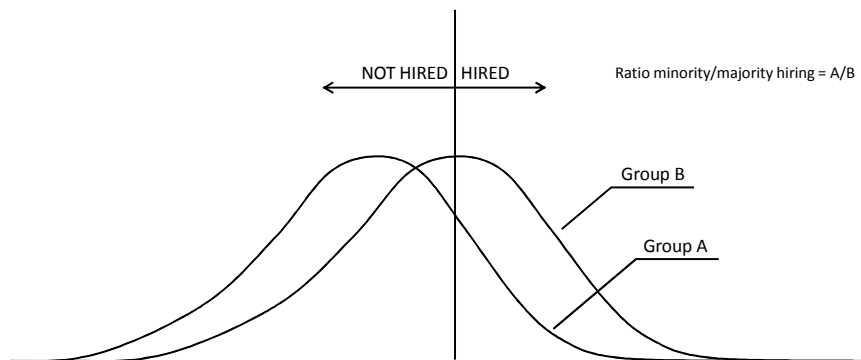
Figure 1A.¹⁵¹ Minority Group Selection Ratios and Four-Fifths Ratios When the Majority Group Selection Ratio Is 1%, 5%, 10%, 25%, 50%, 75%, 90%, 95%, or 99%

| Standardized group difference (<i>d</i>) | Majority group selection ratio | | | | | | | | |
|--|--------------------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| | 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 |

151. See sources cited *supra* note 150. The table shows the minority selection ratio (number hired/number applicants, non-bold number in each cell) and four-fifths ratio (minority selection ratio/majority selection ratio, bold number in each cell) as a function of the majority selection ratio (percent hired, from 1% to 99%, at top) and *d* (standardized group difference in performance, at left). Tabled values in bold less than .80 represent scenarios that violate the four-fifths rule.

The parameters in Figure 1A reflect the fact that, when an employer is seeking to hire the best qualified candidates based on scores on a personnel screening device and two groups (for example, blacks and whites) differ in the distribution of screening scores, the hires from each group, and the corresponding hiring ratios, will be a function of the magnitude of the group score differences as well as the overall number of positions available. These relationships can be determined from the pattern of two overlapping curves, as depicted in Figure 2. Assuming that the curves in Figure 2 represent each group's distribution of scores on a pre-job screen, the distance between curves will reflect the d value, or standard deviation difference in scores, between the groups (A & B). Smaller d values will show curves with more overlap; larger d values will show curves farther apart. Assuming that candidates will be selected from the top down, the cutoff score and the numbers hired (above the cutoff) from each group will reflect positions available. As d increases, fewer applicants will be hired from the lower scoring group (A) relative to the higher scoring group (B). Likewise, the fraction of candidates hired from each group will vary with the number of jobs available. As positions become more scarce and the vertical line moves to the right, the ratio of candidates selected from the lower scoring group (A) relative to the higher (B) will decline.

Figure 2: Distribution of Performance (Screening Criterion; or On-the-Job)



Corresponding to the relationships depicted graphically in Figure 2, the numbers in the body of Figure 1A represent the fraction of hires and the minority-majority hiring ratio (the numbers in bold type) expected in a competitive top-down selection process. Those parameters are a function of (1) the observed value of d , or the average standardized difference between the minority and majority groups in performance on a job screen (left side of the table); and (2) the 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), which reflects the number of jobs available. In sum, the table identifies the minority-majority 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 to select similarly (and the best) qualified workers from each group. Although the table does not incorporate data on the predictive value of particular job screens, the calculated relationships hold for all devices, regardless of validity. As already noted, estimates of the validity of conventional job screens is available in the IOP literature, which also reports evidence on racial differences in performance (d values) for those screens.¹⁵²

An examination of this table yields some important insights. First, for any degree of group disparity (d) in performance on a job test, 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.¹⁵³ Second, the ratio of blacks to whites hired varies with d , or the gap between whites and blacks in

152. See *supra* notes 97, 110 and accompanying text.

153. Accordingly, it is well recognized that placing significant weight on g -related predictors, especially for competitive jobs for which a relatively small percentage of applicants are hired, will severely limit or even eliminate minority hires. See, e.g., Neal Schmitt & Abigail Quinn, *Reductions in Measured Subgroup Mean Differences: What Is Possible?*, in ADVERSE IMPACT, *supra* note 50, at 425, 426 (“Selecting the highest-scoring individuals and a small proportion of applicants ... will virtually eliminate members of lower-scoring subgroups.”).

the pertinent qualification, such as the score on a job screening test. The greater the discrepancy—that is, the lower the average performance for blacks compared to whites—the smaller the black to white hiring ratio and the greater the adverse impact.

This analysis reveals an important defect in the four-fifths rule as currently applied: the existing benchmark does not vary with job selectivity—that is, the ratio of available job candidates to those hired. A fixed ratio makes sense under the assumption implicit in *Griggs* and its progeny, which is that there are no discrepancies in expected job performance between groups.¹⁵⁴ 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, and job competitiveness will have no effect on the relative number of hires from each demographic category. As noted, the empirical evidence reveals that this assumption of equal job readiness is unrealistic.¹⁵⁵ When one group lags behind another in actual performance, selectivity can significantly influence the expected hiring ratios from each group.¹⁵⁶ 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 *justified* (lawful and job-related) adverse impact will vary depending on real-world group disparities in performance.

The table in Figure 1A also shows that valid—and presumptively legitimate—job screens, including those in common use, will routinely violate the four-fifths rule. The zig-zag line superimposed on the table separates the combinations that satisfy the four-fifths 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 four-fifths. At the extremes—for example, using a pure test of cognitive ability generating a typical black-white disparity (d) of one standard deviation—only positions for which more than 95 percent of applicants (that is, virtually *all* applicants) are hired would achieve the target

154. See *supra* Part I (explaining *Griggs* and the development of the four-fifths rule).

155. See *supra* Part II.B (discussing measured performance disparities between blacks and whites).

156. See *supra* note 151.

ratio. When fewer applicants are hired, the ratio of black to white hires expected would always fall below the .8 fraction dictated by the four-fifths rule. Thus, hiring from the top down based on such a test would virtually always expose an employer to a possible disparate impact challenge.¹⁵⁷

As noted above, many selection devices show smaller group disparities than cognitive ability tests. Nonetheless, if employers hire the best candidates and applicants exceed positions, relatively few persons from lower scoring groups will be selected. The data reveal that the adverse impact for commonplace screening methods is still large enough to run afoul of the four-fifths rule on a regular basis. For instance, black-white d values as low as 0.25 have been reported for structured interviews.¹⁵⁸ For a difference of this magnitude, the four-fifths ratio is achieved only if enough positions are available to hire more than 50 percent of job candidates. A d value reported for educational credential (bio data) screens is 0.33.¹⁵⁹ To meet the four-fifths criterion, 75 percent of job seekers must be hired. Finally, one meta-study reports an average black-white difference on job sample task simulations of 0.38 standard deviations.¹⁶⁰ Satisfying the four-fifths rule would require hiring more than 75 percent of majority job candidates. These numbers indicate that compliance with disparate impact targets is possible only for relatively unselective positions. Yet employers make the heaviest use of personnel screening when staffing is competitive—the very situation in which results are likely to generate a prima facie case of disparate impact. And competitive hiring is commonplace in our society today.

As summarized by a review article on disparate impact compliance, “[c]learly, optimal use of valid selection tests that demonstrate subgroup differences cannot occur without significant impact on the realization of equal representation of different groups.”¹⁶¹ In other words, employers who seek out the most productive workers will

157. See Schmitt, Clause & Pulakos, *supra* note 96, at 116 (“[W]ith [job] selection ratios of 0.10, 0.50, and 0.90 and a subgroup standardized predictor difference of 1.00 [$d=1$ standard deviation], the proportion of a lower scoring group hired [relative to applicants from that group] would be 0.013, 0.159, and 0.61 respectively.”).

158. See Huffcutt & Roth, *supra* note 110, at 184.

159. Bobko, Roth & Potosky, *supra* note 123, at 565.

160. See Schmitt, Clause & Pulakos, *supra* note 96, at 119 tbl.4.1.

161. *Id.* at 116.

routinely hire too few minorities. 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 selection criteria are valid and important to the conduct of their business. As discussed in greater detail below, however, that possibility is cold comfort, because defending a disparate impact lawsuit is a costly and risky endeavor.¹⁶² The pervasive threat of disparate impact liability generates considerable unfairness, imposes ongoing burdens, and encourages evasions and distortions of the underlying goals of the doctrine. And these costs yield few benefits for minority workers, whose underrepresentation is ultimately traceable to factors that disparate impact was never designed to address.

IV. REFORMING DISPARATE IMPACT: ALTER OR ABOLISH IT

The IOP literature indicates that disparate impact law, in its current state, is flawed. Although legal scholars have noted the uncertainties created by the doctrine's ambiguities,¹⁶³ its disappointing record as an instrument for vindicating employee rights,¹⁶⁴ and the onerous requirements it imposes on employers,¹⁶⁵ they have paid almost no attention to group differences in skill and the challenges these pose for the nuts and bolts of personnel selection. A clear-eyed look at these realities reveals that the empirical assumptions underlying *Griggs* and its progeny are unfounded. In light of these shortcomings, this Part proposes a significant reform of the doctrine as applied, or, in the alternative, its 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 structured operates on the implicit premise that persons from varying racial and ethnic

162. See *infra* Part IV.

163. See *supra* note 74.

164. See, e.g., Michael Selmi, *Was the Disparate Impact Theory a Mistake?*, 53 UCLA L. REV. 701, 705-06 (2006) (discussing how disparate impact litigation has failed to deliver on the promise of increasing minority representation in the workforce and may have hindered efforts to curb discrimination).

165. See *supra* notes 57-71.

groups are similar in their skills, aptitude, or productivity.¹⁶⁶ This is not the case at present.¹⁶⁷ 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 four-fifths rule or that shows a statistically significant imbalance in workforce composition by race. Because commonly used personnel methods routinely produce adverse impacts, many employers' procedures will be vulnerable to challenge and to the requirement of demonstrating business necessity or job-relatedness.

This situation exposes many, if not most, employers to the threat of a Title VII disparate impact lawsuit. And defending such a lawsuit will necessarily be expensive, risky, and complex.¹⁶⁸ The need to establish the job-relatedness defense places businesses at the mercy of ambiguous rules and unsettled standards.¹⁶⁹ Formal validation is costly and often practically impossible, and the extent to which courts will demand such validation is unpredictable.¹⁷⁰ Efforts to establish content validation frequently depend on expert testimony, which is expensive and resource-intensive.¹⁷¹ The outcome of any validation inquiry is uncertain. These vagaries encourage businesses to engage in inefficient, counterproductive, and potentially illegal maneuvers designed to avoid disparate impact challenges.

Reforming disparate impact to relax the criteria for presumptive liability, or abolishing the doctrine altogether, will have the desirable effect of reducing wasteful litigation and lowering incentives to engage in perverse self-protective measures. Neither reform will undermine equal workplace opportunities nor decrease minority access to desirable jobs. The IOP evidence indicates that social forces beyond the reach of employment discrimination laws determine workplace demography. To put it bluntly, the paucity of minorities, and especially blacks, in most job categories is overwhelmingly due to their lower level of job-related skills compared to other groups in the workforce.

166. *See supra* note 6 and accompanying text.

167. *See supra* text accompanying notes 109-11.

168. *See supra* note 70.

169. *See supra* notes 56-70.

170. *See supra* note 70.

171. *See supra* note 70.

A. Alter It: Disparate Impact Realism

The discussion so far has focused on job selection and screening: the methods that employers use to decide whom to hire or promote into various positions. But the objective of such decisions is to select the best workers—those who will ultimately perform the job well. The proposal advanced here hones in on that goal and takes advantage of what we know about worker productivity. The suggestion is that disparate impact law should be revised to reflect group disparities currently measured in actual worker performance. The evidence 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 data available estimates that blacks and whites differ, on average, by roughly 0.3 standard deviations in performance outcomes for most jobs, with a reported range between about 0.24 and 0.39 standard deviations.¹⁷² That observation suggests, first, that screening practices currently in use are generating a workforce in which the productivity of black workers lags somewhat behind whites, but with smaller gaps than exist 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 of 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, assuming that employers are trying to operate as a competitive meritocracy geared towards maximizing productivity, the data suggest that they are missing the mark. Current levels of racial balance, whether in compliance with the disparate impact

172. See *supra* text accompanying notes 116-17.

rule or not, are being achieved through the employment of minorities who are, on average, somewhat less productive than whites.¹⁷³ This situation is out of sync with the meritocratic ideal that lies at the heart of the disparate impact rule.¹⁷⁴ A perfectly functioning meritocratic system should ideally show no racial gaps in job performance. All persons, regardless of their group identity, should be selected for jobs based on their actual productivity. But this is not quite what is happening today.¹⁷⁵

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 four-fifths rule—is too stringent. Although legal expectations are probably not the only factor behind this overshoot in minority hiring, they might well contribute to it.¹⁷⁶ And the status quo evidence shows that current disparate impact doctrine is not alleviating its effects.¹⁷⁷

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 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 work performance disparities through an appropriate downward adjustment in expected minority-majority hiring ratios in some cases. In effect, this would allow employers to screen out more minority workers than under the current rule. Ideally, this would cause race differences in job performance to narrow and perhaps even converge. Whether this would happen—and, for reasons explained below, it might not—the reform would still be

173. *See supra* note 104.

174. *See supra* note 41.

175. *See supra* note 103.

176. *See infra* note 241 and accompanying text.

177. *See supra* text accompanying note 149.

desirable because it would decrease employers' vulnerability to disparate impact lawsuits.

How would realism work? The first step is to make the stylized assumption that the productivity pattern of the background population matches the pattern IOP experts observe for workers themselves—as reflected specifically in observed average group differences in performance such as the 0.3-0.4 standard deviation disparity measured between black and white workers. This means that, if employers were to hire randomly from the background population (or appropriate pool of job candidates) regardless of race, the workforce would display group (for example black-white) differences in actual job performance that matched disparities currently observed. The second step is to calculate the group hiring ratios that would be *expected* under a valid and predictive meritocratic selection system as a function of the designated productivity gap and various levels of job selectivity. Here it is assumed, in keeping with the understandings underlying the current disparate impact rule, that employers are entitled to hire workers who will be most, and 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 for work productivity, with no race performance gaps observed among actual workers. Given current patterns of productivity, this will necessarily involve relaxing the assumption of racially proportionate hiring and promotion. Fewer blacks than whites will be selected from the applicant pool, and fewer will occupy many, if not most, jobs relative to their numbers in the population. In many cases, these ratios may fall significantly short of the four-fifths standard.

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.¹⁷⁸ A critical insight, however, is that these relationships between expected hiring ratios and job selectivity also hold when the parameter of group differences (*d* value) relates to *actual*, expected job performance, rather than to scores on job screening tests, which are

178. See *supra* Figure 1A.

merely predictive of performance.¹⁷⁹ Assuming that the d value represents the actual racial gap in expected job performance rather than the gap in performance on a job screen, the appropriate ratio of minority to majority hires for positions of varying selectivity, corresponding to the slots available, can once again be determined from the example graphically depicted in Figure 2.¹⁸⁰

This example assumes that the goal is to select the most productive persons available from the entire pool of job candidates, regardless of group identity. This entails selecting persons from each group who will perform the job equally well and should produce a workforce in which there are no detectable differences in productivity between workers from different groups. As previously noted, Figure 2 reveals that, when two groups differ in the level and distribution of expected job performance, the hires from each group, and the hiring ratios, can be determined from the pattern of two overlapping curves. As d (the difference in average productivity between groups) increases, the curves will move further apart, which will result in fewer hires from the less productive group. The number of candidates chosen from each group (above the vertical cutoff) will in turn be a function of the positions available. As positions become scarcer, the cutoff line will move to the right, which corresponds to an increase in the level of performance demanded from persons selected. Figure 2 reveals that fewer persons from the less productive group (A) are chosen relative to the more productive group (B) as the performance cutoff increases.

The values corresponding to these numbers can be identified from the charts depicted in Figures 1A and 1B, which contain identical values. In contrast with the example discussed above,¹⁸¹ the value of d can here be taken to represent actual job performance ratings, or productivity, rather than scores on pre-selection screens. The tabulation in Figure 1B reveals that, given a standardized group difference in actual productivity of about one-third of a standard deviation (which is roughly the black-white difference reported in

179. This application is anticipated by Sackett & Wilk, *supra* note 105, at 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”).

180. See *supra* Figure 2.

181. See *supra* notes 150-52 and accompanying text.

the IOP literature),¹⁸² 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 position in question. For example, for a job in which 25 percent of the majority applicants were hired, a black-white gap of 0.3 in expected productivity would predict a ratio of hiring percentages between blacks and whites of no more than .66. Although not negligible, this is significantly lower than the .8 ratio that satisfies the four-fifths rule. Likewise, for a d value of 0.4 (which is also within the range of on-the-job disparities reported in the literature), where 25 percent 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 percent of majority candidates, and assuming a d value between 0.3 and 0.4, the ratio of blacks to whites hired relative to the candidate pool 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 black to white hires of this magnitude or higher—would be immune from disparate impact challenge. Figure 1B depicts this point. The box defined by the broken line includes hiring ratios that would be expected for two groups with an average difference in job productivity (d value) of 0.3 or less, which is a rough estimate, perhaps on the low side, of the measured black-white difference. If that value were selected as the standard, then any hiring ratios within the box would be permissible under disparate impact realism and would not trigger presumptive disparate impact liability.

182. See *supra* text accompanying note 116.

Figure 1B.¹⁸³ 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 (<i>d</i>) | Majority group selection ratio | | | | | | | | |
|--|--------------------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| | 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 |

183. See sources cited *supra* note 150. The table shows the minority selection ratio (number hired/number applicants, non-bold number in each cell) and four-fifths ratio (minority selection ratio/majority selection ratio, bold number in each cell) as a function of the majority selection ratio (percent hired, from 1% to 99%, at top) and *d* (standardized group difference in performance, at left). Tabled values in bold less than .80 represent scenarios that violate the four-fifths rule.

It is important to reiterate that the implicit assumption behind disparate impact realism is that each group's pattern of productivity for the background job-eligible population corresponds to the distribution, and group differences, in job performance *actually* observed. This is a highly conservative assumption, because, as explained below, job incumbents often have a narrower range of abilities than the background population. Most jobs are somewhat selective, so job incumbents are better matched for ability, and show less racial disparity, than the population as a whole.¹⁸⁴ There is, however, no ready method, apart from random hiring and ex post screening,¹⁸⁵ to determine the performance profile of an unscreened population for particular jobs. Thus, the assumption that the background population matches job incumbents on this measure is, at best, a rough estimate that is highly favorable to underrepresented groups, because it almost certainly understates racial group differences in on-the-job performance that would be expected from random hiring across the full range of potential job-seekers.

Despite these limitations, disparate impact realism possesses several 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 four-fifths 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 reported 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 glean the expected hiring ratios from compilations of data similar to the charts depicted in Figures 1A and 1B. 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

184. See *infra* note 230 and accompanying text, including the discussion of "range restriction."

185. See *infra* notes 297-302 and accompanying text.

liability altogether. If he were sued, the case could be expeditiously resolved through summary judgment.

The hope is that any additional complexity will be more than compensated by the narrower scope of liability employers would face. Indeed, the most compelling reason to adopt this reform is to potentially reduce the number of workplace practices that are vulnerable to disparate impact challenges and thus narrow employers' exposure to potential liability. For cases that satisfy revised criteria but would previously have fallen short of the four-fifths 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.

To be sure, disparate impact realism does not give employers a completely free hand. Hiring ratios that fall short of the revised standards are still subject to challenge. Once again, the relevant benchmark should be roughly a d value of between 0.3 and 0.4—the average black-white on-the-job performance difference currently reported in the literature. As already noted, if the relevant value of d is set conservatively as 0.3 or below, the broken line in Figure 1B indicates the minority-majority hiring ratios expected in order to match groups for productivity. This means that any ratios within the box defined by the broken line are permissible under disparate impact realism. But Figure 1B also reveals that only situations falling *below* the solid line and *above* the broken line (that is, between those two lines) will produce expected hiring ratios that are *less* than those now required by the four-fifths rule.¹⁸⁶ Only in those instances will employers have more leeway than at present. In cases depicted on the rest of the chart, the standard will remain the same.¹⁸⁷ Moreover, even when hiring ratios expected under disparate impact realism (i.e., those within the broken line box) do fall significantly below those mandated by the four-fifths rule, they may still be substantial. For example, for a job in which a quarter of white candidates are hired, the expected ratio of black to white hires

186. It is nonetheless worth observing that d values greater than about 0.5-0.6 are not observed for most commonly used job screens. Therefore, the entire bottom portion of the chart will simply be irrelevant to most personnel selection situations.

187. See *supra* Figure 1B.

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. In addition, employers may still struggle to achieve target ratios, because most valid job screens in current use—such as job interviews, educational credentials, or tests of skill proficiency—produce group disparities that can exceed the average gaps in productivity presently 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 are rather modest. Nonetheless, the hope is that realism 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 when staffing is most selective. As Figures 1A and 1B indicate, the greater the number of candidates that apply for a position, the fewer the number of minorities employers will be expected to hire.¹⁸⁸ More specifically, the ratio of blacks to whites selected will be smaller for the most competitive positions, and black job-holders will be scarcer relative to whites. For example, if only 5 percent of white candidates are hired, the expected black-white hiring ratio is .52. For 1 percent, the ratio is .4.¹⁸⁹ These ratios represent a significant dialing back of the four-fifths standard. Although 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.

In sum, if one group lags behind another in performance on a job screening criterion, the adverse impact on that group will increase

188. *See supra* Figures 1A, 1B.

189. *See supra* Figure 1B.

as the percentage of candidates hired declines.¹⁹⁰ That is likewise true if one group lags behind another in actual job performance.¹⁹¹ It follows that such patterns should not be actionable, let alone result in liability. Under disparate impact realism, employers would be immune from challenge for a greater number of outcomes that are in line with expected workplace patterns.

Another reason to adopt this approach is that disparate impact 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.¹⁹² 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.¹⁹³ 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.¹⁹⁴

In suggesting, without deciding, that the Equal Protection ban on intentional discrimination might limit the race-conscious steps public or private employers could take to avoid violating the disparate impact rule, the *Ricci* decision potentially casts doubt on the legality of the disparate impact doctrine, at least in some circumstances, and adds to employers' uncertainty.¹⁹⁵ The specter of unconstitutional action is a compelling reason to 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 on on-the-job performance

190. *See supra* Figures 1A, 1B.

191. *See supra* Figure 1B.

192. *Ricci v. DeStefano*, 129 S. Ct. 2658, 2664 (2009).

193. *Id.*

194. *Id.*

195. *See id.* at 2683 (Scalia, J., concurring) ("[T]he 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"); *see also* Richard Primus, *The Future of Disparate Impact*, 108 MICH. L. REV. 1341, 1343-44 (2010) (discussing the constitutional implications of *Ricci*).

show that employers are not arbitrarily screening out minorities from jobs for which they are competitive.¹⁹⁶ As discussed below, there are several possible explanations for existing job productivity gaps by race. One factor may be that employers are responding to legal mandates and other societal pressures by overshooting the mark. That is, they may be modifying or relaxing selection practices to achieve greater racial balance.¹⁹⁷ However, the disparate impact rule does not require bending over backwards to achieve diversity, and this outcome is actually at odds with its stated purpose of equal opportunity.¹⁹⁸ In fact, the situation that best comports with the goals of disparate impact is a workforce that is equally productive regardless of race. That is precisely what disparate impact realism aims to achieve.

Fourth, given the demographic and workplace reality, virtually every employer today operates under the threat of potential disparate impact liability. This situation tempts employers to make choices that are either undesirable in themselves or run contrary to the objectives of the disparate impact rule. Disparate impact realism potentially reduces the incentive to adopt perverse strategies, at least in some cases.

Personnel practices that produce adverse impacts are pervasive. Although the elements of a *prima facie* case are up for grabs, many plaintiffs can easily meet the standard and employers are widely vulnerable to suit. Virtually no aspect of the business necessity doctrine is well-established, and uncertainties abound.¹⁹⁹ Regardless of the standard applied, establishing a defense of job-relatedness is cumbersome, difficult, risky, and expensive.²⁰⁰ Although courts can choose to impose strict or lenient requirements for demonstrating business necessity, the potential for failure to meet the standard is great. If the demand is for formal validation, an employer's 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 a high

196. *See supra* Part III.

197. *See infra* text accompanying note 284.

198. *See supra* text accompanying notes 39-43.

199. *See supra* notes 56-71 and accompanying text.

200. *See supra* notes 69-71 and accompanying text.

degree of expertise, including an awareness of methodological pitfalls and controversies.²⁰¹ 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 can withstand legal challenge, and those ratings must be 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 restriction²⁰² and other study design limitations. Ideally, this analysis should be conducted separately for distinct jobs. Even if feasible, this process is, at best, protracted, expensive, and riddled with the potential for error. Employers are often forced to engage experts to advise them on their practices and to assist them in satisfying these data-intensive legal requirements.²⁰³

In addition, outcomes are uncertain. The courts have left many questions surrounding the business necessity defense unaddressed or unresolved. They have provided little guidance as to when different standards of job-relatedness are appropriate, and judicial practice in this area is erratic and unpredictable.²⁰⁴ Indeed, the spate of challenges to civil service exams for police and firefighters—tests that are designed to reflect the specifics of the job and thus are good candidates for content-validation—demonstrates that there is no safe haven.²⁰⁵ Some courts have accepted content validation for civil service exams, whereas others have demanded a more rigorous showing.²⁰⁶ In any event, content validation, which often rests on the testimony of experts, is also complicated and costly. As the voluminous record in *Ricci v. DeStefano* reveals, satisfying a court

201. See *supra* note 76 and accompanying text.

202. See *infra* notes 230-31 and accompanying text.

203. See *supra* text accompanying note 82.

204. See *supra* notes 56-71 and accompanying text.

205. See *supra* note 75.

206. Compare *Ricci v. DeStefano*, 129 S. Ct. 2658, 2682 (2009) (accepting the content validity of the firefighters' test at issue), with *United States v. City of New York*, 637 F. Supp. 2d 77, 110 (E.D.N.Y. 2009) (rejecting content validity and faulting the absence of formal validation).

that a job selection test is facially valid can be a protracted, complex, and expensive undertaking that consumes considerable private and judicial resources.²⁰⁷ In addition, the Civil Rights Act of 1991 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.²⁰⁸ 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 loath to become embroiled in disparate impact disputes.

This reluctance creates a strong incentive for employers to engage in evasive tactics. Employers seeking to avoid the risks of litigation are forced to select from an unattractive menu of options, including abandoning the most valid hiring and promotion practices, engaging in covert affirmative action, or “race-norming” their selection criteria. Some employers may choose to sacrifice validity for diversity, either by relaxing job requirements 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 can serve as a cover for the application of different standards, criteria, or cut-offs across racial groups.²⁰⁹ Such race-conscious practices, including affirmative action or “race-norming,” are legally dubious under Title VII, contrary to the meritocratic underpinnings of the disparate impact doctrine, and, for public entities at least, arguably run afoul of the Equal Protection guar-

207. See *supra* notes 75-76 and accompanying text.

208. See *supra* note 73 and accompanying text.

209. See, e.g., Rutherglen, *supra* note 43, at 107 (“The 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 Issacharoff & Scharff, *supra* note 72, at 6 (“As disparate impact law took hold ... [e]mployers turned increasingly to affirmative action to buttress the representation of historically excluded groups.”).

antee.²¹⁰ Further, in the wake of the Supreme Court's decision in *Ricci v. DeStefano*, public employers have limited freedom to take race-conscious steps to avoid liability for racial imbalances in their workforces, and these restrictions may also extend to private employers.²¹¹ 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 ploys.²¹²

Perhaps the most important virtue of disparate impact realism is that it functions as an information-forcing device by enhancing

210. Congress amended Title VII of the Civil Rights Act in 1991 to outlaw so-called "race norming" or racial adjustments in scores on job tests. See 42 U.S.C. § 2000e-2(l), (m) (2006); see also *supra* note 43.

211. For a discussion of these issues, see generally Primus, *supra* note 195.

212. 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 limited as a function of factors like geography, local demographics, and standard job credentials—such as years of education and experience, or specialized training and licensing in a particular field—the four-fifths target 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 educations and backgrounds 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 point is that the law is unclear on the 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. See *supra* note 56 and accompanying text. 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. See *supra* note 54 and accompanying text. 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. See *supra* notes 111-13 and accompanying text. 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 meet basic criteria—will not necessarily perform equally well. Thus, adjusting the baseline candidate pool to account for such criteria will not necessarily erase group differences in job success. The current situation illustrates this point.

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.²¹³ The contention is that refinements in personnel selection technology could reduce the validity-diversity tradeoff.²¹⁴ Although the search for such methods has so far proven disappointing and there is good reason to believe it will ultimately fail, some researchers continue to suggest that more intensive efforts are needed and might bear fruit.²¹⁵ To the extent that better metrics might be found, however, they are best developed 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 liability in force today, the great majority of valid personnel practices can be expected to generate enough racial disparity to trigger liability under a broad range of real-world conditions.²¹⁶ Thus employers will rarely avoid a presumption of unlawful disparate impact despite their best efforts at alleviating the validity-diversity tradeoff. Because there is so little chance of finding substantially predictive screens that also satisfy the four-fifths rule, employers currently have no reason to search for and devise ways to meet the threshold requirements.

In contrast, under the somewhat lower expectations of disparate impact realism, there may be more circumstances in which employers can identify criteria that, although still substantially predictive of productivity, generate a workforce that is diverse enough to meet target ratios. Because 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 disparate impact standard

213. See *infra* Part IV.C.

214. See *supra* Part II.B.

215. See, e.g., Schmitt, Clause & Pulakos, *supra* note 96, at 117 (suggesting that more research is needed on measurement methods and approaches to “minimizing subgroup differences”).

216. See *supra* Part II.

potentially 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 for a more forgiving standard.²¹⁷ 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 should they be tied to continuously updated information? If not subject to frequent updates, does not disparate impact realism run the risk of freezing current racial inequalities in place, to the potential detriment of protected minorities? 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 of 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.²¹⁸ If disparate impact realism became the standard, experts could put less effort into the complicated business of validating job screens (by establishing their ability to predict job success). Instead, they could focus their attention simply on measuring on-the-job performance, including documenting group differences for various occupations. In fact, although those parameters are already part of the validation process, information on group

217. *See supra* text accompanying note 167.

218. *See, e.g.*, Mary D. Baker, Hunter B. Hughes, Gregory Mitchell & Philip E. Tetlock, *Proactive Responses to Second-Generation Risks in Labor and Employment Cases*, 37 EMP. REL. L.J. 28 (2011).

disparities could be improved. Data on the validity of employment selection methods is plentiful, but there is somewhat less evidence on the adverse impact of personnel screens and on group differences in actual productivity.²¹⁹ The research studies are not as refined and up-to-date as they could be.²²⁰ 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 information on group differences, in response to the proposed disparate impact reform. Although the need for high quality, reliable evidence imposes an initial burden on employers, this would presumably 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 could potentially deter many lawsuits. This type of evidence should be deemed probative at the initial stages of litigation, such as motions to dismiss and for summary judgment.²²¹

One oft-heard claim is that assessments of job performance are inherently unreliable because they are vulnerable to unconscious biases and subjective distortions.²²² A large literature has been

219. See Schmitt, Clause & Pulakos, *supra* note 96, at 120 (noting the relative paucity of adverse impact data relative to evidence on validity and speculating that this is due to “a concern regarding the legal implications of presenting data on subgroup differences”).

220. See, e.g., McKay & McDaniel, *supra* note 115, at 539 (noting the problem of relying on older data when estimating job performance differences); see also Schmitt, Clause & Pulakos, *supra* note 96, at 120 (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, the actual jobs examined, the selection screens employed, and statistical techniques and corrections applied. See *supra* note 110 and accompanying text. There is clearly room for improvement.

221. 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 as 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, *supra* note 63, at 1215 n.160 (“A small but not insignificant 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.

222. 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., Wax, *supra* note 121, at 981 (noting the contention that subjective job evaluations are vulnerable

devoted to scrutinizing the accuracy, reliability, and reproducibility of job performance assessments with particular attention to whether subjective ratings are tainted by impermissible factors such as race. Although some research claims to detect race effects in worker evaluations, the expert consensus is that subjective supervisor ratings are reliable and a fairly accurate reflection of work performance overall.²²³ In particular, studies reveal a high concordance between subjective appraisals and objective, quantifiable measures, such as absenteeism, tardiness, work errors, tasks completed, and work unit output.²²⁴ 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

to distortion by inadvertent or unconscious anti-minority bias); *see also* John Monahan, Laurens Walker & Gregory Mitchell, *Contextual Evidence of Gender Discrimination: The Ascendance of "Social Frameworks,"* 94 VA. L. REV. 1715, 1715, 1733 n.51 (2008) (describing challenges to subjective and discretionary personnel evaluations at Home Depot and Walmart).

223. *See, e.g.,* Outtz & Newman, *supra* note 97, at 76-78 (concluding that racial bias in subjective job assessments is marginal and has little effect on rating accuracy). Even when the data suggest 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 workers or in favor of same-race workers, or whether black or white supervisors (or both) are biased. That is because there is often no stable, independent standard against which to measure the accuracy of subjective ratings. *See, e.g.,* J. Kevin Ford, Kurt Kraiger & Susan L. Schechtman, *Study of Race Effects in Objective Indices and Subjective Evaluations of Performance: A Meta-Analysis of Performance Criteria*, 99 PSYCHOL. BULL. 330, 334 (1986) (summarizing and discussing data on rater and ratee race effects); Kurt Kraiger & J. Kevin Ford, *A Meta-Analysis of Ratee Race Effects in Performance Ratings*, 70 J. APPLIED PSYCHOL. 56, 58 (1985) (noting that the 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").

224. *See, e.g.,* Kelman, *supra* note 63, at 1211 (noting concerns about bias in subjective workers ratings, but discussing one important study in which subjective and objective measures were aligned). 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. In general, they find that subjective assessments tend to be more reliable when procedural checks and safeguards against inconsistencies and arbitrariness are introduced. *See, e.g.,* Baker, Hughes, Mitchell & Tetlock, *supra* note 218, 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, for example,* McKay, *supra* note 97, at 251-52.

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.²²⁵

In any event, the objection 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.²²⁶ But that defense rests on demonstrating a correlation between a job selection method and occupational job performance, which requires rating workers on the job. 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,²²⁷ 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 background pool of eligible workers.²²⁸ Does that assumption make any sense?

Addressing this question requires a more nuanced understanding of racial differences in job performance in relation to the spectrum of abilities among job candidates and the winnowing function of job

225. See Outtz & Newman, *supra* note 97, 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 Baker, Hughes, Mitchell & Tetlock, *supra* note 218, 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 field studies by [IOP] psychologists to the contrary.”); Ford, Kraiger & Schechtman, *supra* note 223, at 335 (“[C]onsistent effects on all ... criterion types [for performance assessments] suggest that there are race differences in job performance in the organizations sampled.”).

226. See *supra* note 4 and accompanying text.

227. See *supra* note 58 and accompanying text.

228. See *supra* Part IV.A.

selection devices. In light of employers' interest in finding workers with the best predicted future performance, employers will try to match applicants to the requirements of the job and screen out people who will perform less well. Accordingly, job screening protocols are designed to choose people with a similar ability to do a job, regardless of their racial identity. Thus, effective staffing practices should ideally minimize or eliminate background group differences in the people hired for particular positions.²²⁹

Because job incumbents, regardless of group identity, are often chosen to possess similar levels of skill, 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.²³⁰ Range restriction tends to be greatest for jobs that turn away the most applicants and are the most competitive.²³¹ The most competitive jobs often, although not always, require higher levels of skill. Likewise, job screening may be minimal for some lower level jobs. But even unskilled workers are not hired entirely at random. Although the situation varies widely with economic conditions and the position at issue, 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 other standard job selection criteria, which tend to be narrower than for pure ability tests, can also exceed black-white performance gaps seen on the job, although by

229. See *supra* text accompanying note 175.

230. See, e.g., John E. Hunter, Frank L. Schmidt & Michael K. Judiesch, *Individual Differences in Output Variability as a Function of Job Complexity*, 75 J. APPLIED PSYCHOL. 28, 29 (1990) ("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."). Range restriction is an important methodological problem in the IOP field, affecting estimates of all parameters relevant to personnel management and their correlations. See, e.g., Sackett, Borneman & Connelly, *supra* note 116, at 217 ("Failure to take range restriction into account can dramatically distort research findings [on personnel selection].").

231. See *infra* note 240.

less.²³² Second, as noted, IOP experts have observed that performance ratings for jobs at every level depend in part on personal, non-cognitive attributes that are not well-captured by conventional selection methods.²³³ These attributes are de-emphasized in present job selection relative to cognitive skills because they do not correlate as strongly with performance and cannot be measured as precisely or accurately.²³⁴ To the extent that commonly used personnel screens fail fully to gauge the range of performance-related behavior, scores on selection criteria will not perfectly predict job success.

These observations help explain why racial job performance gaps are smaller than on standard selection criteria. They also help account for the fact that racial disparities in job ratings persist despite some degree of screening. As noted, a perfectly functioning meritocratic system using omnisciently predictive metrics should produce no racial gaps in job performance at all. 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.²³⁵

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.²³⁶ Several factors probably contribute to this. First, existing personnel devices are inherently imperfect, and thus fail to control fully for all background differences or attributes that bear on performance.²³⁷ Skill disparities in the background population will thus tend to

232. The average difference in cognitive ability for blacks and whites hovers around one standard deviation, whereas most measured differentials in job performance range between about 0.24 and 0.39 standard deviations. *See, e.g.,* Outtz & Newman, *supra* note 97, at 85 (“[S]ubgroup differences on cognitive ability tests are far larger than subgroup differences on actual job performance.”); *see also* Murphy, *supra* note 123, at 138 (“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 McKay, *supra* note 97, at 249-70.

233. *See* discussion of “task performance” and “contextual performance,” *supra* note 123 and accompanying text.

234. *See supra* text accompanying notes 136-48.

235. *See supra* text accompanying note 186.

236. *See supra* notes 116-17 and accompanying text.

237. *See supra* notes 39-42 and accompanying text.

carry over into the workplace.²³⁸ Second, employers operating in the shadow of current antidiscrimination 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. As noted, a commonly relied-on job credential is years of education or specialized training ("biodata"). However, hiring people with similar years of schooling will not eliminate group disparities because years of education are an imperfect proxy for actual skills, on which blacks and Hispanics tend to be deficient.²³⁹ Thus, matching workers based on years of schooling will not necessarily eliminate racial gaps in proficiency on the job.²⁴⁰ In sum, although workers in particular jobs should ideally be equally able regardless of race, they are not. Group differences persist in measured job success.

238. The target of performance parity is an ideal that, even in a race-blind system, will likely never be perfectly achieved as long as skill differences persist between groups. Employers typically hire people who are best qualified first but then proceed to fill slots from the top down. Alternatively, they hire candidates who possess minimal qualifications but who represent a spectrum of abilities above a designated threshold. Either method can produce a range of actual productivity among those hired, because not everyone will perform exactly the same. Given differences in the background population, the *distribution* of skills among persons hired using even race-blind, top-down protocols will likely not match by race. For all these reasons, the practical reality will deviate somewhat from the ideal.

239. See Robert E. Ployhart, *The Diversity-Validity Dilemma: Strategies for Reducing Racioethnic and Sex Subgroup Differences and Adverse Impact in Selection*, 61 PERSONNEL PSYCHOL. 153, 167 (2008) (noting that years of education are not a good proxy for actual ability when used without gauges of actual knowledge, skill, and achievement); see also Philip L. Roth & Philip Bobko, *College Grade Point Average as a Personnel Selection Device: Ethnic Group Differences and Potential Adverse Impact*, *supra* note 199, at 402 tbl.1 (noting 0.78 standard deviations difference between blacks and whites on college grade point average). See generally ESPENSHADE & RADFORD, *supra* note 112.

240. 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, and thus will more closely reflect, their background population. Because, as noted, *g*-loaded criteria are the most predictive of job success, and the black-white difference on such measures approaches one standard deviation, the race gap in performance should be greater for workers who are not stringently screened for ability. See *supra* notes 101-02 and accompanying text. 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 comparing race performance gaps across jobs as a function of selectivity, and there is little useful data analysis on this point.

How do these observations bear on disparate impact realism? Realism ties hiring ratios to group differences in productivity that are 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.²⁴¹ 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 crucial, and the fact that employers must hire from a range of abilities to fill jobs also likely plays a role. 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 the question of whether measured patterns of job performance precisely reflect group profiles in performance ability in the background workforce is almost certainly “no.” It also reveals why this does not 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. By allowing employers to be more selective than under the four-fifths rule, realism is designed to advance the objective of equal standards, and equal job performance, regardless of race. Although employers’ target ratios are predicted on the basis of actual job performance gaps, the realism rule is not designed to preserve them. Rather, it is designed to close them. The hiring ratios expected, as illustrated in 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).²⁴² Thus, the smaller minority-majority ratios permitted by disparate impact realism would ideally bring the

241. See *supra* note 105 and accompanying text.

242. See *supra* Figures 1A, 1B.

measured job performance ratings of blacks and whites closer together than they currently are. In any event, the ultimate goal is to create a workforce in which people from different groups perform the same, and there are no measurable differences in productivity by race.

There is a problem lurking here: 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 observed performance gaps, that will potentially alter the magnitude of the very benchmark (racial differences in 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 itself, which is based on the existence of group skill differences in the background population—a situation that disparate impact realism, and the original disparate impact rule itself, has no power to address or alleviate.

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 a change in employment selection practices under disparate impact reform. After all, a relaxation of the disparate impact rule does not in itself alter the distribution of human capital among potential job candidates. On the other hand, if blacks improve 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 four-fifths rule despite the danger of incurring liability.²⁴³ Indeed, given the size of racial skill gaps and the adverse impact of many routinely employed job filters, many workplaces can be expected to fall short of the diversity required even by the relaxed standards of 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 measurable effect on minority representation in the workforce or patterns of 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.

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 continuous updating of evidence is required, then any narrowing of incumbent performance from the operation of disparate impact realism would tend to 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 protocol 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.²⁴⁴ Although those differences might significantly narrow in the future with improvements in education, experience, or training, this has yet to occur.

243. *See supra* note 148 and accompanying text.

244. *See supra* note 112 and accompanying text.

Such developments, or other evidence pointing to significant society-wide shifts in the distribution of human capital, would force a reevaluation of the rule and the benchmark ratios used in applying it. Short of that, however, currently observed productivity profiles 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 disproportionately excluding *able* minorities. Rather, the evidence suggests just the opposite.²⁴⁵ To the extent employers 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. Even if realism does not close racial gaps in work performance completely, employers will still be better off. The data shows that employers are tolerating existing group

245. It could be objected that the observed pattern of somewhat lower average productivity for blacks relative to whites does not rule out the possibility that employers are imposing arbitrary hurdles contrary to disparate impact's requirements. For example, an employer could adopt a selection method that disproportionately (and selectively) screened out more capable blacks in favor of the less capable. This scenario is farfetched in light of what is known about commonly used job screens, 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* notes 97-102 and accompanying text. 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 racist employer might want to keep blacks from attaining success, but rejecting the most capable job candidates is a peculiarly self-defeating way to accomplish this. The more likely strategy is just to exclude blacks from a specific 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 an unlikely scenario under current conditions, in which the percentage of qualified blacks is lower than whites for many, if not most, occupations.

disparities on the job.²⁴⁶ They are in fact living with them. And even if workplace demographics do not 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 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 and the relative shortage of qualified minority workers.²⁴⁷

246. See *supra* notes 114-18 and accompanying text.

247. 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., Robert G. Schwemm & Jeffrey L. Taren, *Discretionary Pricing, Mortgage Discrimination, and the Fair Housing Act*, 45 HARV. C.R.-C.L. L. REV. 375, 416-17 (2010) (describing lawsuits alleging unlawful disparate impact in mortgage lending practices); Sara Aronchick Solow, *Racial Justice at Home: The Case for Opportunity-Housing Vouchers*, 28 YALE L. & POL'Y REV. 481, 488 (2010) (noting that the federal Fair Housing law "outlaws disparate impact in housing just as Title VII does in employment"); see also Ayres, *supra* note 55, at 32-44 (describing lawsuits under the Equal Credit Opportunity Act (ECOA) against car dealers, car loan underwriters, and mortgage lenders for practices that have an adverse impact on minorities); *id.* at 26-27 n.47 (quoting a commentary on ECOA regulation stating that "[t]he act and regulation may prohibit a creditor practice that is discriminatory in effect because it has a disproportionately negative impact on a prohibited basis" (citing Official Staff Interpretations, Regulation B (Equal Credit Opportunity), 12 C.F.R. pt 202, Supp. I, § 202.6(a)-2 (2009))).

As for claims involving gender, these tend to target requirements for physical strength, appearance, or fitness, on which the genders do differ significantly. Disparities in cognitively related abilities between men and women tend to be small or nonexistent, although there are exceptions (for example in fields like math or engineering and especially for jobs requiring very high ability). Thus, adjustment of the four-fifths rule would generally not be indicated. See Selmi, *supra* note 164, at 746 (discussing lawsuits claiming disparate impact by gender, including class actions against Walmart and Home Depot). On disparate impacts generally, see Harris & West-Faulcon, *supra* note 41, at 112 n.147 (noting contexts in which the courts have "declined to intervene to ameliorate racial disparities" generated by facially neutral

B. Abolish It

Disparate impact realism does not change much. Although it relaxes previous requirements in some circumstances, it preserves them in many. 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 observed work 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. As noted, one hope is that realism will give employers greater leeway to search for qualified workers without fear of being sued. Another is that it will at least protect businesses from potential liability for some existing practices. Alternatively, realism may have little or no effect on employers' vulnerability to disparate impact challenge. Many prevalent job screens—such as educational credentials and structured interviews—show black-white differences greater than 0.3 standard deviations, which place them outside the “realism box” as depicted in Figure 1B²⁴⁸ and thus beyond the reach of the modest safe harbor the proposal creates. If most employers are currently using screens that exceed the adverse impacts permitted even under realism's more forgiving standards, the reform will not only fail to alter actual hiring practices but will also not shield more employers from lawsuits. Whether realism will shrink the actual scope of potential liability relative to the current rule is ultimately an empirical question.

In sum, it is possible that the realism proposal detailed above will make no practical difference. This observation suggests that a better, albeit a more radical, approach would be to abolish the disparate impact doctrine altogether, at least as applied to race in employment.²⁴⁹ This would be equivalent to overruling *Griggs v.*

laws).

248. See *supra* Figure 1B.

249. See *supra* note 247 (noting that disparate impact challenges can also be made against employment practices with differential effects by gender and against practices in other

Duke Power Co. and amending Title VII to remove racially disparate impact as a basis for employment discrimination liability.

The same points in favor of relaxing disparate impact's requirements also strongly support the doctrine's repeal. If the doctrine is applied correctly, workplace diversity will not increase. If the doctrine is applied incorrectly, which will happen with regularity, the effects will be perverse and counterproductive. As already noted, a shortage of qualified minority workers, not the widespread use of non-merit based selection, is the principal culprit behind the lack of workplace diversity. As long as the present demographic situation persists, adverse impacts are to be expected. Moreover, the vast majority of commonly employed selection criteria are valid and job-related and produce the greatest diversity consistent with workplace efficiency. Because the prevailing status quo generally comports with disparate impact's requirements, litigation should not lead to greater racial balance and retaining the disparate impact rule will advance no equal opportunity goals. Thus, even if employers are able successfully to defend their personnel practices and manage to resist changing them, disparate impact enforcement is wasteful and pointless and imposes unnecessary burdens.

Disparate impact lawsuits also carry too great a risk of unjustified liability. Because most sound, useful, and predictive personnel practices show enough adverse impact to violate the four-fifths rule, the standard for triggering a prima facie case is fatally overbroad and ensnares far too much conduct in its net. This means that most employers, if sued, will face the burden of justifying their staffing practices. Given the legal uncertainties and practical difficulties surrounding this task, employers run a significant risk of being found liable regardless of whether their methods are valid or whether they are actually violating the rule. To be sure, relatively few disparate impact cases are filed compared to those that could be, and employers usually prevail.²⁵⁰ But employers who are sued still

contexts).

250. See, e.g., Selmi, *supra* note 164, at 739-40; Wendy Parker, *Lessons in Losing: Race Discrimination in Employment*, 81 NOTRE DAME L. REV. 889, 899 (2006) (noting that disparate impact employment actions are relatively uncommon and that impact claims are almost always coupled with disparate treatment allegations); see also Sherwyn & Heise, *supra* note 35, at 906 ("[T]he vast majority of discrimination cases are disparate treatment or intentional discrimination cases.").

face the prospect of protracted, expensive, uncertain, and resource-intensive litigation. This encourages them to engage in perverse, inefficient, and evasive tactics, including adopting convoluted and less reliable job selection criteria or using race-conscious selection techniques.

There are other good arguments in favor of repealing the disparate impact doctrine. First, most claims brought under Title VII allege unlawful disparate treatment.²⁵¹ 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.²⁵² But the evidence suggests that the costs imposed by these cases are not worth the alleged benefits. These lawsuits place an enormous burden on local governments and consume considerable time, attention, and resources that could be devoted to other purposes. Although these legal challenges may marginally increase diversity in select instances, the objective could be accomplished more simply by other means, including abolishing civil service exams entirely, or selecting randomly from minimally eligible pools of applicants. 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 research on performance patterns of police and firefighters by race is relatively sparse, the existing data indicate that civil service exams are good predictors of success in these types of jobs.²⁵³ The familiar validity-diversity tradeoff applies as much to firefighter and

251. See Parker, *supra* note 250, at 898-99.

252. See, e.g., Ricci v. DeStefano, 129 S. Ct. 2658 (2009); MacDonald, *supra* note 75 (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 Norton, *supra* note 75, at 254, nn.233-34 (detailing cases presenting challenges to civil service exams).

253. See, e.g., MICHAEL G. AAMODT, RESEARCH IN LAW ENFORCEMENT SELECTION 34 (2004) (noting that cognitive ability, and to a lesser extent civil service exams, predict evaluations of police performance); Gerald V. Barrett, Michael D. Polomsky & Michael A. McDaniel, *Selection Tests for Firefighters: A Comprehensive Review and Meta-Analysis*, 13 J. BUS. & PSYCHOL. 507, 507 (1999) (finding that commonly used civil service tests predict training expertise and supervisor ratings of firefighters).

police positions, and civil service jobs generally, as to other occupations in the economy.²⁵⁴ Given racial gaps in developed skills and abilities, racial imbalances in these jobs will likely persist within any kind of selective 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.²⁵⁵ This pattern is not confined to elite and lucrative positions. The U.S. government has long used tests of cognitive ability, or tests 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.²⁵⁶ Yet such practices persist without serious challenge. It may be argued that plaintiffs do not

254. Of course, disparate impact plaintiffs challenge or simply ignore this evidence. See MacDonald, *supra* note 75 (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 firefighters' qualifying exam given by New York City). Plaintiffs in these cases routinely 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. See *supra* notes 136-48 and accompanying text.

255. Under *Wards Cove Packing v. San Antonio*, 490 U.S. 642 (1989), and section 105 of the Civil Rights Act of 1991, plaintiffs must ordinarily specify the practice or procedure that is the source of the observed disparate impact. However, the credentialing and training requirements for every white collar and professional job—such as nurse, lawyer, law clerk, accountant, physician, pilot, engineer, computer programmer, college professor, teacher, or administrator—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 HARV. L. REV. 945 (1982). This author is unaware of any cases featuring disparate impact challenges to professional credentialing requirements.

256. According to a recent report, whereas 16 percent 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 percent of African American applicants scored below the cutoff. CHRISTINA THEOKAS, THE EDUC. TRUST, SHUT OUT OF THE MILITARY: TODAY'S HIGH SCHOOL EDUCATION DOESN'T MEAN YOU'RE READY FOR TODAY'S ARMY 3 (2010), http://www.edtrust.org/sites/edtrust.org/files/publications/files/ASVAB_4.pdf. Among those achieving the minimum threshold score for admission to the military, over 43 percent of white test takers, but fewer than 18 percent 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. *Id.* at 5.

bother to sue because they believe judges will defer to the military or to professional standards. Although that may be true, this situation is arguably unfair to private employers of nonprofessional staff 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 impose hurdles or revert to selection criteria 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 antidiscrimination 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 for repeal rests on a clear-eyed assessment of the main forces underlying racial imbalances in the workplace today.

The underrepresentation of minorities in large segments of the job market is overwhelmingly the result of real skill disparities rather than employer indifference to unjustified racial impacts. Indeed, measured patterns are far more consistent with de facto affirmative action than with unlawful adverse impact. And even if employers moved to significantly *less* diversity in the wake of disparate impact repeal, that would 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 could well generate a greater racially adverse impact than current practices. Strictly speaking, this would comport with current doctrine: the courts have never barred employers from adopting the *strongest* and most *valid* predictors of job success, regardless of adverse impact, and using such predictors would presumably be consistent with the job-relatedness requirement. If the disparate impact doctrine were repealed, however, employers would not have to demonstrate the validity of their screens or worry that their job-relatedness defense would be rejected. They would not have to defend their practices at all. This would significantly lighten their burden and be more efficient in the long run.

In sum, existing differentials are more than accounted for by supply-side differences in job preparation or other cognitive or noncognitive group-based factors. Despite the weaknesses of prevalent job selection devices, the holy grail of more accurate selection with less adverse impact is unobtainable. At present, there is no escape from the validity-diversity tradeoff. This is not just a matter of limitations in the technology of personnel selection, although those limitations exist. It is grounded in the realities of the current distribution of human capital.

Workforce imbalances are likely to persist without significant changes in the distribution of skill. But it is important to remember that inequalities in job qualifications are not of employers' making, and that employers are ill-equipped to correct them. Likewise, supply side disparities are no concern of disparate impact law, and the doctrine is not designed to address 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."²⁵⁷

Those who view a racially balanced workforce as desirable in itself may feel little compunction about forcing employers to "redress the cumulative disadvantages" that minorities suffer.²⁵⁸ On this view, a disparate impact rule that forces firms to employ a more racially balanced workforce is all to the good, regardless of the niceties of the law's actual requirements. The problem with this approach is that the desired result can only be achieved by ignoring or misapplying the disparate impact rule itself. Racial balance depends on encouraging businesses to engage in self-protective affirmative action or in inducing them to relax their standards across the board. But it is perverse to use disparate impact to accomplish a result that is at odds with the doctrine's stated purpose, which is to enforce a race-blind meritocracy. Disparate impact was not meant to accomplish by subterfuge and indirection what the doctrine itself rejects. Moreover, imposing affirmative action through the threat of disparate impact liability represents a

257. Rutherglen, *supra* note 43, at 110.

258. *Id.*

judicial usurpation of the lawmaking function. By achieving a result at odds with stated statutory principles, courts effectively circumvent ordinary political channels. If racial balance and greater diversity are good policy, those priorities should be enacted directly and defended on their merits, with a clear articulation of expectations and requirements.

There is currently no general legal or constitutional requirement that employers operate as a meritocracy. Managers need not hire and promote on the basis of workers' ability to do the job, so long as they do not rely on forbidden factors.²⁵⁹ It is a paradox of disparate impact that employers are restricted to some form of meritocratic selection—that is, selection on criteria shown to be job-related—only if the workplace shows a racial imbalance. Although meritocratic staffing is fundamentally at odds with using race to achieve demographic diversity, antidiscrimination law does not close off all channels for achieving more racial balance than meritocratic selection would allow. For example, as discussed more below, the solution for civil service jobs may lie in discarding competency or performance-based screens altogether, and adopting other neutral and non-meritocratic approaches—such as lotteries or random selection procedures—that ensure maximal workforce diversity without perverting the law.²⁶⁰

One frequently voiced argument in favor of retaining the disparate impact doctrine is that disparate impact liability is needed to “smok[e] out” forms of subtle race- or sex-based discrimination that might be “cloaked in race-neutral selection processes.”²⁶¹ In other words, disparate impact challenges provide “a way of finding the stealth disparate treatment case” in which workers suffer discrimination *because of* a protected characteristic.²⁶² One problem

259. See, e.g., Sherwyn & Heise, *supra* note 35, at 910 (noting the Supreme Court's holding in *Texas Department of Community Affairs v. Burdine*, 450 U.S. 248 (1981), that there is “no obligation to hire the best candidate for a job. Instead, the employer simply could not discriminate”).

260. See *infra* notes 294-97 and accompanying text.

261. Harris & West-Faulcon, *supra* note 41, at 114.

262. *Id.*; see also *Ricci v. DeStefano*, 129 S. Ct. 2658, 2682 (2009) (Scalia, J., concurring) (noting the “smoking out” rationale); Richard A. Primus, *Equal Protection and Disparate Impact: Round Three*, 117 HARV. L. REV. 493, 498-99 (2003); Sherwyn & Heise, *supra* note 35, at 906 (“Disparate treatment cases are particularly difficult to assess because fact finders must ascribe motivation to the actions of the employer... [M]ost employers are now

with this rationale, as noted, is that the amount of racial imbalance that triggers a prima facie case of disparate impact also provides little support for unlawful disparate treatment. As already argued, racial disparities alone are at present exceedingly weak evidence for forbidden conduct under either theory because supply side factors can be expected to generate pronounced disparities even in the absence of unlawful discrimination.²⁶³

Nonetheless, the “smoking out” argument embodies the perception that the disparate impact framework is a 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.²⁶⁴

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 excluding minorities would clearly fall within the ambit of forbidden disparate treatment under Title VII, because the policy was adopted “because of race.”²⁶⁵ And the contention that a disparate impact rule is necessary because the disparate treatment doctrine covers only

sophisticated enough to avoid creating the proverbial smoking gun that would easily establish unlawful intent.”).

263. See, e.g., Wax, *supra* note 121, at 1022.

264. See *Watson v. Fort Worth Bank & Trust*, 487 U.S. 977, 988, 990-91 (1988) (discussing hiring decisions made using subjective criteria); Carle, *supra* note 43, at 258 (noting that “[s]upporters of disparate impact analysis also advance arguments based on the difficulty of proving hidden prejudice” and “the problems of subtle and subconscious bias”); Patrick S. Shin, *Liability for Unconscious Discrimination? A Thought Experiment in the Theory of Employment Discrimination Law*, 62 HASTINGS L.J. 67, 75-83 (2010) (hypothesizing that an employer might inadvertently rate “work experience” as more important when white candidates possess comparatively more work experience than blacks); Elaine W. Shoben, *Disparate Impact Theory in Employment Discrimination: What’s Griggs Still Good For? What Not?*, 42 BRANDEIS L.J. 597, 607-13 (2004) (asserting that a disparate impact theory might succeed in imposing liability in meritorious cases where disparate treatment allegations would fail).

265. 42 U.S.C. § 2000e-2(a) (2006) (providing that it is unlawful “to fail or refuse to hire or to discharge any individual, or otherwise to discriminate against any individual ... because of such individual’s race, color, religion, sex, or national origin”).

“intentional”—that is, conscious or deliberate—discrimination is not supported by the statute’s language, which forbids adverse treatment “because of race.”²⁶⁶ 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 courts are somewhat confused on the question of whether Title VII covers unconscious as well as conscious disparate treatment, the statutory language fits best with the broader interpretation.²⁶⁷ Thus, the disparate impact rule is not necessary to ensure that inadvertently biased conduct is actionable under Title VII. Although unconscious disparate treatment may be difficult to prove, that problem exists regardless of the theory of discrimination 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 nothing would be lost by abandoning a disparate impact analysis.

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 or bias that is “intentional.”²⁶⁸ But that limitation is properly understood to differentiate between 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 is 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

266. See Shin, *supra* note 264, at 80-82 (noting that Title VII’s “statutory language does not expressly say that an action must involve conscious consideration of a factor before it can constitute discrimination because of that factor” and that “[t]he semantics of the ‘because of’ construction seem similarly open to the possibility of liability for unconscious discrimination”); Wax, *supra* note 121, at 982-83.

267. See Shin, *supra* note 264, at 80-82; see also Amy L. Wax, *Discrimination as Accident*, 74 IND. L.J. 1129, 1137-39 (1999).

268. See, e.g., *Washington v. Davis*, 426 U.S. 229, 239-40 (1976); see also Wax, *supra* note 267, at 1138 n.20.

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 of constitutional protections. Once again, a disparate impact analysis should not be needed to get at situations in which 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 race-conscious, pro-diversity regime, is based on a critique of many competitive job selection methods as fundamentally unfair to poorly performing groups—and especially to blacks. Because strict enforcement of disparate impact doctrine will presumably discourage competitive practices that screen out minorities, preserving the status quo will minimize this unfairness.

Although initially directed at pure tests of cognitive ability, this type of accusation has been leveled at all competitive job selection methods that show significant disparities by race.²⁶⁹ The thrust of the critique is that such devices generate too many false negatives among lower-performing minority groups. That is, these selection methods end up excluding too many minority candidates who could adequately perform the job in question.²⁷⁰

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 prescreening devices,²⁷¹ have a correlation of around .5 with measured job performance.²⁷² If combined with the common practice of hiring from the top down or above a given cutoff score, these tests will generate a significant amount of error in the form of false negatives (people who fail the test but could perform

269. See, e.g., Kelman, *supra* note 63.

270. See *id.* at 1125 (describing false negatives as the failure to hire or promote candidates who would “in fact succeed on the job”).

271. See Linda S. Gottfredson, *The Science and Politics of Race-Norming*, 49 AM. PSYCHOLOGIST 955, 955 (1994).

272. See *supra* note 97 and accompanying text.

the job) and false positives (people who pass the test and are hired but perform the job poorly).

Moreover, it can be demonstrated numerically that, for any valid selection device on which one group outperforms another, the absolute (population adjusted) number of false negatives among applicants in the higher-performing group is smaller than in the lower performing group,²⁷³ which, in the case of cognitively loaded tests, is blacks.²⁷⁴ The reason for this is that a relatively larger number of persons from the lower-performing group will fall below the selection cutoff.²⁷⁵ Likewise, the absolute number of false positives will be greater among the higher-performing group (whites).²⁷⁶ 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 or the distribution of performance.²⁷⁷

273. See Kelman, *supra* note 63, at 1223-27 (showing numerically the effect of false negatives on a group that performs worse on a test relative to a higher performing group); see also Gottfredson, *supra* note 271, at 956 (noting the National Academy of Sciences Committee's critique regarding the effect of cognitive ability tests: "Minority workers at a given level of job performance have much less chance of being selected than majority workers at the same level of job performance, and thus are burdened with higher false-[negative] rates").

274. See, e.g. Gottfredson, *supra* note 271, at 956 (noting that blacks scored one standard deviation below whites on the GATB); Sackett & Wilk, *supra* note 105, at 929.

275. For a test on which blacks score lower than whites, more blacks will fail, and there will be more false negatives among this lower performing group. The false negative rate at any given score is applied against many more negatives (failing scores) from the low performing group, thus generating a relatively higher number of persons who could do the job but are not hired. See Kelman, *supra* note 63, at 1223-27; Sackett & Wilk, *supra* note 105, at 935.

276. See Kelman, *supra* note 63, at 1226 ("[W]hites 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 do not make the cutoff score (false negatives). Likewise, relatively more people from the higher performing group will pass the test, including more people who ultimately cannot do the job (false positives).

277. See Sackett & Wilk, *supra* note 105, at 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"); see also Kelman, *supra* note 63, at 1230 ("[I]n any ex ante probabilistic screening system, the existence of false positives and negatives ensures that factual equals will not be treated equally.").

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 adopted a number of tacks. First, some have challenged the notion that a fair test is one that produces the same number of false negatives, regardless of average group performance. Rather, many experts embrace the notion of a fairly meritocratic test as one that is both valid and unbiased—in being equally predictive of productivity for persons from all groups.²⁷⁸ In fact, there is extensive evidence that commonly used employment screens meet this criterion and 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 from lower-performing groups.²⁷⁹

One approach IOP specialists suggest for dealing with false negatives, as well as performance disparities generally, is subgroup norming.²⁸⁰ By applying different standards or cutoffs for candidates from different groups, race-norming can reduce or eliminate adverse impact through the adjustment of the number of candidates selected. One common method is to employ “dual lists” and to choose the best candidates from each group through a top-down selection

278. This is the widely used “Cleary” model of test fairness, which looks at whether the test has the same predictive validity for different sociodemographic 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, Borneman & Connelly, *supra* note 116, at 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, which would tend to work in blacks’ favor); *see also* Kelman, *supra* note 63, at 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”); Daniel A. Newman, Paul J. Hanges & James L. Outtz, *Racial Groups and Test Fairness, Considering History and Construct Validity*, 62 *AM. PSYCHOLOGIST* 1082, 1082 (2007) (discussing the Cleary criterion of bias in testing); Sackett, De Corte & Lievens, *supra* note 130, at 468 (discussing the Cleary criterion of fairness).

279. *See, e.g.*, Gottfredson, *supra* note 271, at 961 (discussing the unavoidable tradeoff between false negatives and false positives).

280. On subgroup norming in the race context (race-norming), *see, for example*, Dianne C. Brown, *Subgroup Norming: Legitimate Testing Practice or Reverse Discrimination?*, 49 *AM. PSYCHOLOGIST* 927, 927-28 (1994). *See also* Pyburn, Ployhart & Kravitz, *supra* note 103, at 148-49; Sackett & Wilk, *supra* note 105, at 929.

process.²⁸¹ Another is to establish distinct score or performance cutoffs for members of each group or otherwise to relax qualifications for one group relative to another.²⁸²

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.²⁸³ Indeed, in 1989 the National Academy of Sciences (the Academy) issued a report recommending race-conscious score adjustments on the General Aptitude Test Battery (GATB), a test written exam that was widely used by the U.S. Employment Service of the Department of Labor to screen potential government employees.²⁸⁴ The Academy endorsed the GATB as a valid, unbiased predictor of job success across the board, and acknowledged that the test's racially disparate impact "is due not to [the test's] imperfections[,] but to substantial racial differences in the job-related skills, abilities, and knowledge[] [the test] reveal[s]."²⁸⁵ In justifying race-norming of 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.²⁸⁶ It thus

281. See Gottfredson, *supra* note 271, at 957 (discussing "within-group scoring" schemes).

282. See, e.g., Kelman, *supra* note 63, 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 low test scores").

283. See, e.g., Wayne F. Cascio, Rick Jacobs & Jay Silva, *Validity, Utility, and Adverse Impact: Practical Implications From 30 Years of Data*, in *ADVERSE IMPACT*, *supra* note 50, at 271, 282 (noting that "race-norming" is "the single best way to maximize validity and utility simultaneously, while minimizing adverse impact"); Sackett & Wilk, *supra* note 105, at 931 (noting that race-norming best reconciles the competing goals of "achiev[ing] 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 & Roth, *supra* note 128, at 566, 570 (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").

284. COMM. ON THE GEN. APTITUDE TEST BATTERY, *FAIRNESS IN EMPLOYMENT TESTING: VALIDITY GENERALIZATION, MINORITY ISSUES, AND THE GENERAL APTITUDE TEST BATTERY 6-8* (John A. Hartigan & Alexandra K. Wigdor eds., 1989); see also Gottfredson, *supra* note 271, at 955, 956-57.

285. Gottfredson, *supra* note 271, at 955 (citing *ABILITY TESTING: USES, CONSEQUENCES, AND CONTROVERSIES PT. I.* (A.K. Wigdor & W.R. Garner eds., 1982)).

286. See COMM. ON THE GEN. APTITUDE TEST BATTERY, *supra* note 284, at 278.

maintains overall employee quality better than the generalized lowering of requirements that would otherwise be necessary to avoid an adverse impact.²⁸⁷

In further defending its recommendation to race-norm the GATB, the Academy report acknowledged the “false negatives” objection that a single race-blind cutoff excludes too many minorities who could do the job.²⁸⁸ 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, however, lowering the cutoff score for minorities is not without costs. Although previously false negatives are now pushed into the positive category, more minorities with lower scores are hired. This increases the number of false positives from the minority, lower-scoring group. If the race-norming is pronounced, most of the people observed to fail on the job will be minorities.²⁸⁹ 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

287. This argument is similar to that made by Jeff Rosen for allowing affirmative action in education. He asserts that a ban on achieving diversity by lowering college admission standards selectively for lower-performing minority groups will lead to relaxing them across the board, thus compromising the academic quality of institutions generally. See Jeff Rosen, *How I Learned To Love Quotas*, N.Y. TIMES MAG., June 1, 2003, <http://www.nytimes.com/2003/06/01/magazine/how-i-learned-to-love-quotas.html?pagewanted=all&src=pm> (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 wholesale lowering of academic standards that will inevitably accompany pressures to achieve greater diversity); see also John R. Lott, Jr., *Does a Helping Hand Put Others at Risk?: Affirmative Action, Police Departments, and Crime*, 38 ECON. INQUIRY 239, 249 (2000) (“Changing tests to employ a greater percentage of blacks can make it more difficult to screen out lower-quality candidates generally.”).

288. See COMM. ON THE GEN. APTITUDE TEST BATTERY, *supra* note 284, at 277-78.

289. On this point see, for example, Gottfredson, *supra* note 271, at 961, which 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.” Under Gottfredson’s hypothetical numerical example of hiring using race-norming for black job candidates, the number of false negatives falls significantly, “from 50% to 13%, [but] the rate of false positives increases from 17% to 59%.” *Id.* In addition, “two thirds of all poor workers [among hires] would ... be Black, despite Blacks composing less than one third of all workers hired.” *Id.*

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.²⁹⁰

In imposing a less exacting requirement for lower-performing groups while maintaining a higher and different standard for others, race-norming amounts to a form of affirmative action that incorporates a race-conscious double standard. This places the practice at odds with the meritocratic underpinnings of the disparate impact rule. It is also expressly illegal under the Civil Rights Act of 1991.²⁹¹ 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.

Like the problem of adverse impact generally, the problem of false negatives has generated continuing concern, fueling repeated attempts to devise new personnel procedures that more accurately predict job performance while mitigating or even eliminating racially disparate impact. One hope is that the false negative problem can be solved by developing more sensitive and accurate job screens. Unfortunately, attempts to modify selection methods to generate fewer false negatives have not borne fruit. And given current realities, devising more “sensitive” criteria—that is, those that are better able to identify successful workers—is unlikely to reduce racially adverse impacts significantly and may sometimes have the opposite effect. Most criteria that improve predictions for one group will do so across the board. This means that the racial balance of workers selected will not change much, even if the relative number of false negatives among a less qualified group declines. As the literature clearly indicates, the main driver of adverse impacts is the availability of qualified workers, not the fact

290. *See id.* (noting that under race-norming schemes that increase false positives among minorities, “[b]eing Black would ... be strongly associated with being a poor worker”).

291. 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 supra* note 43. 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. *See Ricci v. DeStefano*, 129 S. Ct. 2658, 2677 (2009) (requiring an employer to have “a strong basis in evidence to believe it will be subject to disparate impact liability” before engaging in intentional discrimination to avoid such liability).

that potentially successful workers are not being identified. Numerically, the magnitude of racial differences far outweighs the imperfections in existing predictive instruments.²⁹²

The effort to exclude fewer blacks by achieving greater accuracy in job screening is at bottom another attempt to evade the validity-diversity tradeoff—a project that, as discussed above, has failed so far and, without significant upgrades in minority human capital, is destined to fail. Additionally, there are inherent limits to the accuracy of even the most predictive screens. As one IOP expert has noted, unavoidable imperfections in prediction and measurement mean that “there is clearly nothing an employer can do to design a better selection system.”²⁹³ Multiple factors affect individual performance, but they are not all “knowable before hire”²⁹⁴ and thus are not amenable to accurate identification ahead of time. Moreover, even conceding that current methods sometimes err by excluding some blacks (and other candidates) who could do the job, “the optimal selection system” remains one with substantial disparate impact.²⁹⁵ In sum, given the limits inherent in identifying good workers, efforts to produce more diversity by fiddling with job criteria or creating more predictive screens are 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, current levels of adverse impact are likely to persist.²⁹⁶

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, or staffing by lottery. Managers would assess employees’ job performance directly and discharge individuals whose work falls short.²⁹⁷ This proposal has

292. See *supra* text accompanying note 122.

293. Sackett & Wilk, *supra* note 105, at 934.

294. *Id.*

295. *Id.*

296. See, e.g., WAX, *supra* note 112, at 73-76 (discussing the “skills deficits” among minority workers and attempts to address this problem); see also *supra* text accompanying note 111.

297. See Kelman, *supra* note 63, at 1222, 1245-46; see also Sackett & Wilk, *supra* note 105, at 994 (observing that “if one gave all applicants a job tryout and kept the highest performers

serious drawbacks. Where there are fewer positions than job-seekers, employers will miss out on applicants who are better matched to jobs than those likely to be chosen randomly, with significant costs to efficiency.²⁹⁸ 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 could be accomplished ex ante, some degree of imbalance will inevitably 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. This result would encourage familiar challenges to the fairness of on-the-job assessments and raise suspicions concerning the “discretion inherent in ex post screening systems.”²⁹⁹ In any event, any diversity gains 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 job training would be lost, and the high risk of being fired would deter employees from developing job-related skills.³⁰⁰

Despite its considerable costs, 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 the disparate impact rule is unlikely to achieve greater workforce racial balance, and indeed might move the situation in the opposite direction.³⁰¹ If diversity is the goal, however, there are other ways to achieve it. Directly enacting race-based affirmative action in the

... one would retain considerably more Blacks than would be hired” using screens with greater black-white differences than actual performance differences).

298. See Kelman, *supra* note 63, at 1233 (acknowledging the “additional costs of ex post screening”).

299. *Id.*

300. See, e.g., Douglas O. Staiger & Jonah E. Rockoff, *Searching for Effective Teachers with Imperfect Information*, 24 J. ECON. PERSP. 97, 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).

301. See *supra* text accompanying notes 163-71.

workplace might prove difficult politically and may run afoul of the Equal Protection Clause of the Constitution. But there is a race-neutral alternative. As noted above, there is no general requirement to adopt a system of competitive meritocratic job selection.³⁰² A more diverse government workforce—including fire and police departments—can be achieved by relaxing job requirements or abandoning meritocratic criteria altogether.

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 redesigned 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 validity-diversity 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 specific test results.³⁰³ The opinion does not necessarily cast aspersions on the full range of choices that employers might make to achieve greater diversity. Specifically, it does not rule out the political or administrative choice to give a very easy test or to use criteria that are minimally dependent on cognitive ability.

The validity-diversity tradeoff indicates that any test that significantly reduces adverse impact will also seriously compromise validity.³⁰⁴ The question is then: why bother with a test at all? Diversity might be better achieved, at least for government jobs, by dropping civil service exams entirely in favor of minimal threshold requirements coupled with lotteries or random selection from available applicants. Because many civil service laws and labor contracts mandate competitive exams, a shift to this system would require some political will.³⁰⁵ This method might ultimately prove more efficient, and certainly would be much simpler, than elaborate litigation based on the chimerical, elusive quest for performance-predictive civil service screens with less adverse impact. Adopting

302. See *supra* note 259 and accompanying text.

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

304. See, e.g., *supra* note 289 and accompanying text.

305. See Rutherglen, *supra* note 43, at 113.

this method would require facing up to its consequences and giving up on the fiction that the validity-diversity tradeoff can be overcome. It may impose burdens of personnel supervision, training, and evaluation that are greater than at present. And a more diverse civil service may come at the cost of a less capable one. But if the citizenry considers diversity a top priority, these may be tradeoffs it is willing to make.

CONCLUSION

Under present job market conditions, the validity-diversity 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 group 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 roots, both present and historical.³⁰⁶

The argument for abolishing the disparate impact doctrine is that it does not currently comport with reality and does little to change it. Its goals and assumptions are at odds with what we know about educational disparities, productivity, job performance, and human capital. The four-fifths rule and other elements of the doctrine embody this mismatch between theory and reality. Above all, disparate impact litigation represents a costly, misplaced effort that distracts from the root causes of workforce imbalance and draws

306. See *supra* Part II.B (discussing factors behind skill gaps between whites and minorities).

resources away from the initiatives needed to address it.³⁰⁷ In light of this reality, disparate impact should be altered or abolished.

307. *See, e.g.*, ESPENSHADE & RADFORD, *supra* note 112, at 377 (noting that “improving the academic performance of underrepresented 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 societal 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 the college degree requirement that employers currently impose serve as an elaborate system of screening. 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*, WKLY. STANDARD, Sept. 27, 2010, at 22, 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).