The Dead End of “Disparate Impact”

Amy L. Wax

In 2003, the city of New Haven, Connecticut, sought to fill 15 vacancies for supervisory positions in its fire department by promoting from within. As required by law, the city administered to applicants a written and oral civil-service exam created with the help of personnel experts and fire-department officials. In all, 118 firefighters took the exam; when the test scores came back, it turned out that white applicants had passed at roughly twice the rate of black applicants. If the fire department had followed the city’s civil-service placement rules, no black applicants, and at most two Hispanic applicants, would have been promoted to fill the 15 vacancies.

To avoid this outcome, the city eventually threw out the exam results. Officials were concerned, in part, that the promotions mandated by the test results would prompt a lawsuit by minority applicants. But some of the applicants who had passed the exam protested the city’s decision, claiming they were being denied a fair chance at a promotion for which they had proved themselves qualified. Seventeen successful white test-takers and one successful Hispanic test-taker sued to have the results reinstated; in 2009, their lawsuit reached the United States Supreme Court as *Ricci v. DeStefano*.

At the heart of the *Ricci* case was the doctrine of disparate-impact discrimination, which the Supreme Court first articulated in its 1971 decision in *Griggs v. Duke Power Company*. At issue in *Griggs* was the requirement that employees hired into service jobs at the power company’s facilities had to possess a high-school diploma and achieve a minimum score on an IQ test. The plaintiffs argued that these rules disqualified too many black job applicants, thereby violating Title VII of the Civil

*Amy L. Wax is the Robert Mundheim Professor of Law at the University of Pennsylvania Law School.*
Rights Act of 1964, which prohibits employment discrimination based on race, color, religion, sex, or national origin.

The Supreme Court agreed, ruling that job criteria with an adverse or exclusionary effect on minorities— even if those criteria were “neutral on their face, and even neutral in terms of intent”— could violate the Title VII ban on race discrimination in hiring. The Court further stipulated that employers could escape liability for “disparate impact” only if they demonstrated that their adverse selection practices had “a manifest relationship to the employment in question” or that they were justified by “business necessity.” In examining the criteria for positions at the Duke Power Company, the Court found insufficient evidence to satisfy the job-relatedness defense, and so ruled against the utility.

According to the Griggs Court, the purpose of the newly established disparate-impact rule was to “achieve equality of employment opportunities” by removing “built-in headwinds” and “barriers that had operated in the past” to impede minorities’ workplace advancement. In Griggs and several subsequent cases, the Court has repeatedly stressed that the doctrine’s goal is fully consistent with a competitive meritocracy—one in which businesses remain free to seek out, hire, and promote the best and most productive workers regardless of race and to adopt personnel practices that best achieve that result. The purpose of the rule, according to the Court, is not to enact affirmative-action or group quotas for employment, but simply to eliminate arbitrary disadvantages suffered by minority job-seekers.

Despite this assertion, the development of the Griggs doctrine has proved anything but friendly to meritocratic objectives. Although the Supreme Court has never held that all workplaces must be racially balanced, lower courts and the Equal Employment Opportunity Commission (EEOC), which is charged with administering Title VII, have firmly embraced the presumption that the racial profiles of particular workplaces should reflect the racial composition of the broader population.

This presumption makes no sense, however, unless people from all racial groups are equally qualified for positions at all levels of the economy; only then will every racial group be represented in each occupation exactly in proportion to its share of the broader population. If members of one racial group are more qualified for particular positions than others, they will be hired in disproportionately greater numbers; persons from a less qualified group will be under-represented in those jobs.
The unfortunate reality is that there today exist pronounced differences in worker qualifications by race. That pattern is rooted in historical and social circumstances that may well call for policy reforms and other remedies. But the Court’s disparate-impact doctrine does nothing to change those circumstances or to bring about such reforms; indeed, it stands only to further disadvantage minority groups by setting their members up to underperform and by draining attention and resources away from the true causes of minority under-representation. Moreover, by burdening employers with an arcane tangle of perverse requirements — and by making it virtually impossible for companies to match the most qualified candidates to available jobs — the disparate-impact rule clearly does more harm than good.

These insights have so far had little influence on the law of disparate impact. In its decision in the Ricci case, a 5-4 majority of the Court read the facts narrowly to conclude that New Haven’s civil-service exam was sufficiently related to the jobs in question to survive scrutiny and ultimately sided with the firefighters who had sued to have their scores reinstated. The opinions in that case assumed the continuing vitality of the disparate-impact framework, suggesting that the Court is disinclined to question its decision in Griggs.

But a review of the premises and implications of the disparate-impact doctrine shows that, where the Court has chosen not to act, Congress should step in. The legislative branch should revise Title VII to abolish liability based on adverse impact, at least as applied to race in employment. Doing so would revive the core anti-discrimination principle of the law — a principle that has been undermined by the misguided conflation of equal opportunity and equal results arising from Griggs and its aftermath.

CHOOSING WORKERS

Understanding the perversities of the disparate-impact rule requires a review of the ways in which employers make personnel decisions and of how these practices shape the composition of the work force.

In evaluating candidates for hiring or promotion, companies rely on a panoply of selection criteria, both formal and informal. These include years of education, type of educational experience, and specialized training (collectively known in the field of industrial and organizational psychology, or IOP, as “biodata”), with entry to higher-level jobs often restricted to persons who have obtained high-school, college, or graduate
degrees. Although the use of standardized tests of pure intelligence or cognitive ability has declined in the wake of *Griggs*, many employers still rely on specialized assessments of job knowledge, competence, and skill (including civil-service and professional qualifying exams), as well as on standard personality tests. Many employers also conduct structured or unstructured interviews and solicit letters of recommendation. Recently, prompted by the racially adverse impact of measures of verbal and abstract analysis—areas in which some minority groups underperform—experts have also developed alternative instruments that employ audio or video techniques, or that make use of so-called “assessment center” protocols based on job simulations, real-time problem solving, or actual work samples.

By collecting data on screening methods and correlating the scores on these measures with actual on-the-job ratings, IOP experts have documented the factors that best predict work performance over a wide range of occupations. A strong consensus has emerged, based on hundreds of studies performed over decades, that general cognitive ability—known alternatively as IQ or \( g \)—is the best predictor of work performance for all types of positions, from least to most skilled. (For a more extensive discussion of this evidence and other points raised here, see my 2011 article “Disparate Impact Realism,” 53 *William and Mary Law Review* 621.) Such measures are also “unbiased,” in that the correlation of cognitive ability with job outcomes is independent of a candidate’s race, background, or identity.

The process of demonstrating a link between hiring criteria and subsequent work outcomes is known in IOP as “validation.” And the measured validity of \( g \) is in the range of approximately 0.5 to 0.6 (on a scale that runs from -1 for a total negative correlation to 1 for a complete positive correlation), which represents a relatively powerful social-scientific prediction. Moreover, the usefulness of job-selection criteria is observed to vary with their emphasis on IQ-dependent skills. Criteria that rely more on intelligence are the most effective predictors of occupational success, while screening methods that de-emphasize intelligence in favor of other personal attributes or factors are less accurate in selecting the best workers.

These observations spell trouble for employers’ ability to maximize workforce productivity while meeting legal expectations for diversity. Although psychologists have been interested in personnel selection for some time, the *Griggs* case caused a surge in research designed to help businesses comply
with the new requirements while maximizing the effectiveness of their work forces. The considerable body of social-science evidence generated in the decades since *Griggs*—including extensive research in IOP, labor economics, and educational demography—has established that black workers, and to a lesser extent Hispanic workers, lag behind white and Asian workers in the measures that predict proficiency in a broad range of job-related tasks. The most important of these attributes is general cognitive ability, or IQ.

In describing ethnic disparities in performance on various job-screening tests or criteria, IOP researchers refer to the standardized ethnic-group differences (designated as \(d\)) associated with a given measure of performance. As reported by University of Michigan psychologist Richard Nisbett in his book *Intelligence and How to Get It*, the average measured IQ difference between blacks and whites remains large at this point, standing at a \(d\) value of about one standard deviation from the mean. This pattern has been stable for decades and is well documented in the IOP literature.

Because blacks today have significantly lower average IQ scores than do whites and Asians, they tend to underperform these groups on measures that draw heavily on cognitive ability, including the range of skills that determine academic achievement as well as job success. Racial disparities in proficiency, knowledge, and learning persist at all levels of schooling, with black students consistently observed to enter elementary school, high school, college, graduate school, and professional schools with test scores and grades lower than those of other groups. For instance, scores on a 2009 national test of academic skills (the National Assessment of Educational Progress) revealed that the average black 12th grader reads at the level of the average white eighth grader. Richard Arum, in his recent book *Academically Adrift*, documents that black students, on average, enter college with significantly lower scores and grades than do whites; moreover, he presents evidence that black students learn significantly less in college than do students from other ethnic groups, even when they possess the same entering qualifications.

These disparities do not stop at the schoolhouse door. Rather, they have serious consequences for job placement and, ultimately, for workplace performance. Because the most useful job-selection devices tend to be significantly “\(g\)-loaded”—in that they assign considerable weight to cognitive ability—it is not surprising that minority applicants, and especially black applicants, tend to score worse than whites on job-screening and placement criteria that have been shown
to predict workplace success. These racial disparities are traceable to real, average group differences in what is being predicted: the work-related skills that employees bring to the job. Indeed, direct on-the-job performance ratings also show fairly consistent racial gaps, with existing evidence indicating that black workers lag behind white workers in job performance by a bit less than a third of a standard deviation. A 2006 meta-analysis by Patrick McKay and Michael A. McDaniel in the Journal of Applied Psychology estimated the discrepancy as between 0.24 and 0.39 standard deviation across a spectrum of jobs.

The g-dependency of the best job-selection criteria both contributes to their usefulness and accounts for their adverse impact on protected minorities. The problem is that the power of IQ and IQ-related measures as job screens, combined with lower average scores for minorities, is radically at odds with the Griggs assumption that meritocratic staffing will maximize work-force diversity. In fact, the opposite is true: Meritocracy and diversity are almost always inversely related. The screening devices with the strongest or most “valid” links to job success will tend to generate the least diverse work forces. This reality is well understood by industrial psychologists, who even have a name for it: the diversity-validity tradeoff. Given the current distribution of skill and human capital across groups, it is impossible for most employers to escape this tradeoff.

All told, the tendency of valid job-selection methods to screen out minorities is not an artifact of measurement. Rather, it reflects real deficits in human capital that affect the ability to function in the workplace. The reasons are surely complex and deserving of attention and action, but the bottom line is clear: Too many minority workers, and especially black workers, either are unable to perform many jobs currently available in the economy or lack the ability to compete effectively with persons from other groups. Unfortunately, the existence of these patterns has had, at best, an imperfect influence on the law, which remains largely oblivious to the implications for workplace diversity and for the difficulties employers face in trying to meet minority targets while maintaining work-force quality.

**A N U N W O R K A B L E S T A N D A R D**

Over the years, the Griggs expectation of racially proportionate workplace representation has come to be embodied in a technical standard, the so-called four-fifths rule, which the courts have developed and the
EEOC has embraced. Under this rule, if a workplace employs members of a protected minority group at less than four-fifths, or 80%, of their proportion of the local population, the employer is deemed presumptively liable for a disparate-impact violation. When a lawsuit satisfies this threshold condition, the employer can escape liability only by advancing the defenses identified in the *Griggs* decision. Employers must show that their staffing methods are either necessary to the conduct of their businesses or related to the jobs in question.

This framework has proved virtually unworkable in practice, largely because the doctrine fails to take account of how job success is actually predicted and has resisted awareness of group differences in job-related skills. Judges and plaintiffs’ lawyers frequently disparage the importance of cognitive ability and assume that intelligence is minimally related, or even irrelevant, to many occupations. They often uncritically accept the notion that special traits or “constructs” peculiarly associated with specific types of work are more highly correlated with job performance than is general cognitive ability. This tendency is illustrated by the argument, endorsed by some of the justices in *Ricci*, that the firefighters’ exam was likely flawed for its failure to assess the “command presence” or leadership skills that are central to a fire captain’s job.

More broadly, written civil-service exams are often claimed to be inferior to alternative methods for choosing public-safety personnel like policemen and firefighters, as exemplified by the amicus brief filed in *Ricci* by five individual IOP experts. That brief argued that other screens, including unwritten “assessment center” procedures, could achieve more diversity while selecting equally able or superior employees. In the same vein, political scientists Desmond King and Rogers Smith, in a *New York Times* op-ed in 2011, insisted that workplace diversity could be increased by “adopt[ing] employment tests that are fair and inclusive and do a better job at predicting job performance than many Civil Service exams now do.”

Unfortunately, the notion that the best candidates are identified by selecting for job-peculiar traits rather than for more general abilities, although intuitively appealing, is a product of wishful thinking unsupported by hard data or empirical research. Although non-cognitive capacities make some difference, general intelligence is simply a more important variable for achieving proficiency in a wide range of occupations. This is true even in professions, such as nursing or teaching, that would seem to depend heavily on special non-cognitive skills like compassion.
or patience. Indeed, it is safe to say that cognitive ability better predicts on-the-job performance than does any personality trait or talent that IOP experts have yet identified. Conscientiousness — the personality trait with the strongest documented link to job success — shows a correlation with job performance in the range of about 0.2 to 0.4, in contrast with the significantly higher correlation of 0.5 or more for IQ. Contrary to the Supreme Court’s assumption in Griggs, the comparative power of IQ extends even to relatively uncomplicated positions requiring modest skills, such as clerical or retail work. What this means is that hiring on the basis of intelligence — as opposed to other, non-cognitive personal attributes or talents — will almost always produce better-performing workers.

As for alternatives to written tests, it is revealing that no relevant research findings were adduced (either by New Haven or in the amici briefs for the city’s case) to support the assertion in Ricci that alternative job screens can achieve better results while promoting diversity. Although studies show that written tests of job knowledge are robust predictors of performance in public-safety jobs like policing and firefighting, “assessment center” procedures like those touted in the Ricci briefs rarely achieve comparable validity. And even when such procedures reduce adverse impact, the effects are usually too modest to satisfy legal standards. All told, there is essentially no credible evidence that “better” selection methods — ones that are equally or more valid but have less adverse impact — exist or can be readily devised.

This is not for lack of trying. An entire cottage industry is now devoted to refining personnel selection with the goal of increasing workforce diversity without compromising an employer’s search for the most able employees. This quest has spawned a voluminous literature; the basic approach in nearly every case is to de-emphasize the academic and analytic measures on which minorities lag behind in favor of other abilities that yield smaller or non-existent racial differences.

This work has produced uniformly disappointing results. Except in highly specialized circumstances or in staffing for the least competitive jobs, adopting alternative screening methods that minimize the significance of abilities related to intelligence almost always results in the selection of less capable workers. The reason is simple: The paucity of non-Asian minorities in competitive positions reflects real differences in human capital and skill. Thus changing entry requirements to create a more diverse work force, including scrapping existing civil-service
exams, will generally not result in a more qualified work force. For now, the diversity-validity tradeoff remains the iron law of personnel selection.

The bottom line, therefore, is that most employers who engage in genuinely meritocratic skill-based hiring using a broad range of valid personnel practices will fail to meet the disparate-impact doctrine’s threshold diversity targets. Indeed, as IOP experts Paul Sackett and Jill Ellingson have observed, most employers have no hope of even coming close to satisfying the Griggs four-fifths requirement. Group performance differences, or $d$ values, commonly viewed as small—for example, 0.2, which is far lower than the one standard deviation black-white difference in pure tests of cognitive ability—can produce violations of the four-fifths rule in even modestly competitive hiring situations. This means that businesses that strive to hire the best workers will routinely violate the four-fifths proportionality rule for minority hires and thus expose themselves to potential disparate-impact challenges.

These facts can be demonstrated using specific data that are readily available in the IOP literature. As a general matter, when two groups differ in the distribution of performance on a valid job screen, an employer seeking the best-qualified candidates for a limited number of positions will hire relatively fewer persons from the lower-performing group. This relationship can be represented graphically, as in the figure below.

![](image)

As the figure illustrates, the hires from each group and the corresponding hiring ratios will be a function of both the average score gap between the groups on a screening criterion and the overall number of positions available. As positions become more scarce (and more competitive), the “cutoff” score for hiring increases, and the vertical cutoff
line moves to the right. If the average group score diverges, the distributional curves move farther apart. In both cases, fewer people from the lower-scoring group will be chosen.

Paul Sackett and his colleagues have constructed a table that reflects these patterns. They calculate the precise expected ratio of hires from a minority group relative to a majority group as a function of the selectivity of a job (percent hired relative to the number of job-seekers), as well as the average performance of the two groups on some valid job-screening measure. On the assumption that employers will choose to hire the best candidates, the chart numbers in bold represent the expected ratio of minority to majority hires relative to available applicants. This is expressed as a function of the selectivity of the job (represented at the top of the chart by the percentage of majority candidates hired) and the average racial-group difference (d value) in performance on the job-selection device (represented by the numbers on the left).

The combinations represented below and to the left of the line on the table fall short of the disparate-impact rule’s threshold expectation for diversity (as defined by the four-fifths rule, or a greater than 0.8 ratio of minority to majority hires). Meritocratic hiring for jobs that reflect these combinations exposes employers to potential Title VII lawsuits. Only the numbers above and to the right of the line satisfy the Griggs rule’s workplace-balance requirements. But this “safe harbor” covers only situations in which average group differences in measured performance are relatively small (say 0.1 or 0.2 standard deviation), or where almost all job-seekers are offered jobs. In other words, it is possible for employers who hire meritocratically to escape disparate-impact lawsuits—but only in unusual circumstances.

The calculations on the chart also demonstrate another key defect in the disparate-impact doctrine. In a meritocratic system, the expected ratio of hires from groups that differ in measured job-related abilities is not fixed. Rather, it will vary with the job’s skill demands and the scarcity of positions relative to applicants. By imposing a uniform workplace-balance requirement for all positions, the four-fifths rule fails to reflect this sliding scale. Even apart from other serious flaws, this alone is a major shortcoming of existing law.

**Proving Innocence**

For purposes of considering the law’s effects, the bottom line is that employers seeking to maximize job productivity will routinely generate a
## MINORITY GROUP SELECTION RATIOS AND FOUR-FIFTHS RATIOS

<table>
<thead>
<tr>
<th>Standardized group difference ($d$)</th>
<th>1%</th>
<th>5%</th>
<th>12%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>90%</th>
<th>95%</th>
<th>99%</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>0.1</td>
<td>0.08</td>
<td>0.41</td>
<td>0.84</td>
<td>2.21</td>
<td>4.60</td>
<td>7.16</td>
<td>8.81</td>
<td>9.38</td>
<td>9.87</td>
</tr>
<tr>
<td></td>
<td>0.80</td>
<td>0.68</td>
<td>0.24</td>
<td>0.88</td>
<td>9.23</td>
<td>9.45</td>
<td>9.58</td>
<td>9.90</td>
<td>9.99</td>
</tr>
<tr>
<td>0.2</td>
<td>0.06</td>
<td>0.33</td>
<td>0.69</td>
<td>1.92</td>
<td>4.21</td>
<td>6.81</td>
<td>8.60</td>
<td>9.25</td>
<td>9.83</td>
</tr>
<tr>
<td></td>
<td>0.60</td>
<td>0.66</td>
<td>0.69</td>
<td>0.77</td>
<td>0.84</td>
<td>0.91</td>
<td>0.96</td>
<td>0.97</td>
<td>0.99</td>
</tr>
<tr>
<td>0.3</td>
<td>0.04</td>
<td>0.26</td>
<td>0.57</td>
<td>1.66</td>
<td>3.82</td>
<td>6.44</td>
<td>8.37</td>
<td>9.10</td>
<td>9.78</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>0.32</td>
<td>0.38</td>
<td>0.48</td>
<td>0.65</td>
<td>0.76</td>
<td>0.87</td>
<td>0.92</td>
<td>0.98</td>
</tr>
<tr>
<td>0.4</td>
<td>0.03</td>
<td>0.21</td>
<td>0.46</td>
<td>1.42</td>
<td>3.45</td>
<td>5.06</td>
<td>6.81</td>
<td>8.93</td>
<td>9.73</td>
</tr>
<tr>
<td></td>
<td>0.20</td>
<td>0.42</td>
<td>0.46</td>
<td>0.57</td>
<td>0.69</td>
<td>0.82</td>
<td>0.90</td>
<td>0.94</td>
<td>0.98</td>
</tr>
<tr>
<td>0.5</td>
<td>0.02</td>
<td>0.16</td>
<td>0.38</td>
<td>1.21</td>
<td>5.09</td>
<td>7.68</td>
<td>9.28</td>
<td>9.73</td>
<td>9.96</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>0.32</td>
<td>0.38</td>
<td>0.48</td>
<td>0.65</td>
<td>0.76</td>
<td>0.87</td>
<td>0.92</td>
<td>0.98</td>
</tr>
<tr>
<td>0.6</td>
<td>0.02</td>
<td>0.13</td>
<td>0.30</td>
<td>1.02</td>
<td>3.24</td>
<td>5.28</td>
<td>7.52</td>
<td>8.51</td>
<td>9.57</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>0.26</td>
<td>0.20</td>
<td>0.41</td>
<td>0.55</td>
<td>0.70</td>
<td>0.84</td>
<td>0.90</td>
<td>0.97</td>
</tr>
<tr>
<td>0.7</td>
<td>0.01</td>
<td>0.10</td>
<td>0.24</td>
<td>0.85</td>
<td>2.42</td>
<td>4.88</td>
<td>7.19</td>
<td>8.26</td>
<td>9.47</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>0.20</td>
<td>0.14</td>
<td>0.34</td>
<td>0.48</td>
<td>0.65</td>
<td>0.80</td>
<td>0.87</td>
<td>0.96</td>
</tr>
<tr>
<td>0.8</td>
<td>0.01</td>
<td>0.07</td>
<td>0.19</td>
<td>0.71</td>
<td>2.12</td>
<td>4.48</td>
<td>6.84</td>
<td>8.20</td>
<td>9.36</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>0.14</td>
<td>0.19</td>
<td>0.28</td>
<td>0.42</td>
<td>0.60</td>
<td>0.76</td>
<td>0.84</td>
<td>0.95</td>
</tr>
<tr>
<td>0.9</td>
<td>0.01</td>
<td>0.06</td>
<td>0.15</td>
<td>0.58</td>
<td>1.84</td>
<td>4.29</td>
<td>6.48</td>
<td>7.70</td>
<td>9.22</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>0.22</td>
<td>0.15</td>
<td>0.37</td>
<td>0.54</td>
<td>0.72</td>
<td>0.81</td>
<td>0.93</td>
<td>1.00</td>
</tr>
<tr>
<td>1.0</td>
<td>0.00</td>
<td>0.04</td>
<td>0.01</td>
<td>0.04</td>
<td>0.07</td>
<td>0.15</td>
<td>0.37</td>
<td>0.61</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.08</td>
<td>0.11</td>
<td>0.19</td>
<td>0.32</td>
<td>0.49</td>
<td>0.68</td>
<td>0.78</td>
<td>0.91</td>
</tr>
<tr>
<td>1.1</td>
<td>0.00</td>
<td>0.03</td>
<td>0.09</td>
<td>0.03</td>
<td>0.13</td>
<td>0.34</td>
<td>0.57</td>
<td>0.80</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.06</td>
<td>0.09</td>
<td>0.15</td>
<td>0.27</td>
<td>0.45</td>
<td>0.65</td>
<td>0.74</td>
<td>0.90</td>
</tr>
<tr>
<td>1.2</td>
<td>0.00</td>
<td>0.02</td>
<td>0.07</td>
<td>0.13</td>
<td>0.15</td>
<td>0.25</td>
<td>0.48</td>
<td>0.67</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.04</td>
<td>0.07</td>
<td>0.12</td>
<td>0.23</td>
<td>0.40</td>
<td>0.59</td>
<td>0.71</td>
<td>0.88</td>
</tr>
<tr>
<td>1.3</td>
<td>0.00</td>
<td>0.02</td>
<td>0.05</td>
<td>0.24</td>
<td>0.09</td>
<td>0.14</td>
<td>0.49</td>
<td>0.63</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.04</td>
<td>0.06</td>
<td>0.10</td>
<td>0.19</td>
<td>0.35</td>
<td>0.55</td>
<td>0.67</td>
<td>0.85</td>
</tr>
<tr>
<td>1.4</td>
<td>0.00</td>
<td>0.01</td>
<td>0.04</td>
<td>0.01</td>
<td>0.08</td>
<td>0.13</td>
<td>0.23</td>
<td>0.42</td>
<td>0.59</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.02</td>
<td>0.02</td>
<td>0.00</td>
<td>0.01</td>
<td>0.02</td>
<td>0.03</td>
<td>0.04</td>
<td>0.05</td>
</tr>
<tr>
<td>1.5</td>
<td>0.00</td>
<td>0.02</td>
<td>0.03</td>
<td>0.06</td>
<td>0.13</td>
<td>0.27</td>
<td>0.46</td>
<td>0.59</td>
<td>0.80</td>
</tr>
</tbody>
</table>

*Selection ratio = number of applicants hired/number of applicants applied. Per cell, two values are given. The first value refers to the minority group selection ratio. The second value in bold represents the four-fifths ratio (i.e., the minority group selection ratio/majority group selection ratio). Tabled values in bold less than .80 represent scenarios that violate the four-fifths rule.

work force that falls short of the stringent four-fifths standard. And this, in turn, means that their staffing methods will be vulnerable to legal challenge. Of course, the law of disparate impact does give employers a defense: They must prove the “business necessity” or “job-relatedness” of their practices. Unfortunately, supporting this defense is always an uphill battle, even for employers who have done nothing wrong. The pertinent rules surrounding the employer’s responsibilities are a morass of unsettled standards and ambiguous requirements that are often impossible to meet.

The courts and the EEOC have recognized at least three methods by which employers can justify selection procedures that produce racially lopsided results. All three involve variations on the process of validation, or demonstrating a correlation between a qualifying criterion for a particular type of work and the ability to perform the job.

“Content” validation requires establishing a manifest relationship or plausible match between a screening assessment and key job tasks. This form of validation relies largely on individual judgments rather than on showing a formal or statistical link between a job screen and subsequent productivity. For example, the relationship between proficiency in reading and being hired as an English teacher appears obvious on its face, and many courts will not question that occupational requirement. In contrast, the so-called “construct” and “criterion” validation methods require an employer to demonstrate a quantitative relationship between a job-selection method and measures of specific job-related skills (in the case of construct validation) or workers’ actual on-the-job performance (in the case of criterion validation). Criterion validation, which links a screen directly to job ratings, is considered the most rigorous and exacting “gold standard” and is the focus of the most intensive IOP research.

Complicated methodological challenges stand in the way of meeting these validation requirements. One important problem is that of “range restriction,” which stems from limitations in the populations of workers for whom pertinent performance data are available. Range restriction is reflected in the oft-heard, but frequently misleading, contention that scores on various qualifying tests are largely irrelevant to the success of individuals within a given occupation.

One example is the case of licensing exams for doctors. Because admission to most medical schools requires high scores on IQ-related measures — such as pre-admissions tests and undergraduate grades — enrolled students, especially in selective schools, are drawn from
a small, elite band of the population. Since doctors are relatively similar in intellectual ability compared to people at large, attributes such as drive, focus, work ethic, and compassion will appear to loom larger than scores on licensing exams in predicting success in post-graduate training and beyond within that restricted population. But if medical students were unscreened and admitted by lottery—and if intelligence varied as widely among them as it does among the general population—licensing-exam scores would show much higher measured correlations with professional success.

In the same vein, people who choose to work in particular occupations or who seek specific types of jobs are usually drawn from a fairly narrow slice of the work force and resemble one another more closely than do people chosen at random. Moreover, performance correlations can be observed only for candidates who are actually hired to do a job, which further restricts the available evidence. The effect of this clustering is to reduce the apparent measured correlation between intelligence-related criteria and success. IOP experts have developed statistical methods for correcting for these and related distortions in the actual data, and their investigations over a broad range of situations have enabled them to get fairly reliable results. But individual employers are not research psychologists. Given the limitations of the information that is often available about the jobs in question in particular lawsuits—and given the expense and difficulty of gathering and analyzing the appropriate evidence—the employer’s task of validating specific staffing methods in order to mount a job-relatedness defense can be difficult, expensive, and infeasible in practice.

What this means for employers is that disparate-impact litigation is almost always a costly, risky, and error-prone venture. Uncertainties and ambiguities abound. Although validation methods vary widely in their burdens and feasibility, the EEOC and the courts have never clarified which are appropriate in which cases. Judges retain considerable discretion, with some deferring to employers and others demanding rigorous data. Regardless of the standard applied, establishing job-relatedness is always expensive and sometimes methodologically impossible, and the outcome is seldom predictable. Parties are drawn into obscure and specialized inquiries culminating in protracted and costly “battles of the experts.” The voluminous record in Ricci, the New Haven firefighters’ case, contains thousands of pages of fact-finding and expert testimony, much of it devoted to abstruse questions surrounding the predictive validity of the firefighters’ exam.
Other critical doctrinal questions also remain unresolved, which only adds to the uncertainty. For instance, the Supreme Court has never definitively identified the baseline population against which the proportions of particular racial groups are to be measured. Is it all work-eligible adults? Only those who actually apply for the job in question? Or those with specific, designated job qualifications? A broader baseline population draws in a greater number of less-qualified minorities, making it more difficult for employers to hire on a meritocratic basis without running afoul of racial-proportionality rules. But restricting the population pool has been deemed by some courts to be inconsistent with the rationale of the disparate-impact rule, which questions the need for qualifications of any kind. At the same time, the courts have also failed to agree on when and whether employers being sued for their job-selection criteria are obligated to identify alternative job-screening methods that would have less of an adverse impact on minorities.

Although few disparate-impact lawsuits are actually filed, and though defendants win most of those that are, the specter of disparate-impact lawsuits nonetheless dangles like a Sword of Damocles over employers’ heads. The prospect of onerous, unpredictable, and protracted litigation provides employers with a strong incentive to find some way to avoid being sued. One temptation is to satisfy diversity targets by relaxing personnel-selection standards across the board. Firms can also switch to more haphazard staffing methods that tend to obscure informal affirmative action or other race-conscious practices. Some of these tactics pose the risk of yielding a less effective work force, while others (such as the use of race-based criteria) are legally suspect or even expressly forbidden under Title VII.

But even if employers refrain from these tricks, the disparate-impact rule still does more harm than good. The notion that the job market is riddled with capricious hurdles and arbitrary job requirements is, in our day, mostly a myth. Commonly used selection devices do tend to predict future performance, and most employers are sincerely interested in matching workers to jobs. Given that the vast majority of standard screening methods are largely “valid,” disparate-impact challenges will increase diversity only if the law is misapplied and employers are forced to drop legitimate job requirements in favor of those that are less meritocratic. By the same token, if most lawsuits are resolved correctly and fairly—recognizing that firms’ selection criteria are in fact valid and essential to hiring—employers will almost always prevail in court. This
The Dead End of “Disparate Impact”

means that workplace diversity will not increase as a result of Title VII lawsuits, and that existing racial imbalances in the work force will persist. Indeed, without significant changes in the distribution of human capital—which is where our attention should really be focused—the law will accomplish very little.

Of course, workplace diversity will increase if employers respond to the threat of litigation by adopting covert affirmative action or race-conscious selection methods—a result favored by those who prioritize a racially balanced work force. But using disparate-impact lawsuits to accomplish this result is perverse: Since the inception of the disparate-impact doctrine, the courts have consistently stated that it should not serve as just another form of affirmative action. Rather, as the Supreme Court made clear in Griggs, the doctrine’s goal is to establish a neutral, color-blind meritocracy in which employers are free to impose uniform, work-related requirements.

The problem, of course, is that this ideal is at odds with the race-conscious double standards that Griggs has created in practice. The law was never intended to correct underlying disparities in human skill and capital, which employers are not in a position to address. It was instead premised on the implicit assumption that such race-based disparities do not exist. But that assumption was mistaken; these disparities are a fact of life. To make them disappear through hiring practices is to require affirmative action. And if affirmative action is the objective, it should be enacted explicitly and directly—not accomplished through subterfuge using a legal device that was never intended for that purpose.

A COSTLY DISTRACTION

The divide between the assumptions underlying the disparate-impact rule and the realities of the American work force has not only meant that the rule is difficult to follow, but also that its enforcement is selective and erratic.

The federal government has long used tests of cognitive ability—including the Armed Forces Qualifying Test—to determine admission to the military and assignments within it. A report recently released by the liberal non-profit Education Trust reveals that the AFQT has a pronounced disparate impact by race, with 39% of otherwise qualified black applicants scoring below the cutoff as compared to 16% of white applicants. Even starker gaps prevail among top scorers eligible
for special training and elite positions. The sheer number of blacks who apply to serve in the military masks these disparities, which persist without serious challenge.

Likewise, the criteria for entry into the elite professions have long generated a pronounced racial imbalance. Ironically, the dissenting Supreme Court justices who voted to invalidate the firefighters’ exam in *Ricci v. DeStefano* routinely hire law clerks with perfect grades, top class ranks, and law-review board positions. The racially disparate effects of those formidable hurdles are accepted without question. In short, barriers to entry abound in the fields of business, law, medicine, technology, academia, and finance, and all generate racially adverse effects. The world is full of disparate-impact lawsuits waiting to happen—but very few ever do.

None of this is to say that existing personnel practices are unproblematic. As Stanford law professor Mark Kelman showed in an influential 1991 *Harvard Law Review* article on ability testing, no screening device can perfectly predict a job applicant’s eventual workplace performance. Many valid qualifying measures generate a disproportionate number of “false negatives”—that is, persons who could perform adequately but are rejected because of low test scores—among groups with lower average performance. Indeed, larger numbers of false negatives among underachieving groups is endemic to the very structure of top-down competition, because more candidates from those groups will fall below a chosen cutoff. At any given error rate, more will thus be eliminated erroneously. Although the IOP community has extensively debated how to deal with this problem, there is no easy solution. As some experts have noted, schemes that tend to reduce the number of false negatives from low-scoring groups will tend to generate an excessive number of “false positive” workers from the same cohorts—people who are hired but perform inadequately. A system in which most underperforming employees are disadvantaged minorities is clearly undesirable, so reducing false negatives at the cost of increasing false positives may not be worth the price.

There is no such thing as foolproof personnel selection. All criteria are imperfect and all systems come with unavoidable tradeoffs. Given the best available techniques identified to date, the diversity-validity tradeoff prevails, which means that the most predictive job-selection devices tend to generate the most adverse impacts among lower-scoring
minorities. There is currently no known way around this dilemma, which reflects existing distributions of human capital that are rooted in complex historical and social circumstances. As James Outtz and Daniel Newman, two prominent IOP experts, noted in a 2010 volume on adverse impact, “there are many realistic disadvantages that distinguish racial subgroups, and these disadvantages logically have some implications for job performance.”

The unfortunate reality is that too few minorities possess the qualifications and training needed to compete effectively for the jobs available in our economy. Disparate-impact litigation thus represents a costly, misplaced effort that fails to address the true causes of existing workplace imbalances and draws resources away from the initiatives needed to correct them.

As applied to race and employment, the disparate-impact rule should be repealed by Congress or abolished by the courts. The Supreme Court’s vast expansion of the concept of discrimination in Griggs was a mistake: well intended, perhaps, but not well suited to realities of America’s work force and society. The civil-rights laws should return to their original purpose, which was to eliminate double standards and adverse treatment targeted at disfavored groups. After all, Title VII, by its own terms, forbids discrimination “because of” race, gender, and other characteristics—and says nothing about eventual outcomes.

In the sphere of employment, the key questions are: “Why do some people compete more effectively than others for jobs and social rewards?” and “What can be done about it?” These questions are complicated and pressing, and the law of disparate impact does nothing to address them. It in fact only distracts us from finding urgently needed answers.