SCALING THE WALL AND RUNNING THE MILE: THE
ROLE OF PHYSICAL-SELECTION PROCEDURES
IN THE DISPARATE IMPACT NARRATIVE

Yiyang Wu†

Since the Supreme Court’s landmark decision in Dothard v. Rawlinson in 1977, gender-based disparate impact litigation has been limited in scope, but there remains room for growth. This Comment focuses on one particularly successful subset of gender-based disparate impact cases, physical-selection procedures. An examination of these decisions shows that plaintiffs have faced an uphill battle in combating unfounded assumptions, both in establishing a prima facie case as well as in rebutting the affirmative defense. Indeed, some lower courts have relied on arguments that are inconsistent with the Supreme Court case law as it has progressed since Griggs v. Duke Power Co.

At the same time, the success of physical-selection procedure cases offers hope for expansion going forward. By contextualizing an industry’s practices, referring to narratives of female applicants, and providing examples of reasonable alternatives, advocates have succeeded in positively framing their arguments in a manner that factfinders are likely to welcome. In doing so, advocates can help reclaim the ideals of Title VII and the disparate impact movement.

† Articles Editor, University of Pennsylvania Law Review, Volume 160. J.D. Candidate, 2012, University of Pennsylvania Law School; A.B., 2007, Harvard University. I would like to thank Professor Serena Mayeri for her invaluable insights in developing this Comment. I owe additional gratitude to the editors and board of the University of Pennsylvania Law Review. Finally, I am grateful for the support of my family and Benjamin Cheng, my soon-to-be family.
INTRODUCTION ................................................................................... 1196
I. THE STANDARD DISPARATE IMPACT DOCTRINE ............................. 1199
   A. Griggs and Its Progeny ............................................................ 1199
   B. Wards Cove and the Civil Rights Act of 1991 ......................... 1202
   C. Ricci v. DeStefano: An Update on the Disparate Impact Doctrine........................................................................ 1204
II. THE GENDER-BASED DISPARATE IMPACT DOCTRINE ....................... 1207
   A. Dothard v. Rawlinson: Limits to Its Strengths ......................... 1207
   B. Gender-Based Disparate Impact in Lower Federal and State Courts After Dothard ................................................. 1210
III. A SUBSET OF SUCCESS: PHYSICAL-SELECTION
    PROCEDURE CASES ........................................................................ 1212
    A. Absence of Disparate Impact Statistics .................................... 1214
    B. Failure to Train ....................................................................... 1216
    C. The Catchall Business Necessity Defense ............................... 1221
    D. “Strength Is Not Always Everything,” and “More Is Not Always Better” .............................................................. 1226
IV. BEST PRACTICES FOR GENDER-BASED DISPARATE IMPACT CASES ........................................................................ 1228
    A. Contextualize the Test and the Industry .................................... 1229
    B. Appeal to the Factfinder’s Capacity to Dismiss Unfounded Stereotypes ............................................................... 1231
    C. Address Valid Safety Concerns by Providing Concrete Reasonable Alternatives ......................................................... 1234
    D. The Doctrine Shall Set You Free ............................................... 1236
CONCLUSION ....................................................................................... 1237

INTRODUCTION

The merits and potential of disparate impact theory have been enthusiastically championed and critically debated\(^1\) since its Supreme

---

\(^1\) See, e.g., Lara M. Gardner, A Step Toward True Equality in the Workplace: Requiring Employer Accommodation for Breastfeeding Women, 17 Wis. Women’s L.J. 259, 280 (2002) (contending that disparate impact theory could be used to “make a prima facie showing of discrimination based on sex” for employers’ failure to accommodate breastfeeding women in the workplace); Michelle A. Travis, Equality in the Virtual Workplace, 24 Berkeley J. Emp. & Lab. L. 283, 341-73 (2003) (proposing the use of disparate impact theory to equalize women’s access to telecommuting options); see also Reva Siegel, Why Equal Protection No Longer Protects: The Evolving Forms of Status-Enforcing State Action, 49 Stan. L. Rev. 1111, 1145 (1997) (arguing that disparate impact theory litigation could “move the nation closer to disestablishing . . . gender stratification than current constitutional doctrines now do”).
Court inception in 1971. Indeed, the Court’s most recent decision, *Ricci v. DeStefano*, in which the Court considered the constitutionality of an employer’s reaction to allegations of a disparate impact on African American candidates as a result of written promotion exams, has reinvigorated the discourse but provided few answers. Most of the disparate impact debate after *Ricci* has understandably centered on the future of race-based claims. Professors Mario Barnes, Erwin Chemerinsky, and Trina Jones, for example, have argued that the *Ricci* decision signaled the Court’s intent to enter a realm of “post-race equal protection.” Yet this race-focused discourse obfuscates the fact that female disparate impact claimants face disparities at least as dramatic as those faced by African American claimants. And almost no attention has been paid after *Ricci* to how gender-based disparate impact cases have fared or what the doctrine’s future prospects are for success.

In this Comment, I focus on one increasingly prevalent subset of gender-based disparate impact litigation: physical-selection procedure cases. Physical-selection procedure litigation typically features female plaintiffs who challenge the use of certain employer-instituted tests, such as wall climbs and timed mile runs, which act as barriers to female entry into traditionally male-dominated jobs. In deciding these cases, courts tend to rely on a number of unfounded assumptions, particularly with regard to the relationship between strength and safety, and make

---

2 See *Griggs v. Duke Power Co.*, 401 U.S. 424, 430 (1971) (“Under the [Civil Rights] Act, practices, procedures, or tests neutral on their face, and even neutral in terms of intent, cannot be maintained if they operate to ‘freeze’ the status quo of prior discriminatory employment practices.”).


5 Barnes, Chemerinsky, & Jones, *supra* note 4, at 994-95.

6 To my knowledge, no literature has yet explored how *Ricci* has affected gender-based disparate impact claims. For one example of a district court’s attempt to decide a gender-based disparate impact claim after *Ricci*, see United States v. Massachusetts, 781 F. Supp. 2d 1, 3-4, 11-21 (D. Mass. 2011).
assertions that are inconsistent with disparate impact theory as articulated in Supreme Court opinions from *Griggs v. Duke Power Co.*\(^7\) onward. However, plaintiffs who confront these flawed assumptions have been able to achieve success in a number of lower courts.\(^8\) In this Comment, I identify the roadblocks commonly facing physical-selection procedure claims and suggest best practices that might enable claimants to overcome them.

Part I of this Comment describes the standard (mostly race-based) disparate impact doctrine as it has evolved since *Griggs*.\(^9\) Part II examines the gender-based disparate impact movement, demonstrating that its expansion beyond height and weight requirements has, for the most part, been limited. In marked contrast, litigation surrounding physical-selection procedures has had atypical success in lower courts. Parts III and IV focus on lower court decisions in physical-selection procedure cases and identify patterns that help to explain why certain plaintiffs succeed where others fail. I argue that successful plaintiffs have been able to bypass certain common obstacles, including arguments that females must train for their examinations and arguments that physical-selection procedures are a business necessity. In unsuccessful physical-selection procedure cases, courts have misinterpreted Supreme Court precedent and relied instead on unproven assumptions that ignore the reality of the employer’s actual needs.

I conclude in Part IV with suggestions that may enable future gender-based disparate impact advocates to find greater success. Advocates should first contextualize the industry and the particular employment practice. Second, advocates should attempt to humanize the consequences of these questionable practices and misguided assumptions by referring to the specific narratives of applicants. Third, advocates should provide concrete examples of reasonable alternatives to the challenged procedure. Finally, advocates should consistently ground their arguments in the disparate impact theory as put forth in *Griggs*. In doing so, advocates will further not only the progress of gender-based causes but also the disparate impact movement at large.

---

\(^7\) 401 U.S. 424 (1971). For discussion of the *Griggs* case, see infra Section I.A.

\(^8\) For a list of these cases, see infra note 82.

\(^9\) The majority of this narration focuses on race-based disparate impact doctrine for one simple reason: the Court has granted certiorari almost exclusively in race-based disparate impact cases.
I. THE STANDARD DISPARATE IMPACT DOCTRINE

To understand the current state of disparate impact litigation, it is important to trace its historical roots from its inception. By examining Title VII of the Civil Right Acts of 1964, the Griggs progeny, and the Civil Rights Act of 1991, I aim to make two points: first, the Court intended disparate impact litigation to serve as a counterpart to disparate treatment cases; second, the Court’s analysis of disparate impact theory has overwhelmingly focused on race-based claims. Thus, a separate analysis of gender-based disparate impact case law is needed.

A. Griggs and Its Progeny

In Griggs v. Duke Power Co., the Supreme Court held that Title VII prohibits neutral employment policies that have a disparate impact on African American plaintiffs without a business necessity justification. In Griggs, a power plant required its employees to pass two screening...
tests and have a high school diploma in order to obtain a position outside of the labor department.\(^{15}\) The plaintiffs were able to demonstrate that this policy had a disproportionately adverse impact on African Americans.\(^{16}\) The Court held that the “absence of discriminatory intent does not redeem employment procedures or testing mechanisms that operate as ‘built-in headwinds’ for minority groups and are unrelated to measuring job capability.”\(^{17}\) Because the power plant could not show that the test served an overriding business necessity, the Court ultimately held that the employer’s selection procedures violated Title VII.\(^{18}\)

Relying upon the general language in \textit{Griggs}, the Court decided several disparate impact cases clarifying the doctrine in a decidedly pro-plaintiff manner. In \textit{Albemarle Paper Co. v. Moody}, the Court held that an employer, in order to comply with Title VII, must demonstrate by “professionally acceptable methods” that its discriminatory tests are “predictive of or significantly correlated with important elements of work behavior which comprise or are relevant to the job.”\(^{19}\) Although the employer in \textit{Albemarle}, perhaps taking a lesson from its counterpart in \textit{Griggs}, had conducted a half-day study to prove that its general ability tests were correlated with job performance, the Court noted that the study results were not statistically significant, were highly subjective, and involved an unrepresentative sample group.\(^{20}\) The Court also created a surrebuttal to the \textit{Griggs} framework, holding that plaintiffs could defeat the business necessity defense by showing “that other tests

---

\(^{15}\) \textit{Id.} at 427-28.

\(^{16}\) \textit{Id.} at 431-33, 436. While the Supreme Court noted that the plaintiffs were more likely to fail aptitude tests because they had received “inferior education in segregated schools,” \textit{id.} at 430, such a “present effects of past discrimination” argument has been criticized by at least one scholar. See Steven L. Willborn, \textit{The Disparate Impact Model of Discrimination: Theory and Limits}, 34 Am. U. L. Rev. 799, 804 (1985) (“Under this [past discrimination] theory, . . . a disparate impact opens the door to litigation over ancient, but preserved, race-based discrimination.”).

\(^{17}\) \textit{Griggs}, 401 U.S. at 432.

\(^{18}\) See \textit{id.} at 431 (“[N]either the high school completion requirement nor the general intelligence test is shown to bear a demonstrable relationship to successful performance of the jobs for which it was used.”).


\(^{20}\) \textit{Albemarle}, 422 U.S. at 429-35.
2012] Scaling the Wall and Running the Mile 1201

or selection devices, without a similarly undesirable racial effect, would also serve the employer’s legitimate interest.\(^\text{21}\)

\textit{Dothard v. Rawlinson}, which prohibited height and weight requirements for employment within prison systems that “had a discriminatory impact on women applicants,”\(^\text{22}\) is the only case in which the Court has ever considered the merits of a gender-based disparate impact claim. The Court in \textit{Dothard} set forth the following business necessity test: first, a defendant must articulate a quality “essential to effective job performance,” and second, it must prove that the challenged practice accurately assesses that quality.\(^\text{23}\) The Court also explicitly held that an employer could be liable for using tests that had a disparate impact across gender lines.\(^\text{24}\) However, as I argue in Section II.A, the Court failed to take advantage of the opportunity \textit{Dothard} presented to strengthen the gender-based disparate impact doctrine.

The next set of decisions, \textit{Connecticut v. Teal} and \textit{Watson v. Fort Worth Bank & Trust}, contributed rules to the emerging doctrine but also made moves toward fashioning a more limited cause of action. In \textit{Teal}, the employer attempted to compensate for a discriminatory selection procedure by promoting African Americans and ensuring that the overall result of the process would be an “appropriate racial balance.”\(^\text{25}\) The Court rejected the employer’s contention that its bottom-line result could be a complete defense to a disparate impact claim and required that each specific procedure undergo analysis.\(^\text{26}\) Yet by


\(^{22}\) 433 U.S. 321, 331 (1977). Because Part II examines \textit{Dothard} in greater detail, the description of the case here only addresses that which is relevant to a general overview of the disparate impact doctrine.

\(^{23}\) Id. at 331-32; \textit{see also id.} at 332 (“If the job-related quality that the [employers] identify is bona fide, their purpose could be achieved by adopting and validating a test for applicants that measures strength directly.”).

\(^{24}\) \textit{See id.} at 328-29 (stating that the standard for examining a gender-based disparate impact claim is the same as that used in \textit{Griggs} and \textit{Albemarle} to assess race-based claims).


\(^{26}\) Id. at 442, 452-56.
directing its analysis to discrete elements of the employer’s procedure, the Court’s decision foreshadowed its later efforts to limit disparate impact claims to those in which a specific practice is proved to have caused the alleged wrong.\(^{27}\)

In *Watson*, the Court ruled that subjective evaluation processes, such as interviews, also fall under the purview of the disparate impact doctrine.\(^{28}\) However, the Court also limited the doctrine’s potential for effectuating change by increasing the plaintiff’s burden in making a prima facie case and reducing the defendant’s burden for proving a business necessity defense.\(^{29}\) Taken together, these cases represent a shift in the Court’s posture from pro-plaintiff to pro-defendant—a shift that culminated in *Wards Cove* and prompted Congress to step in and limit these cases’ impact.

### B. Wards Cove and the Civil Rights Act of 1991

The Court strengthened its pro-defendant posture in *Teal* and *Watson* even further in *Wards Cove*, and, in so doing, triggered a congressional response in the form of the Civil Rights Act of 1991. The facts underlying the plaintiffs’ case in *Wards Cove* were similar to previous Title VII cases: a group of cannery workers alleged that a combination of hiring and promotion practices had resulted in a disparate impact against minorities.\(^{30}\) However, for the first time in a disparate impact case, the Court decided in favor of the employer-defendant.\(^{31}\) In finding for the employer, the Court held that the workers’ prima facie case had to identify a specific employment practice that caused the dispar-

---

\(^{27}\) Lye, * supra* note 19, at 329-30.

\(^{28}\) See *Watson v. Fort Worth Bank & Trust*, 487 U.S. 977, 990-91 (1988) (“If an employer’s undisciplined system of subjective decisionmaking has precisely the same effects as a system pervaded by impermissible intentional discrimination, it is difficult to see why Title VII’s proscription against discriminatory actions should not apply.”).

\(^{29}\) See *id.* at 994, 998 (holding that a plaintiff must pinpoint a “specific employment practice” that results in a disparate impact, while a defendant need only show some relationship between its employment selection criteria and the job itself, not “particular criteria [that] predict on-the-job performance”).


\(^{31}\) *Id.* at 655, 661.
In addition, the Court established a relaxed affirmative defense standard: an employer has an affirmative defense as long as it establishes that the practice is reasonably related to the employer’s business justification. As the Court emphasized, “[T]here is no requirement that the challenged practice be ‘essential’ or ‘indispensable’ to the employer’s business for it to pass muster.”

At least partially as a response to Wards Cove, the 102nd Congress passed the Civil Rights Act of 1991. For the first time, Congress provided an explicit statutory basis for disparate impact litigation, stating in no uncertain terms that disparate impact claims have a legitimate place within Title VII. The Act confirmed that a plaintiff’s prima facie burden includes establishing that a specific employment practice, or group of employment practices, caused a disparate impact on a protected class. At the same time, the Act also rejected the “reasoned review” standard for an employer’s business necessity established in Wards Cove. The Act declared that Wards Cove had “weakened the scope and effectiveness of Federal civil rights protections.”

Congress’s insistence on a pre–Wards Cove interpretation of disparate impact helped establish the legitimacy of the modern disparate impact doctrine and provided advocates a substantive analytical

12 See id. at 657 (“As a general matter, a plaintiff must demonstrate that it is the application of a specific or particular employment practice that has created the disparate impact under attack.”).
13 See id. at 659 (“The touchstone of this inquiry is a reasoned review of the employer’s justification for his use of the challenged practice.”).
14 Id.
16 Seiner & Gutman, supra note 4, at 2194.
17 See 42 U.S.C. § 2000e-2(k)(1)(A) (2006) (“An unlawful employment practice based on disparate impact is established . . . [if] a complaining party demonstrates that a respondent uses a particular employment practice that causes a disparate impact on the basis of race, color, religion, sex, or national origin and the respondent fails to demonstrate that the challenged practice is job related for the position in question and consistent with business necessity . . . .”).
18 See Civil Rights Act of 1991, Pub. L. No. 102-166, § 3(2), 105 Stat. 1071, 1071 (stating that one of the Act’s purposes was “to codify the concepts of ‘business necessity’ and ‘job related’ enunciated . . . [in] Supreme Court decisions prior to Wards Cove”).
19 Id. § 2(2).
framework with which to make their cases. The Supreme Court did not address the disparate impact doctrine again until 2009.

C. Ricci v. DeStefano: An Update on the Disparate Impact Doctrine

The Supreme Court’s decision in Ricci v. DeStefano and the implications it holds for disparate impact litigation, affirmative action, and the Equal Protection Clause of the Fourteenth Amendment have generated a lively academic discussion. In Ricci, a group of white and Hispanic firefighters sued the City of New Haven because it refused to certify the results of a promotional exam; the firefighters argued that the failure to certify violated Title VII and the Equal Protection Clause. The City, though, maintained that it had refused to certify the exam results because such a certification would have had a disparate impact on minority firefighters and thus exposed the City to Title VII liability.

The Supreme Court held that an employer is liable under Title VII if it overturns promotion test results because of race unless the employer can demonstrate that it has a strong evidentiary basis that doing so would cause it to lose a disparate impact lawsuit. The Court concluded that the City had not demonstrated a sufficient basis in evidence to show that it would be liable to the unsuccessful test takers in a disparate impact suit had it certified the test results.

The implications of Ricci for the disparate impact doctrine have yet to be seen; predictions have ranged from hopelessly bleak to cautiously optimistic. Professor Richard Primus has summarized three possible interpretations of the Ricci holding. First, it may mean that the “actions necessary to remedy a disparate impact violation are per se in

---

40 See Seiner & Gutman, supra note 4, at 2194 (“With the passage of the Civil Rights Act of 1991, . . . disparate impact finally gained the clear analytic framework it had lacked since its inception in Griggs.”).
41 See sources cited supra note 4.
42 129 S. Ct. 2658, 2664 (2009).
43 Id.
44 Id. at 2676-77. In articulating its reasoning for the holding, the Court explained: But once . . . [a testing] process has been established . . . [employers] may not then invalidate the test results, thus upsetting an employee’s legitimate expectation not to be judged on the basis of race. Doing so, absent a strong basis in evidence of an impermissible disparate impact, amounts to the sort of racial preference that Congress has disclaimed and is antithetical to the notion of a workplace where individuals are guaranteed equal opportunity regardless of race.
45 Id. at 2677 (citation omitted).
conceptual conflict with the demands of disparate treatment doctrine (and, implicitly, the demands of equal protection),” sounding the death knell for disparate impact causes of action.\(^\text{46}\) Second, \textit{Ricci} may be distinguished institutionally, because “public employers, unlike courts, are not authorized to engage in the race-conscious decision-making that disparate impact remedies entail.”\(^\text{47}\) Third, \textit{Ricci} may signify that in situations which produce “visible victims”—the white and Hispanic firefighters—the remedy is not proper.\(^\text{48}\) Because “the standard judicial remedies all avoid creating [such] visible victims,” this last reading allows Title VII’s disparate impact standard to survive future constitutional attack.\(^\text{49}\)

Notwithstanding the disagreement surrounding disparate impact remedies after \textit{Ricci}, the steps involved in disparate impact litigation remain—as of yet—unchanged.\(^\text{50}\) First, the plaintiff must prove that the defendant engaged in a practice that had an adverse impact on a protected class.\(^\text{51}\) Upon the plaintiff’s sufficient pleading of a prima facie case, the burden shifts to the defendant to justify the disparate

\(^{46}\) Richard Primus, \textit{The Future of Disparate Impact}, 108 MICH. L. REV. 1341, 1344 (2010). Professor Primus cites Justice Scalia’s concurrence in \textit{Ricci} and an article by Ronald Dworkin as two primary examples of this interpretation. \textit{Id.}

\(^{47}\) \textit{Id.} at 1344-45.

\(^{48}\) \textit{Id.} at 1345.

\(^{49}\) \textit{Id.; see also} Charles A. Sullivan, \textit{Ricci v. DeStefano: End of the Line or Just Another Turn on the Disparate Impact Road?}, 104 NW. L. REV. 411, 411 (2010) (“Reports of the death of Title VII’s disparate impact theory of discrimination in the wake of Ricci v. DeStefano may be exaggerated.”). \textit{But see} Melissa Hart, \textit{From Wards Cove to Ricci: Struggling Against the “Built-In Headwinds” of a Skeptical Court}, 46 WAKE FOREST L. REV. 261, 262 (2011) (arguing that Ricci “may well have done as much to eviscerate disparate impact’s potential as Wards Cove did twenty years earlier”); Seiner & Gutman, supra note 4, at 2205-09 (contending that certain interpretations of Ricci, in particular the one that allows for a new affirmative defense, would impose significant barriers to plaintiffs attempting to establish disparate impact liability).

\(^{50}\) However, the standards governing the business necessity defense may also be affected by \textit{Ricci}. Most recently, in \textit{Easterling v. Connecticut}, a federal district court suggested that the standard for establishing business necessity in disparate impact cases after \textit{Ricci} is a “Significantly Correlated Standard”—that is, “a hiring practice is job related if the practice is significantly correlated with elements of work behavior that are relevant to the job.” 783 F. Supp. 2d 323, 325 (D. Conn. 2011).

There may also be an additional affirmative defense based on a defendant’s awareness of a disparate impact at the time of the alleged discriminatory action. \textit{See infra} note 96; \textit{see also} United States v. Massachusetts, 781 F. Supp. 2d 1, 3 n.3 (D. Mass. 2011) (mentioning such a potential defense but declining to evaluate its merits).

\(^{51}\) \textit{See, e.g.}, Guardians Ass’n of the N.Y.C. Police Dep’t, Inc. v. Civil Serv. Comm’n, 630 F.2d 79, 86 (2d Cir. 1980) (explaining that “the common mode of Title VII analysis” requires that the plaintiff first “establish a prima facie case on the basis of disparate impact”).
impact by establishing a business necessity defense. The defendant must demonstrate both that the selection procedure adequately tests a certain skill and that this skill is sufficiently necessary to perform the job. Even if the defendant proves a legitimate business necessity, the plaintiff may still prevail if she is able to propose an alternative practice that would still provide for the employer’s legitimate business needs but without the disparate impact.

In theory, these steps apply just as easily to selection procedures based on physical tests (which tend to implicate gender concerns) as they do to procedures based on written tests (which tend to implicate race). Yet the fact patterns in race and gender cases seem intuitively different, and in light of the different standards brought to bear in the Court’s equal protection jurisprudence, it is worth pausing for a

52 See, e.g., id. ("[T]he defendant is required to rebut the plaintiff’s case by proving that the disparity results from legitimate, job-related selection procedures.").
53 See supra text accompanying note 19; see also Griggs v. Duke Power Co., 401 U.S. 424, 432 (1971) (holding that the employer practice or policy in question must have a “manifest relationship” to the employee’s job duties). Indeed, the Equal Employment Opportunity Commission (EEOC) has issued guidelines to direct courts’ analysis of the business necessity defense. The defense, under the guidelines, requires that employers’ discriminatory tests be validated by either (1) content-validity, (2) criterion-validity, or (3) construct-validity studies. 29 C.F.R. § 1607.5(A) (2011). Content validity can be established by producing “data showing that the content of the selection procedure is representative of important aspects of performance on the job for which the candidates are to be evaluated.” Id. § 1607.5(B). Criterion validity can be established by producing “empirical data demonstrating that the selection procedure is predictive of or significantly correlated with important elements of job performance.” Id. Construct validity, meanwhile, can be established by producing “data showing that the [employment] procedure measures the degree to which candidates have identifiable characteristics which have been determined to be important in successful performance in the job for which the candidates are to be evaluated.” Id.
54 See, e.g., Albemarle Paper Co. v. Moody, 422 U.S. 405, 425 (1975) (“If an employer does . . . meet the burden of proving that its tests are ‘job related,’ it remains open to the complaining party to show that other tests or selection devices, without a similarly undesirable racial effect, would also serve the employer’s legitimate interest . . . .”).
55 Courts apply intermediate scrutiny to analyze gender-based classifications under the Equal Protection Clause. See, e.g., Craig v. Boren, 429 U.S. 190, 197-200, 204 (1976) (using intermediate scrutiny to strike down a statute that prohibited the sale of beer to males under the age of twenty-one, while restricting sale to females under the age of eighteen). In contrast, courts analyze race-based classifications under a more rigorous standard of judicial review. See, e.g., Adarand Constructors, Inc. v. Peña, 515 U.S. 200, 227 (1995) (holding that all racial classifications—invidious or benign—imposed by the federal government are subject to strict scrutiny). While traditional disparate impact analysis does not implicate constitutional concerns because it is a statutory cause of action, Ricci has demonstrated that invalidating test results due to perceived disparate impact liability can, in some instances, trigger equal protection concerns. See supra notes
moment to consider gender-specific disparate impact case law. Part II explains the gender-based disparate impact doctrine as it has developed in the Supreme Court and in lower courts and seeks to provide a backdrop against which to consider the smaller subset of cases—those involving physical-selection procedures—that are examined in Part III.

II. THE GENDER-BASED DISPARATE IMPACT DOCTRINE

The disparate impact doctrine applies to any employee who has been discriminated against “because of” or “on the basis of” that person’s “race, color, religion, sex, or national origin.” It is quite logical, then, for advocates championing the expansion of the gender-based disparate impact doctrine to turn to Supreme Court jurisprudence defining disparate impact generally. However, the Supreme Court has overwhelmingly shaped this doctrine in response to race-based cases. This Part focuses on *Dothard v. Rawlinson* and its effects on gender-based disparate impact litigation in the lower courts.

A. Dothard v. Rawlinson: Limits to Its Strengths

Any narrative of the gender-based disparate impact doctrine must begin with *Dothard v. Rawlinson*, the only gender-based disparate impact case considered by the Supreme Court to date. In *Dothard*, Dianne Rawlinson applied to work as a correctional counselor with the Alabama Board of Corrections but was rejected because she did not meet Alabama’s statutorily imposed 120-pound weight requirement. Plain-
tiffs submitted evidence that Alabama’s height and weight thresholds combined to exclude over forty percent of the nationwide female population but less than one percent of the male population. The Court held that this evidence was sufficient to establish that the requirements had a disparate impact on female applicants.

Asserting the business necessity defense, the defendants argued that the height and weight requirements were related to strength. Strength, they argued, was “essential to effective job performance” as a prison guard. The Court rejected this argument because the defendants did not provide appropriate evidentiary support for what “amount of strength” a guard needed to perform effectively or what height or weight would ensure that a guard possessed the requisite strength. The Court criticized the defendants for failing “to offer evidence of any kind in specific justification of the statutory standards” and consequently refused to sustain a business necessity defense.

But, by dismissing this defense on the technical grounds that no statistical analysis was undertaken, the Court did not engage in further analysis of the defendants’ claims. Instead, the Court stated simply that “[i]f the job-related quality that the appellants identify is bona fide, their purpose could be achieved by adopting and validating a test for applicants that measures strength directly.” In so doing, the Court bypassed ruling on the merits of the defendants’ underlying argument of whether strength is an appropriate measure by which to gauge job ability.

More specifically, the Court did not acknowledge the fact that the disparate impact doctrine does not allow all strength measures to pass the business necessity test in all situations. In Dothard, the Court presumed that Dianne Rawlinson, who physically dealt with prisoners on a daily basis, needed to meet a minimum threshold of strength in order to adequately perform her job duties. While this assumption was

---

Marshall’s dissent eloquently demonstrates that the majority’s reasoning “regrettably perpetuates one of the most insidious of the old myths about women—that women, wittingly or not, are seductive sexual objects.” Id. at 345 (Marshall, J., dissenting).

59 Id. at 329-30 (majority opinion). While the defendants argued that the Court should consider statistics based on actual applicants to the positions in question, the Court accepted the plaintiffs’ nationwide population. Id. at 330-31.

60 Id. at 331.
61 Id.
62 Id. at 331-32.
63 Id.
64 Id. at 332.
probably true in that particular case, it is not necessarily true in all gender-related disparate impact cases. Consider, at the extremes, a strength requirement for a high school English teacher, or a test requiring a minimum number of pushups for a job that involves long-distance walking. Such tests should fail—the former because the English teacher does not need physical strength to perform his job (lack of relationship between the skill and the job), and the latter because pushups are not the appropriate method with which to measure aerobics (lack of relationship between the test and the required skill).

On a fundamental level, the disparate impact doctrine requires evidence of a clear correlation between the skill and the test (demonstrated by the English teacher example), and the test and the job (demonstrated by the pushup example). If the Dothard Court had stated that employers must statistically demonstrate that strength and other broadly defined traits were related to the job at hand, and that the test adequately reflected that trait, the Court could have then shown how the employers in the case had (or had not) established these dual requirements. Instead, the Court’s presumptive silence on this matter has led to confusion in lower courts’ analyses of physical-selection procedure cases and has in this way stymied Dothard’s potential.

Dothard’s impact, as stated in later cases, has by and large been limited to three specific propositions: (1) gender is included in Title VII’s enumerated list of protected classes, and the same disparate impact standards used for race should also apply to gender-based disparate

65 Two race-related corollaries might be the literary tests in Albemarle and the general intelligence test in Griggs. In both cases, the Court held that the tests did not adequately correspond to the job duties of low-level workers at paper mills and coal plants, respectively. See Albemarle Paper Co. v. Moody, 422 U.S. 405, 431-32 (1975); Griggs v. Duke Power Co., 401 U.S. 424, 431 (1971).

66 Cf. Brunet v. City of Columbus, 642 F. Supp. 1214, 1220-22 (S.D. Ohio 1986) (concluding that a ladder-climbing test was a poor indicator of endurance and agility, both of which were important when evaluating necessary skills in potential firefighters), rev’d on other grounds, 1 F.3d 390 (6th Cir. 1993).

67 See supra notes 19, 53 and accompanying text; see also United States v. Massachusetts, 781 F. Supp. 2d 1, 17 (D. Mass. 2011) (“The first part of the analysis, business necessity, ‘inquires whether the job criteria arise out of a manifest business need.’ . . . The second part of the analysis, relatedness, ‘inquires whether there is a [statistically proven] correlation between the criteria used and successful job performance.” (quoting Graf-fam v. Scott Paper Co., 870 F. Supp. 389, 400 (D. Me. 1994))).

68 See infra Section III.C.
impact claims; (2) data from the actual applicant pool in question is not always necessary to establish a prima facie case of disparate impact; and (3) height and weight restrictions cannot by themselves establish a business necessity sufficient to rebut a disparate impact claim.

While these holdings have undoubtedly “strengthened the position of disparate impact plaintiffs,” disparate impact has not had nearly as much effect on dismantling gender stratification as advocates had once so forcefully predicted.

B. Gender-Based Disparate Impact in Lower Federal and State Courts After Dothard

Gender-based disparate impact cases have had limited success since Dothard. To be sure, a number of height and weight restrictions—similar to those in Dothard—have been struck down by lower courts. Yet novel claims outside of the traditional testing, patronage, and promotion cases remain “few and far between.” For example, “no-

---

69 See, e.g., Yuhas v. Libbey-Owens-Ford Co., 562 F.2d 496, 498 (7th Cir. 1977) (explaining that in Dothard “the Supreme Court expressly extended Griggs to a case of sexual discrimination”).

70 See, e.g., Berkman v. City of New York (Berkman II), 705 F.2d 584, 594 (2d Cir. 1983) (“Those who have been deterred by a discriminatory practice from applying for employment are as much victims of discrimination as are actual applicants whom the practice has caused to be rejected.”); LeBoeuf v. Ramsey, 503 F. Supp. 747, 756 (D. Mass. 1980) (using Dothard to support the assertion that “[g]eneralized national statistics, such as those offered by plaintiff here, will suffice to make out a prima facie case of violation of Title VII”), rev’d sub nom. Costa v. Markey, 677 F.2d 158 (1st Cir. 1982).


72 Lye, supra note 19, at 326.

73 See supra note 1 and accompanying text.

74 See, e.g., EEOC v. St. Louis-S.F. Ry. Co., 743 F.2d 739, 741-43 (10th Cir. 1984) (finding that a height requirement for the position of switchman-brakeman discriminated against women); Horace v. City of Pontiac, 624 F.2d 765, 769 (6th Cir. 1980) (holding that height requirements illegally discriminated against female police officer applicants); Blake v. City of Los Angeles, 595 F.2d 1367, 1381-83 (9th Cir. 1979) (rejecting a minimum height requirement for applicants to the Los Angeles Police Department (LAPD)).

75 Michael Selmi, Was the Disparate Impact Theory a Mistake?, 53 UCLA L. REV. 701, 742-43 (2006). Professor Selmi’s research, which addresses all disparate impact claims,
"spouse" restrictions prohibiting spouses from working together have on occasion been statistically proven to have an adverse impact on women. However, such findings of disparate impact have rarely resulted in employer liability. As of 2006, only one federal court of appeals had found a no-spouse claim viable under disparate impact theory. Even pregnancy leave claims, once championed by feminist advocates, have found only limited success in the courts. As Professor Joanna Grossman recently opined, “The reality is that plaintiffs almost

and not just gender-related suits, also points to a significant overall decline in the success rate of disparate impact litigation since 1985. Id. at 738.

See, e.g., EEOC v. Rath Packing Co., 787 F.2d 318, 332-33 (8th Cir. 1986) (holding that a no-spouse rule had a disparate impact upon women and was not "justified by business necessity"); Yulas v. Libbey-Owens-Ford Co., 562 F.2d 496, 500 (7th Cir. 1977) (determining that the no-spouse rule had a "substantial discriminatory impact," as evidenced by the fact that "seventy-one of the last seventy-four people disqualified under it were women"). But see Thomas v. Metroflight, Inc., 814 F.2d 1506, 1509 (10th Cir. 1987) (finding that the plaintiff failed to meet her burden of proving that an airline company’s no-spouse restriction had a disparate impact on women); Harper v. Trans World Airlines, Inc., 525 F.2d 409, 412-14 (8th Cir. 1975) (holding that the employee failed to prove that a no-spouse policy adversely affected women).

Notably, married women have brought the “vast majority” of suits challenging no-spouse policies, thus “mak[ing] it clear that antinepotism and no-spouse rules have a disparate impact on women.” Timothy D. Chandler et al., Spouses Need Not Apply: The Legality of Antinepotism and No-Spouse Rules, 39 SAN DIEGO L. REV. 31, 43, 69 (2002).

Sharon Rabin-Margalioth, Love at Work, 13 DUKE J. GENDER L. & POL'Y 237, 245 (2006). Professor Rabin-Margalioth finds the lack of successful no-spouse claims “surprising” because, “[a]lthough neutral on their face,” no-spouse restrictions “disproportionately affect the female partner in the relationship, whether upon the mutual decision of the couple concerned or the unilateral decision of the employer.” Id.


See, e.g., Stout v. Baxter Healthcare Corp., 282 F.3d 856, 861-62 (5th Cir. 2002) (upholding a three-day pregnancy leave policy); Dormeyer v. Comerica Bank-Ill., 223 F.3d 579, 581, 583-84 (7th Cir. 2000) (rejecting a disparate impact challenge when a pregnant employee was discharged); Spivey v. Beverly Enters., Inc., 196 F.3d 1309, 1314 (11th Cir. 1999) (rejecting a disparate impact challenge to an employer’s practice of only assigning light duty to officers injured on the job); EEOC v. Elgin Teachers Ass’n, 27 F.3d 292, 295-96 (7th Cir. 1994) (upholding a leave policy that treated pregnant and nonpregnant teachers equally); United States v. Bd. of Educ. of the Consol. High Sch. Dist. 230, Palos Hills, Ill., 983 F.2d 790, 797-99 (7th Cir. 1993) (upholding the maternal leave policy over the plaintiff’s disparate impact claim); Levin v. Delta Air Lines, Inc., 730 F.2d 994, 1002 (5th Cir. 1984) (upholding, against a disparate impact claim, Delta’s policy of shifting pregnant women to ground duty).
never prevail on such claims in the pregnancy context." Yet, one area of gender-based disparate impact cases has found success in the courts: physical-selection procedure cases.

III. A SUBSET OF SUCCESS: PHYSICAL-SELECTION PROCEDURE CASES

Discriminatory physical-selection procedure cases under the disparate impact theory serve as an excellent case study of both the potential and the limitations of the gender-based disparate impact doctrine. These cases, which parallel race-based disparate impact cases involving written examinations, feature female plaintiffs who challenge the use of physical-selection procedures, such as weightlifting, push-ups, and running, which act as entry barriers to traditionally male-dominated jobs. Although the Supreme Court recently tackled written examination cases, one area of gender-based disparate impact cases has found success in the courts: physical-selection procedure cases.
tions in Ricci, it has yet to rule on the validity of physical-selection procedures. Lower courts have issued conflicting opinions, both in the standards applied and in the holdings reached. This Part will attempt to explain why some cases succeed while others do not.

To that end, this Part will identify the various points at which physical-selection procedure claims tend to break down. First, the majority of cases appear to presume that physical-selection procedures have a prima facie disparate impact on women and so proceed directly to an analysis of the business necessity defense. This approach suggests that courts do not view actual disparate impact evidence as an integral component of the gender-based disparate impact narrative. Second, some decisions point to the plaintiffs’ failure to adequately prepare for exams, an analysis that is inconsistent with the Griggs holding. Third, the dispositive issue in a majority of cases is the business necessity defense. This Part argues that courts have misconstrued the spirit of this defense as put forth in the Civil Rights Acts of 1964 and 1991. Finally, litigants struggle against claims that “strength is everything” and that


Eight out of nineteen cases have upheld physical-selection procedures as valid under Title VII. See Lanning v. Sc. Pa. Transp. Auth. (Lanning II), 308 F.3d 286, 288 (3d Cir. 2002) (validating a test that required applicants to run one-and-a-half miles in twelve minutes); Zamlen v. City of Cleveland, 906 F.2d 209, 213, 217-220 (6th Cir. 1990) (validating a test involving an overhead lift); Cleghorn v. Herrington, 813 F.2d 992, 995-96 (9th Cir. 1987) (validating a physical-fitness test for security personnel in a nuclear facility); Berkman II, 705 F.2d 584, 587, 590 n.5 (2d Cir. 1983) (validating a test that included ladder raising and hose drags); Evans v. City of Evanston, 695 F. Supp. 922, 924, 928 (N.D. Ill. 1988) (validating a ladder-climb, hose-connect, and agility-course test), vacated, 881 F.2d 382 (7th Cir. 1989); United States v. City of Wichita Falls, 704 F. Supp. 709, 711, 715 (N.D. Tex. 1988) (validating an obstacle-course and dummy-drag test); Eison v. City of Knoxville, 570 F. Supp. 11, 12-13 (E.D. Tenn. 1983) (validating a test consisting of sit-ups, push-ups, squats, and a two-mile run); Hardy v. Stumpf, 576 F.2d 1342, 1343-44 (Cal. 1978) (en banc) (validating a wall-scaling test).

“more is better,” stereotypical and anachronistic assumptions regarding the traits needed to perform traditionally male occupations.\textsuperscript{88}\footnote{\textit{See infra} Section III.D.}

This Part suggests that the courts’ proclivities can largely be explained by their tendency to bypass the formal doctrine and instead heed their own traditional assumptions about gender. Courts place undue emphasis on arguments that should not be part of the disparate impact discourse. Courts also rely on presumptions about an employer’s needs without critically analyzing the validity of these presumptions and the context in which they have developed. While it is likely that some of these issues are not unique to gender-based disparate impact claims, it is not my intention to compare these physical-selection procedure plaintiffs to consider and be aware of the following common obstacles.

\footnote{\textit{See} infra Section III.D.}

A. Absence of Disparate Impact Statistics

Under the traditional disparate impact analysis, the first issue typically addressed is the presence or absence of a statistically adverse impact on a protected class. The minimum threshold needed to establish a disparate impact has been hotly contested, with plaintiffs and defendants clashing over the “four-fifths rule,”\textsuperscript{89}\footnote{\textit{See} 29 C.F.R. § 1607.4(D) (2011) (stating that a selection rate “for any race, sex, or ethnic group” that is less than eighty percent “of the rate for the group with the highest rate will generally be regarded by the Federal enforcement agencies as evidence of adverse impact”); \textit{see also} Watson v. Fort Worth Bank & Trust, 487 U.S. 977, 995 n.3 (1988) (describing the EEOC’s eighty-percent standard as having been “criticized on technical grounds” and characterizing it as no “more than a rule of thumb for the courts”).} the minimum sample size,\textsuperscript{90}\footnote{\textit{See}, e.g., Fudge v. City of Providence Fire Dep’t, 766 F.2d 650, 657, 658 n.10 (1st Cir. 1985) (concluding that “[w]here the size of the sample is small . . . the ‘four-fifths rule’ is not an accurate test of discriminatory impact”).} and the relevant labor market.\textsuperscript{91}\footnote{\textit{See}, e.g., EEOC v. Chi. Miniature Lamp Works, 947 F.2d 292, 299-303 (7th Cir. 1991) (finding the use of Chicago as the “relevant labor market” inappropriate because the varying levels of interest for the jobs at issue, based on commuting time and language preference, could not adequately be taken into account); Pina v. City of E. Providence, 492 F. Supp. 1240, 1246 (D.R.I. 1980) (determining that the use of the “general population” as the relevant labor market was “proper” because firefighting is a skill that “the general population may possess”).} Yet practically all of the physical-selection procedure cases in the gender context have completely bypassed the issue of adverse impact. For example, in seven of the eight
cases in which physical-selection procedures were validated, courts first determined that a disparate impact existed but then found an appropriate business necessity. Indeed, most of the courts found the existence of a disparate impact as a matter of course and paid only scant attention to the statistical support advanced by the plaintiffs. As such, “disparate impact” as a statistical inquiry has been noticeably absent from physical-selection procedure cases.

Plaintiffs’ attempts to introduce the stark discrepancies in hiring rates often fail to persuade. In Berkman v. City of New York (Berkman IV), for example, a total of seven women placed in the top 15,316 applicants on a physical-abilities test. Most courts disclose the statistics in their opinions but nevertheless find that business necessity justifies the employer’s discriminatory practices. These courts quickly shuttle through the statistical analyses, so painstakingly put together by previous courts, en route to more “meaningful” analyses. It seems that courts are more interested in analyzing the qualities of the job in question than the statistics that substantiated the adverse impact claim in the first place.

92 See supra note 82 for a list and description of these cases. The only exception to this pattern is Eison v. City of Knoxville in which the plaintiff claimed that a physical-qualification test had a disparate impact on female applicants to the police force. 570 F. Supp. 11, 12 (E.D. Tenn. 1983). If the plaintiff’s specific test class from 1982 had been analyzed separately, it would have fallen below the “four-fifths rule.” Id. at 13. However, the court found that a proper analysis should include all applicants who took the same test throughout the years of its administration; thus, the court found “no adverse impact on women.” Id.

93 In the Lanning I litigation, for example, the district court devoted only five “findings of law” in determining that a prima facie case of disparate impact existed. Lanning v. Se. Pa. Transp. Auth., Nos. 97-0593, 97-1161, 1998 WL 341605, at *55 (E.D. Pa. June 25, 1998), vacated, 181 F.3d 478 (3d Cir. 1999). However, it devoted sixty-five “findings of law” to the question of whether defendants had a proper business necessity defense. Id. at *55-70. Likewise, in United States v. City of Wichita Falls, the court devoted only several sentences to the analysis of statistical adverse impact. 704 F. Supp. 709, 712 (N.D. Tex. 1988). As it hastily continued, “However, simply because a test has an adverse impact under the EEOC Uniform Guidelines on a minority group does not mean that the test necessarily is discriminatory.” Id.

94 812 F.2d 52, 60 (2d Cir. 1987).

95 See supra Part I.

96 In fact, in light of the Court’s decision in Ricci, there may be even more at stake in achieving a comprehensive doctrinal framework for the statistical analyses surrounding disparate impact. The Ricci Court held that an employer is liable for overturning test results unless it can demonstrate that it had a strong evidentiary basis that certifying such results would result in its losing a disparate impact action. 129 S. Ct. 2658, 2676-77 (2009). While the implications of this holding remain debated, Joseph Seiner and Benjamin Gutman have argued that Ricci creates an affirmative defense if the defendants can prove they did not know the practice would have an unlawful disparate impact.
B. Failure to Train

One common pitfall encountered in physical-selection procedure cases is the sufficiency of the claimants’ training and preparation. A number of physical-selection procedure cases, including the Third Circuit’s *Lanning v. Southeastern Pennsylvania Transportation Authority* decisions (*Lanning I* and *II*), added a new component to the disparate impact narrative: an adequate preparation requirement. In *Lanning I* and *II*, female plaintiffs challenged the use of a physical-fitness-screening test that required applicants to run one-and-a-half miles within twelve minutes. The Southeastern Pennsylvania Transportation Authority (SEPTA) conceded that the test produced a disparate impact on women as 55.6% of male applicants passed the test, compared to 6.7% of female applicants. Finding that SEPTA had not shown that its fitness-test cutoff score measured “the minimum qualifications necessary for successful performance of the job in question,” the Third Circuit in *Lanning I* reversed the district court’s ruling for the defendants.

However, Judge Weis’s dissent developed an interesting argument: he pointed out that “nearly all women who trained for [the running test] were able to pass.” Videotapes revealed that plaintiff Cathy Lanning had walked for a portion of the test, demonstrating what Judge Weis deemed a “‘cavalier’ attitude towards the running test.” The lack of training proved crucial to Judge Weis’s analysis of the case:

Here, where applicants have it within their power to prepare for the running test, they may properly be expected to do so. In view of the important public safety concerns at issue, it is not unreasonable to expect

---

Seiner & Gutman, *supra* note 4, at 2204. Thus, under this framework, the statistical analysis of whether a disparate impact exists would be relevant not only for a plaintiff’s prima facie case but also for a defendant’s affirmative defense. It may be even more important for potential litigators and defendants to understand what exactly a “disparate impact” looks like in the first instance.

98 *Lanning II*, 308 F.3d at 288; *Lanning I*, 181 F.3d at 481-83.
99 *Lanning I*, 181 F.3d at 483.
100 *Id.* at 494. The Third Circuit remanded the case and instructed the district court to reassess SEPTA’s business necessity defense under this “minimum qualifications” standard. *Id.*
101 *Id.* at 495 (Weis, J., dissenting) (emphasis added).
102 *Id.*
all applicants—female or male—to take the necessary steps in order to qualify for the positions.\footnote{Id. at 501. In fact, SEPTA’s brief described how Ms. Lanning “ran a portion of the course with her hands in her pockets.” Brief for Appellee Southeastern Pennsylvania Transportation Authority at 9, \textit{Lanning I}, 181 F.3d 478 (Nos. 98-1644, 98-1755), 1999 WL 33617177.}

Despite Ms. Lanning’s affidavit describing her diligent preparation for the test,\footnote{See Brief for Appellants Lanning et al. at 58, \textit{Lanning I}, 181 F.3d 478 (Nos. 98-1644, 98-1755), 1998 WL 34085350 (“Officer Lanning paid to attend a police academy. When she trained to pass the SEPTA run, she developed a 1.5 mile course, practiced running it on a routine basis, and timed herself as she ran.”).} and despite her subsequent successful career with the University of Pennsylvania Police Department’s tactical bicycle unit,\footnote{\textit{Id.} at 31.} Judge Weis was not convinced.

When the case reached the court of appeals for the second time, the majority incorporated portions of Judge Weis’s \textit{Lanning I} dissent into its decision. In “one final note,” the court stated the following:

While it is undisputed that SEPTA’s 1.5 mile run test has a disparate impact on women, it is also undisputed that, in addition to those women who could pass the test without training, nearly all the women who trained were able to pass after only a moderate amount of training. It is not, we think, unreasonable to expect that women—and men—who wish to become SEPTA transit officers, and are committed to dealing with issues of public safety on a day-to-day basis, would take this necessary step.\footnote{\textit{Lanning II}, 308 F.3d 286, 292 (3d Cir. 2002).}

Given this language, it is unclear whether the court intended to make training part of the prima facie analysis—a possible limitation on the use of actual statistics to prove a prima facie adverse impact—or some sort of affirmative defense against “cavalier” plaintiffs, or a mere chastisement, in dicta, of the plaintiffs’ lack of effort.

\textit{Lanning II} is not the only case that has made reference to a plaintiff’s failure to train. Another gender-based disparate impact case, \textit{Berkman v. City of New York (Berkman IV)}, has also alluded to the training aspect. In that case, plaintiffs were given the opportunity to participate in a special training program for women in preparation for a physical-abilities test.\footnote{\textit{Berkman IV}, 812 F.2d 52, 55 (2d Cir. 1987).} In rejecting the plaintiffs’ claim that the test had a disparate impact, the court noted in passing that many of the female applicants had refused this training, and that the training had proven effective for those female applicants who did participate in the
program. Failure-to-train language has also made its way into race-based disparate impact decisions. In its *Ricci* decision, for example, the Supreme Court mentioned in dicta that a possible argument in favor of certifying the New Haven test results might stem from the fact that the test was easy to pass if the candidate had studied. As one firefighter candidate explained, “[e]very one” of the questions on the written examination “came from the [study] material. . . . [I]f you read the materials and you studied the material, you would have done well on the test.” In her dissent, Justice Ginsburg noted that access to such study materials “fell at least in part along racial lines.” Other lower court decisions addressing the validity of written examinations have also discussed the potential impact of failure to study.

This failure-to-train branch of thought has been coined “contributory” disparate impact theory by some scholars. Professor Peter Siegelman, for example, has advocated for a new affirmative defense that would exculpate defendants “if they can show that plaintiffs seeking to establish disparate impact liability failed to make reasonable efforts to meet the job requirement being challenged.”

---

108 Id. at 55, 62. The court, however, did not ultimately rely on this rationale. Instead, it focused on the business necessity affirmative defense. *Id.* at 59-60.
110 *Id.*
111 *Id.* at 2693 (Ginsburg, J., dissenting).
112 See, e.g., *Perry v. Orange Cnty.*, 341 F. Supp. 2d 1197, 1212 n.7 (M.D. Fla. 2004) (arguing that one of the plaintiffs “should have been excluded from the disparate impact analysis” because, among other things, she did not purchase the test preparation materials for the fire department promotion exam); *Gilbert v. City of Little Rock*, 544 F. Supp. 1231, 1253 (E.D. Ark. 1982) (“[T]he [written] test does not improperly impact upon blacks, but instead the low test score resulted from lack of study.”), *aff’d in part, rev’d in part* 722 F.2d 1390 (8th Cir. 1983); *Contreras v. City of Los Angeles*, No. 77-1706, 1977 WL 15509, at *7 (C.D. Cal. Nov. 10, 1977) (holding that because the plaintiffs “did not attempt to seriously perform well” on their accounting exams there was not an “adverse effect on Spanish surnamed persons”), *aff’d*, 656 F.2d 1267 (9th Cir. 1981). *Contreras*, at least, can be distinguished from *Ricci* because it involved an unusually small sample size and a group of professional accountants. See 1977 WL 15509, at *5 (“The Senior Accountant plaintiffs are trained specialists in the accounting profession who have pursued their chosen work for a substantial length of time. . . . Their qualifications to pass a multiple choice test should be substantially greater than minority applicants seeking to be firemen.”).
114 *Id.* at 520; see also Laya Sleiman, *Note, A Duty To Make Reasonable Efforts and a Defense of the Disparate Impact Doctrine in Employment Discrimination Law*, 72 FORDHAM L. REV. 2677, 2682 (2004) (“The theory of disparate impact can thus be strengthened and
parate impact theory is “far from well-established,” Professor Charles Sullivan has predicted that “the extension of the theory to all races and both genders may lead more courts to explore this approach.”

The problem with *Lanning II*, and especially a contributory negligence disparate impact theory, is that it is simply unsupported by the disparate impact paradigm. The current framework of disparate impact centers on three specific questions. First, is there an adverse impact on a protected class? Second, is there a business necessity for this adverse impact? Third, is there an alternative procedure that will satisfy the employer’s needs without having an adverse impact on these individuals? It is far from clear which of these questions would be answered by a plaintiff’s failure to train. It certainly is not relevant to the prima facie case, as the Court has already held that disparate impact can be demonstrated without actual applicant pool statistics, and thus a focus on actual plaintiffs is unwarranted. And it seems even less relevant to a question of business necessity or alternative methods.

More broadly, failure-to-train theories attempt to bring to light the fact that applicant preparation, and not the test itself, was the “actual” cause of the disparate impact. Yet it has been made abundantly clear that disparate impact discrimination focuses on the *consequences* of employment practices, and the presence of an alternative predating factor cannot excuse a discriminatory test. It does not matter whether the disparate impact was caused by the bad faith of employers, lack of applicant preparation, or a poorly designed entrance examination; what is critical in the prima facie case is the impact itself.

As discussed earlier, courts do not spend much time discussing the *actual* prima facie case in physical-selection procedure cases. See supra Section III.A. In light of these tendencies, it is important for plaintiffs to contextualize their arguments, urging con-
After all, the whole point of a disparate impact challenge is that a facially non-discriminatory employment or promotion device—in this case an examination—has a discriminatory effect. It would be odd indeed if a defendant whose facially non-discriminatory examination which has a disparate impact could escape the obligation to validate the examination merely by pointing to some other facially non-discriminatory factor that correlates with the disparate impact.119

Consider the impact of a failure-to-train analysis on the landmark Griggs case—should those plaintiffs have been condemned for not studying hard enough? Would the plaintiffs who had not adequately prepared for the GED exam be excluded from statistics establishing a disparate impact? In Griggs, the Supreme Court noted that the African American plaintiffs were more likely to fail aptitude tests because, as compared to whites, they had received “inferior education in segregated schools.”120 Plaintiffs’ lack of education was an alternative predating factor to their low passage rate on the test in question, but that fact did not preclude the Court from finding a disparate impact.121 If an alternative predating factor could undermine a prima facie case, then even the landmark Griggs case might have been decided differently.122

Introduction of preparation and training evidence should not be taken lightly; it is highly prejudicial and panders to the stereotype that women are weak, both in body and in will, and the perception that hard-working men lose their jobs due to the “preferential treatment”123 of women who fail to meet the bare minimum of expectations. In fact,

119 940 F.2d 1211, 1228 (9th Cir. 1991).
120 401 U.S. at 430.
121 See id. at 436 (“What Congress has commanded is that any tests used must measure the person for the job and not the person in the abstract.”); see also Stagi v. Nat’l R.R. Passenger Corp., 407 F. Supp. 2d 671, 676 (E.D. Pa. 2005) (“[T]he existence of some ultimate cause for which the employer is not legally responsible cannot defeat a plaintiff’s prima facie case.”); Ass’n of Mexican-Am. Educators v. California, 937 F. Supp. 1397, 1410 (N.D. Cal. 1996) (stating that “[d]efendants cannot escape liability by showing that the disparate impact is attributable to particular background factors” such as lack of English language skills and lack of education), aff’d, 183 F.3d 1055 (9th Cir. 1999).
122 Interestingly, the only Supreme Court decision on gender-based disparate impact, Dothard v. Rawlinson, would not have been changed by performing a failure-to-train analysis, because Dothard held that certain statutory height and weight requirements—that is, traits that are indisputably beyond a person’s control—violated Title VII. 433 U.S. 321, 332 (1977). As discussed earlier, Dothard did not extend its reasoning to strength requirements, perhaps further underscoring the weak posture of the Dothard decision. See supra text accompanying notes 61-68.
123 Grossman, supra note 80, at 617.
there may be an evidentiary basis upon which to exclude such arguments as irrelevant to the disparate impact framework,\textsuperscript{124} as they have no place in current prima facie analysis and are even less appropriate in discussing job relatedness and alternative methods.\textsuperscript{125} The courts’ consideration of such evidence is another obstacle faced by gender-based disparate impact litigants.

C. The Catchall Business Necessity Defense

The business necessity affirmative defense is the doctrinal graveyard for gender-based disparate impact claims.\textsuperscript{126} Due to the lack of clarity in the Civil Rights Act of 1991,\textsuperscript{127} the silence of the Supreme Court, and the consequent inconsistency of appellate court standards,\textsuperscript{128} judges have exercised considerable discretion in deciding how

\textsuperscript{124} The Federal Rules of Evidence exclude all evidence that is not relevant. FED. R. EVID. 402; see also FED. R. EVID. 401 (defining “relevant evidence” as “having any tendency to make the existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence”). Additionally, Rule 403 excludes even relevant evidence if “its probative value is substantially outweighed by the danger of unfair prejudice, confusion of the issues, or misleading the jury, or by considerations of undue delay, waste of time, or needless presentation of cumulative evidence.” FED. R. EVID. 403.

\textsuperscript{125} Professor Siegelman has argued that courts should create a new affirmative defense for defendants who can prove that plaintiffs failed to train for their test. See supra notes 113-14 and accompanying text. While that is certainly possible, none of the decisions discussed above has established this affirmative defense in so many words and thus it remains a tentative recommendation at best. However, establishing a failure-to-train defense is overly complicated and unnecessary. First, most plaintiffs already have a non–litigation-related incentive for training: passing the test. Second, while it is relatively easy to train for a timed run, other physical-selection procedure tests require more strenuous preparation. See, e.g., Brunet v. City of Columbus, 642 F. Supp. 1214, 1220 (S.D. Ohio 1986) (describing a test that involved a five-story ladder climb), rev’d on other grounds, 1 F.3d 390 (6th Cir. 1993). Mandating that applicants train adequately for these tests (or, alternatively, allowing defendants to escape liability by proving that applicants did not train) extends an advantage to those with the resources to do so.

\textsuperscript{126} Indeed, the eight cases studied in this Comment that resulted in selection-procedure validation all found a legitimate business necessity in their particular set of facts. See supra note 82.

\textsuperscript{127} See Rosemary Alito, Disparate Impact Discrimination Under the 1991 Civil Rights Act, 45 RUTGERS L. REV. 1011, 1013 (1993) (“[T]he Act is so ambiguous, and leaves so many major questions unanswered, that it is impossible to state with any degree of certainty what its impact will be other than to delay and complicate litigation for the foreseeable future.”).

\textsuperscript{128} See Hollar, supra note 83, at 785 (“Because the Supreme Court precedent is so confusing and the terms ‘business necessity’ and ‘job related’ are seemingly redundant, lower courts have struggled to articulate a rule for analyzing physical test cases.”). Hollar summarizes four major standards that have emerged for evaluating physical-abilities
to administer this defense.\[129\] For advocates hoping to expand the gender-based disparate impact doctrine, this discretion has proven to be more of a curse than a blessing. Indeed, the business necessity doctrine has become a sort of catchall that legitimizes the court’s intuitive reluctance to invalidate discriminatory selection procedures.

Doctrinally, the Civil Rights Act of 1991 codified the business necessity defense, allowing employers to escape Title VII liability if they can prove “that the challenged practice is job related for the position in question and consistent with business necessity.”\[130\] Thus, the affirmative defense analysis essentially focuses on two different, but related, questions. First, the defendant must establish sufficient job validity: is the skill being tested actually necessary for satisfactory job performance?\[131\] Second, the defendant must prove sufficient content validity: is the test in place actually related to this skill?\[132\]

Consider again the height and weight requirements at issue in Dothard. The Dothard Court rejected the requirements because the defendants had failed to produce evidence “correlating the height and weight requirements with the requisite amount of strength thought essential to good job performance.”\[133\] That is, the Court found that this particular test (the height and weight requirement) was not substantially reflective of the skill needed (strength). Thus, the employer’s test in Dothard failed due to its lack of content validity. The defendant also bore the burden of proving that strength was a business necessity for the job in question—that is, its job validity. Strength is not per se important to every occupation, but Dothard failed to clarify this point.

\[129\] Notably, even the least deferential standards have resulted in sustaining business necessity defenses. See, e.g., Lanning II, 308 F.3d 286, 291-92 (3d Cir. 2002) (applying the minimum-qualifications test). And even the most lenient standards have resulted in the rejection of the defense. See, e.g., Harless v. Duck, 619 F.2d 611, 616 n.6 (6th Cir. 1980) (employing the manifest-relationship test).


\[131\] See supra note 53 and accompanying text; see also, e.g., Graffam v. Scott Paper Co., 870 F. Supp. 389, 400 (D. Me. 1994) (requiring an inquiry into “whether the job criteria arise out of a manifest business need”).

\[132\] See supra note 55 and accompanying text; see also, e.g., Griggs v. Duke Power Co., 401 U.S. 424, 436 (1971) (“What Congress has forbidden is giving these devices and mechanisms controlling force unless they are demonstrably a reasonable measure of job performance.”); Graffam, 870 F. Supp. at 400 (requiring an inquiry into “whether there is a correlation between the criteria used and successful job performance”).

Most courts have adopted a single standard to deal with both content validity and job validity. The consequence of this approach is that the question of job analysis is at best conflated with that of content analysis and at worst completely overlooked. In *Hardy v. Stumpf*, for example, the California Supreme Court examined the validity of a physical-performance test for Oakland’s police force that required an applicant to scale a six-foot wall. In finding that the test was justified by business necessity, the court pointed to the fact that the test required applicants to perform a necessary job skill. As the court quipped, “Surely, it is difficult to imagine a more accurate way of testing ability to scale a six-foot wall than to scale one.”

Likewise, in *Evans v. City of Evanston*, the plaintiff challenged a pre-training physical-screening test for becoming a firefighter that required applicants to climb an aerial ladder, drag hoses, and walk through ten tires while carrying a tarp, all within a fixed time period. The defendant’s test creators consulted at least two “experts in firefighting” to ensure the screening test simulated on-the-job tasks. The court rejected the plaintiff’s argument that the test “was constructed without an effort to quantify the physiological requirements of the job of [a] firefighter.” The court reasoned:

Under [the plaintiff’s] theory, one might say of the typist that he or she needs hand and finger speed, and coordination, eye control and the capacity for sustained concentration; and the best way to test these is to measure the physiological requirements of each of these elements in typing and then apply appropriate tests for each. All this rather than ask the candidate to sit and type. 

Despite a male passage rate of over ninety percent and a female passage rate of less than sixteen percent, the court—citing the similarity

---

135 576 P.2d 1342, 1343-44 (Cal. 1978) (en banc). The wall-scaling test was specifically devised because, by law, Oakland fences could be no more than six feet. *Id.* at 1345.
136 See *id.* (“The record clearly establishes a reasonable relationship between the physical performance test—particularly the six-foot wall climb—and job performance.”).
137 *Id.* at 1347.
138 695 F. Supp. 922, 923-24 (N.D. Ill. 1988), *vacated*, 881 F.2d 382 (7th Cir. 1989). The City of Evanston set its cut-off time for passing the test at one standard deviation above the mean, which typically allowed about eighty-four percent of the applicant pool to pass. *Id.* at 929. The district court ultimately held that the time limitation was unreasonable. *Id.*
139 *Id.* at 924.
140 *Id.* at 927-28.
141 *Id.* at 927.
between the components of the test and the requirements of the job—validated the test. 142

The Evans and Hardy courts seemed to find it intuitive that a test replicating a job-related skill should be validated, and they consequently dismissed alternative tests for falling short of the ideal. 143 Had these courts performed a two-step analysis, however, they might have understood the problem: the accuracy of the test is dependent on what it is trying to measure. 144 If the point of the test is to establish the applicant’s ability to perform one specific task (in Hardy, the applicant’s ability to scale a wall), then the test is certainly a valid indicator of that particular skill. However, the defendant would then need to prove that the wall-scaling skill is substantially related to the job at hand. 145 For example, the defendant might be required to introduce evidence of the number of times wall-scaling is actually performed by the average police officer.

Alternatively, if the point of the test is to determine the applicant’s general physical fitness, rather than the applicant’s wall-climbing ability, then the job-relatedness issue would be easily established. After all, some level of physical fitness is without question an important aspect of a police officer’s job. 146 However, the defendant would still need to prove that the examination was an appropriate test for measuring physical fitness. In this alternative, the fact that the test is closely related to an actual job responsibility is confounding and likely irrelevant. 147 To continue with the analogy used in Evans, rejecting a test that asked a

142 Id. at 925, 928.
143 Id. at 925; Hardy v. Stumpf, 576 P.2d 1342, 1347 (Cal. 1978).
144 The EEOC has established three types of validity: criterion, content, and construct. See 29 C.F.R. § 1607.5(A) (2011). All three types relate to this general point. For definitions of these validities, see supra note 53.
145 The standard necessary to determine a substantial relationship between the skill and the job will vary by jurisdiction. See supra note 128. In Harless v. Duck, for instance, the Sixth Circuit found under the lenient “manifest relationship” standard that a test must present concrete evidence of the “amount” of the certain skill needed. 619 F.2d 611, 616 & n.6 (6th Cir. 1980). Applying this rationale in Hardy, a defendant would need to demonstrate an appropriate “amount” of wall-scaling to justify that component of the job-application test.
146 See Harless, 619 F.2d at 616 (“Undoubtedly, police officers must meet certain physical standards to be capable of performing their jobs safely and effectively.”).
147 See, e.g., Easterling v. Connecticut, 783 F. Supp. 2d 323, 340-41 (D. Conn. 2011) (explaining that the mile-and-a-half run required for becoming a corrections officer (CO) “was only meant to measure aerobic capacity” and that while “[a]erobic capacity is a prerequisite to a number of activities a CO might have to perform . . . [r]unning itself [was] not a close approximation to the typical duties of a CO”).
Scaling the Wall and Running the Mile

typist to type may not seem as absurd if the overall skill being tested were the ability to manage an office.

Moreover, both Evans and Hardy merely required the test as a pre-screening device prior to entering a training academy. In other words, following this selection procedure, applicants were sent to a training academy to ensure that they learned the required skills. The relevant analogy is thus not a typist being tested on her ability to type, but rather a typist being denied entrance into a training school because of her inability, at that precise moment, to type fast enough. In fact, the EEOC regulations specifically instruct employers to "avoid making employment decisions on the basis of measures of knowledges, skills, or abilities which are normally learned in a brief orientation period, and which have an adverse impact."

It may be that, even under this doctrinal analysis, defendants will ultimately demonstrate that tests that accurately assess job-related responsibilities also sufficiently reflect a skill that is necessary for job performance. Using this two-step process, however, will help displace some of the adjudicator's tendencies to appeal to her own intuitions of how things ought to be. It will force the defendant to provide more than just simple assertions of gut feelings or repetitions of the status quo. Most importantly, it will reorient the conversation to focus on the Griggs disparate impact doctrine.

---

148 In fact, studies have shown that "[w]omen benefit greatly from training because most tests have a significant skill component which can readily be taught." Ruth Colker, Rank-Order Physical Abilities Selection Devices for Traditionally Male Occupations as Gender-Based Employment Discrimination, 19 U.C. DAVIS L. REV. 761, 803 (1986). In Berkman v. City of New York (Berkman I), one of the plaintiffs' expert witnesses testified that for selection procedures, "what must be identified are not those who are strongest or fastest but, instead, those who, with the benefit of training in pacing ... can perform the punishing tasks of firefighting as they are actually required to be performed." 536 F. Supp. 177, 212 (E.D.N.Y. 1982). These witnesses also testified that "firefighting takes its toll, not as a result of failures of maximum strength or speed, even at critical moments, but rather through the physical demands extending over long periods of time which necessitate paced performance at less-than-maximum levels." Id.

149 29 C.F.R. § 1607.5(F).

150 At least one court has written approvingly of the two-step inquiry, though the court itself did not follow this approach. See Easterling, 783 F. Supp. 2d at 341 ("A second viable interpretation of the foregoing precedent is that determining whether a hiring practice is 'job related and consistent with business necessity' is a two-step inquiry." (emphasis added)).
D. “Strength Is Not Always Everything,” and “More Is Not Always Better”

Plaintiffs in physical-selection procedure cases struggle against at least two other stereotypical assumptions regarding female applicants that undermine the Supreme Court case law as put forth in Griggs and its progeny. First, plaintiffs must combat the assumption that occupations requiring physical exertion, such as firefighting and policing, must necessarily involve only such traditional physical traits as speed, size, and strength.151 This “strength is everything” assumption has been repeatedly questioned by historical anecdotes and personal testimony.152 Some male SEPTA officers interviewed in the process of the Lanning I and II litigation, for example, stated that it was their experience that when arresting male perpetrators, sometimes a female officer could de-escalate the situation better than a male officer could.153 Indeed, the defendants were unable to produce any evidence at trial suggesting that they had ever taken disciplinary action against an officer because she could not perform the physical aspects of her job.154

Similarly, courts often make the assumption that in terms of physical skill, “more is better.” Under this theory, physical-abilities tests should be scored on a rank-order basis rather than simply pass or fail.155 In 1979, Brenda Berkman challenged the New York City Fire Department’s rank-order entrance examination for its adverse impact

151 Colker, supra note 148, at 776-79.
152 See, e.g., Blake v. City of Los Angeles, 595 F.2d 1367, 1379 n.7 (9th Cir. 1979) (“Shorter officers may have certain advantages in observing field situations (e.g., the ability to look under objects, the ability to squeeze through narrow passageways) that taller officers lack.”); David Holmstrom, Women Officers Arrest the Gender Gap, CHRISTIAN SCI. MONITOR, Jan. 12, 2000, at 11, 14 (asserting that women may improve police forces due to their ability to resolve neighborhood and family disputes as well as other problems).
153 E-mail from Lisa Rau, Att’y of Record for Lanning I & II to author (Oct. 12, 2011, 11:37 AM) (on file with author). It should be noted, however, that this was not a position advanced by Rau or any other counsel for the Lanning I and II plaintiffs. Rather, this belief “was what was reported by experienced male transit police officers. The Lanning Plaintiffs did not assert that they would be better than their male police officer counterparts but simply that they should be not be discriminated against in hiring when they were fully qualified to do the job.” Id.
154 Lanning I, 181 F.3d 478, 483 (3d Cir. 1999). According to the plaintiff’s brief on appeal, the defendants were unable to demonstrate in the court below that any officer provided inadequate backup assistance due to an inability to run a mile and a half in twelve minutes. Brief for Appellants Lanning et al., supra note 104, at 2. SEPTA discontinued the use of the contested running test in the fall of 2009. Telephone Interview with David Scott, Deputy Chief, SEPTA Transit Police Dep’t (Nov. 23, 2011).
155 See, e.g., Lanning I, 181 F.3d at 493-94 nn. 22-23.
on female applicants. Initially, the Department did not attempt to perform any content or job validity. Instead, it justified the rank-order with the “obvious” conclusion “that every increment in the abilities tested for in its physical exam necessarily represents a better performance as a firefighter.” The district court rejected this argument and noted that “rank-ordering should be used only if it can be shown that a higher score . . . is likely to result in better job performance.” Here, the physical tasks measured by the entrance exam were not representative of actual tasks performed by firefighters because they required candidates to perform the tasks anaerobically, while firefighters in the field use both anaerobic and aerobic skills simultaneously. The Second Circuit nevertheless validated the test. Basing its conclusion partially on “self-evident” intuitions, the Second Circuit found that “the Fire Department is entitled to select those who are endowed with the physical abilities to act effectively in the first moments of arrival at a fire scene, where immediate speed and strength literally concern matters of life and death.”

The two stereotypes apparent in Berkman and Lanning reflect a “nested doll” problem: unfounded assumptions are nested within each other, and a victory that peels away one assumption (that “more is better”) reveals yet another assumption (that applicants who are naturally “endowed” aerobically will be more effective “in the first

---


157 Berkman I, 536 F. Supp. at 211.

158 Id. at 210 (alteration in original) (quoting Guardians Ass’n of the N.Y.C. Police Dept’, Inc. v. Civil Serv. Comm’n, 630 F.2d 79, 100 (2d Cir. 1980)). The Second Circuit in Guardians held that “it is reasonable to insist that the test measure important aspects of the job.” 630 F.2d at 99.


160 Berkman IV, 812 F.2d 52, 59-60 (2d Cir. 1987).

161 Id. at 59. Rank-order tests have also been held valid in Cleveland. See Zamlen v. City of Cleveland, 906 F.2d 299, 220 (6th Cir. 1990).
moments of arrival at the scene of a fire"\textsuperscript{162}). In order to transcend these stereotypes, then, advocates must confront each assumption head on. For example, in challenging the “more is better” assumption, Professor Ruth Colker has asserted that the job analysis undertaken to substantiate rank-order tests often does not include women.\textsuperscript{163} As she argues, “In order to assess whether a test is valid on a rank-order basis it is crucial to include women in the job analysis. To date [1986], no job analysis has included female participants.”\textsuperscript{164} Advocates must demonstrate that by relying on these analyses, courts base a business necessity defense on a set of data—here the job analysis—that is itself a product of stereotypes. Relying on this type of analysis defeats the very purpose of disparate impact theory.

IV. BEST PRACTICES FOR GENDER-BASED DISPARATE IMPACT CASES

Because roughly half the physical-selection procedure cases studied in this Comment have resulted in the judicial validation of the test in question,\textsuperscript{165} the sample lends itself to a comparison between the cases that have invalidated tests and those that have not. I have identified four best practices that have proven beneficial in the physical-selection procedure realm and that may also be applicable to gender-based disparate impact cases in general.

First, an advocate should contextualize the problem at hand. Since a vast majority of physical-selection procedure plaintiffs seek positions in traditionally male occupations, contextualizing the industry as such will give the factfinder a narrative in which to frame the litigation. Second, an advocate should appeal to the factfinder’s capacity to dismiss unfounded stereotypes. This can be achieved by calling as witnesses female applicants who failed the test in question but ultimately proved to be skilled and successful in their jobs. Third, an advocate should

\textsuperscript{162} Berkman IV, 812 F.2d at 59.

\textsuperscript{163} Colker, supra note 148, at 798; see also Easterling v. Connecticut, 783 F. Supp. 2d 323, 328 (D. Conn. 2011) (“The plaintiff has presented expert testimony that the . . . percentile rankings used by the [Department of Corrections] were based on measurements from a sample of women who possessed a higher level of fitness than the overall female population.”); Brief for Appellants Lanning et al., supra note 104, at 48 (arguing that because SEPTA’s simulation included only five percent of women, this fact “increases the necessary scrutiny” given to the test and the design procedure).

\textsuperscript{164} Colker, supra note 148, at 798.

\textsuperscript{165} See supra note 82.
address valid concerns by providing concrete examples of effective alternative-selection procedures from other municipalities. Presenting such reasonable alternatives can help the factfinder overcome public safety concerns. Finally, and perhaps most importantly, an advocate should consistently refer back to the purpose and intent of the disparate impact cause of action, emphasizing its stated neutrality and its consequent potential for social change.

A. Contextualize the Test and the Industry

To understand the discourse surrounding physical-selection procedures, it is informative to review the tradition of exclusion faced by women seeking to enter the law enforcement and firefighting professions. In many major cities, women were not permitted to apply for certain positions within police forces or firefighting squads until the 1970s. Indeed, the City of New York did not open its firefighter applicant exam to women until 1977. When women were eventually allowed into these professions, they were relegated to second-class status. For example, Los Angeles policewomen were barred from regular police patrol duty, generally performed “tasks relating to women and children, desk duty, and administration,” and were ineligible for promotion above sergeant. Even the Supreme Court has acknowledged the obstacles faced by women in traditionally male professions; for example, in United States v. Virginia, the Court noted that “women seeking careers in policing [have] encountered resistance based on fears that their presence would . . . deprive male partners of adequate assistance . . . . Field studies [do] not confirm these fears.”

While this historical discrimination is not in and of itself a factor in disparate impact analysis, it provides an essential context in which to explain subsequent actions taken by city employers and testmakers. Simply put, physical-selection procedures are often used in jobs where it is still a “man’s world,” and this correlation is not a coincidence. When an industry consists of an overwhelming majority of men, then men are more likely to be in hiring and supervisory positions, and—

---

166 See, e.g., Blake v. City of Los Angeles, 595 F.2d 1367, 1371 (9th Cir. 1979) (explaining that the LAPD excluded entry-level policewomen from regular patrol work until 1973).
168 Blake, 595 F.2d at 1371.
consciously or subconsciously—they are more likely to base their standards of success around the manner in which men traditionally perform these jobs. They will likely be less understanding of how the job can be performed equally well in alternative manners. Because of the lack of internal pressure to adapt, courts should examine male-dominated occupations critically and vigilantly, looking out for decisions made on the basis of unfounded assumptions.

Using this lens, courts will find much about which to be suspicious. In Berkman I, a proposed test that considered an applicant’s flexibility was rejected because the fire department wanted a test “for firefighters, not for ballet dancers.” More recently, the employer in Lanning I and II hired a test developer who had previously been criticized by the Supreme Court in United States v. Virginia for his overtly stereotypical categorizations of women and the impact on all-male institutions if women were to be admitted. When asked rhetorically at a deposi-

---

170 See, e.g., Harless v. Duck, 619 F.2d 611, 616-17 (6th Cir. 1980) (invalidating a test developed through an “intuitive process”); Brunet v. City of Columbus, 642 F. Supp. 1214, 1249 (S.D. Ohio 1986) (rejecting “anecdotal evidence” as insufficient to validate a test), rev’d on other grounds, 1 F.3d 390 (6th Cir. 1993).

171 The Supreme Court has emphasized the importance of including members of the group who are disparately excluded in studies that are used to validate tests. See Albemarle Paper Co. v. Moody, 422 U.S. 405, 435-36 (1975) (citing 29 C.F.R. §§ 1607.5(b)(1), (b)(5) (1974)).


173 The expert in question, Dr. Paul Davis, testified extensively on the point that because of physiological differences between men and women, “[f]emale cadets would not be able to perform the tasks in the [Virginia Military Institute (VMI)] rat training program at a level comparable to that of male cadets,” that women would reduce the “intensity and aggressiveness of the current program,” and that as a result, “[t]he program’s benefit to the morale of male participants would be adversely affected.” United States v. Virginia, 766 F. Supp. 1407, 1438 (W.D. Va. 1991), vacated, 976 F.2d 890 (4th Cir. 1992), aff’d, 518 U.S. 515 (1996). When United States v. Virginia eventually reached the Supreme Court, the United States countered Davis’s testimony by arguing that “time and again since this Court’s turning point decision in Reed v. Reed, [this Court has] cautioned reviewing courts to take a ‘hard look’ at generalizations or ‘tendencies’ of the kind pressed by Virginia.” 518 U.S. at 541 (citation omitted). The Court agreed, finding that

[t]he notion that admission of women would downgrade VMI’s stature, destroy the adversative system and, with it, even the school, is a judgment hardly proved, a prediction hardly different from other ‘self-fulfilling proph[ecies],’ once routinely used to deny rights or opportunities. When women first sought admission to the bar and access to legal education, concerns of the same order were expressed.

Id. at 542-43 (alteration in original) (footnotes omitted) (citation omitted) (quoting Miss. Univ. for Women v. Hogan, 458 U.S. 718, 730 (1982)); see also Charles J. Russo & Susan J.
tion if there were any physical activities in which women’s abilities exceeded those of men, Dr. Davis responded affirmatively—women are better, he allowed, at “having children and giving milk.”

Proving test-developer bias, or inadequate content analysis prior to the application of the test, cannot invalidate a test if the analysis performed ex post ultimately substantiates the examination. Nevertheless, understanding the context may support the argument that courts should examine these decisions more critically in light of the fact that test developers are likely not to have undertaken such analysis themselves. As Judge O’Hern noted in his In re Vey dissent, when a woman is excluded from a traditionally “male” occupation, it should “give[] rise to concern whether a double standard is being invoked when such [facially neutral] traits are implemented to disqualify a woman from police duty.” It is important for the advocate to frame the litigation in a manner that brings context to the obstacles faced by female applicants and the test results that subtly reflect such obstacles.

B. Appeal to the Factfinder’s Capacity to Dismiss Unfounded Stereotypes

Secondly, it is important to identify concrete ways in which discriminatory stereotypes hurt both female applicants and society in general. Courts have positively received specific examples of female applicants who had failed a test, but ultimately proved to be skillful law enforcers. In fact, the Sixth Circuit in Zamlen v. City of Cleveland expressed the possibility of overturning trial verdicts that exclude the testimony of female employees hired prior to the implementation of the selection procedure. Because these employees would have spoken to the fact that the 1983 “physical performance examination [was] not representative of their actual duties,” the court found that their statements could have been “relevant to a central issue in this case—the validity of the selection device.”


639 A.2d 718, 719 (N.J. 1994) (O’Hern, J., dissenting). Judge O’Hern explained that “[a]mong [the] traits that the employing authority deems disqualifying for police work are [being] bold, suspicious, and easily frustrated.” Id.

906 F.2d 209, 215-16 (6th Cir. 1990).

Id. at 215. However, the court did not overturn the verdict due to the exclusion of this testimony because similar evidence was entered by male firefighters testifying on
In *Lanning I*, the plaintiff produced the testimony of a SEPTA officer who had not passed the timed mile-and-a-half run but had been hired due to a clerical error.\(^{176}\) In its statement of the facts, the Third Circuit noted:

This officer has subsequently been “decorated” by SEPTA and has been nominated repeatedly for awards such as Officer of the Year and Officer of the Quarter. SEPTA has commended her for her outstanding performance as a police officer and has chosen her to serve as one of SEPTA’s two defensive tactics instructors.\(^{179}\)

A similar commendation was put forth in *United States v. City of Erie*, where the court contemplated the ability of female officers hired under a previous standard that had allowed for more female hires.\(^{180}\) Courts have proven receptive to the fact that officers can perform at exceptional levels despite having failed the selection procedure put forth by the employer.

Courts have also responded positively to expert testimony critiquing the merits of tests. In *Brunet v. City of Columbus*, female firefighters challenged a physical test that, inter alia, required applicants to climb and descend a five-story high ladder.\(^{181}\) The plaintiffs’ expert concluded that

---

\(^{176}\) 181 F.3d 478, 483 (3d Cir. 1999). The Third Circuit noted, moreover, that “SEPTA has promoted incumbent officers who have failed . . . the physical fitness test,” and that “SEPTA has never disciplined, terminated, removed, reassigned, suspended or demoted any transit officer for failing to perform the physical requirements of the job”—all while women comprised only sixteen of its 234-member police force. *Id.*

\(^{179}\) *Id.*

\(^{180}\) See 411 F. Supp. 2d 524, 566-67 (W.D. Pa. 2005). As the court stated, [*T]he majority (five) of the eight female officers currently on the City’s police force were hired before the City began using the [physical-agility test] in 1994. . . . The City has presented no evidence that Detective Kemling or Detective Sergeant Mangan were ever unable to perform the physical tasks required by their jobs. On the contrary, Detective Sergeant Mangan has been promoted twice and received a commendation for her police work. Detective Kemling has been a field training officer and has been promoted to the rank of detective. *Id.* (citations omitted).

\(^{181}\) 642 F. Supp. 1214, 1220 (S.D. Ohio 1986), *rev’d on other grounds*, 1 F.3d 390 (6th Cir. 1993). In *Brunet*, the city administered the ladder climb to two sets of firefighter applicants—one in 1980 and a second in 1984—and in those years respectively four and two females, but 105 and 124 males, eventually met the necessary requirements to become firefighters. *Id.*
the test was a poor metric of endurance and agility.\textsuperscript{182} The court explained that “[t]he inevitable result of this narrowed focus upon strength is that relatively small differences in strength will tend to determine whether an individual is selected as a firefighter. There is no guarantee . . . that in selecting stronger individuals, individuals with greater endurance and agility are also being selected.”\textsuperscript{183} Finding that it was important for the test to cover the range of abilities necessary to perform a firefighter’s job, the court directed the city to redesign its test accordingly.\textsuperscript{184}

In fact, scholars have argued that advocates in race-based disparate impact cases should take note of the ability of claimants in gender-based cases to demonstrate that a discriminatory procedure hinders the accuracy of the test and therefore hurts the overall quality of employees.\textsuperscript{185} Professor Helen Norton, for example, asserts that the \textit{Ricci} decision stems from an assumption of a zero-sum state of equality—that “a decision maker’s concern for the disparities experienced by members of one racial group (‘empathy’) inevitably includes the intent to discriminate against others (‘prejudice’).”\textsuperscript{186} In order to rebut the notion of zero-sum equality, Professor Norton points to arguments used in \textit{Harless, Berkman}, and other cases involving gender-based physical-selection procedures.\textsuperscript{187}

\begin{flushright}
\textsuperscript{182} Id. at 1222.  \\
\textsuperscript{183} Id.  \\
\textsuperscript{184} Id. at 1222, 1250.  \\
\textsuperscript{185} Professor Ann McGinley argues that assessment centers may be “the preferable method[.] of measuring job skills while simultaneously permitting cities to promote a [racially] diverse population.” Ann C. McGinley, Ricci v. DeStefano: A Masculinities Theory Analysis, 33 HARV. J.L. & GENDER 581, 621 (2010). But assessment centers may have a harmful effect on female applicants if the employer does not sufficiently control for the so-called “double bind” predicament. \textit{Id}. The “double-bind” predicament was first introduced in Price Waterhouse v. Hopkins and refers to the fact that women who seek employment in a traditionally male occupation are criticized for their aggressiveness and thus placed “in an intolerable and impermissible catch 22: out of a job if they behave aggressively and out of a job if they do not.” 490 U.S. 228, 251 (1989) (plurality opinion). For example, Professor Madeline Heilman and her coauthors examined gender norms for women by having subjects review information about an employee they were told recently completed a training program. The researchers found that “it is only women, not men, for whom a unique propensity toward dislike is created by success in a nontraditional work situation.” Madeline E. Heilman et al., Penalties for Success: Reactions to Women Who Succeed at Male Gender-Typed Tasks, 89 J. APPLIED PSYCHOL. 416, 421, 426 (2004).  \\
\textsuperscript{186} Helen Norton, The Supreme Court’s Post-Racial Turn Towards a Zero-Sum Understanding of Equality, 52 WM. & MARY L. REV. 197, 202 (2010).  \\
\textsuperscript{187} See id. at 255-56 (emphasizing that successful arguments have pointed to the lack of a “meaningful relationship” between physical-ability tests and actual job requirements).
\end{flushright}
Here, “attention to disparate impact spurred better hiring for key public safety positions.” Advocates in race-based cases may do well to learn from their gender-based counterparts in framing arguments to demonstrate that disparate impact analysis protects “classes who may not be recognized as productive because of traditional yet unexamined assumptions.” In so doing, those advocates may ultimately produce a net social gain.

C. Address Valid Safety Concerns by Providing Concrete Reasonable Alternatives

Undoubtedly, there are valid public safety concerns at stake in public safety occupations. Likewise, there are valid concerns, both fiscal and social, when proposing expansions of gender-based disparate impact claims. It is important that advocates assuage these fears by demonstrating (1) that such changes have occurred elsewhere without adverse consequence to public safety and (2) that such an overhaul is not as drastic as it may at first appear.

In many of the cases that have successfully invalidated selection-procedure examinations, plaintiffs have been able to propose a reasonable alternative, fulfilling the third prong of the traditional disparate impact analysis. In *Blake v. City of Los Angeles*, for example, the district court found that the LAPD “could be so structured as to permit the employment of an equal percentage of male and female applicants,” but that it “could not be required to adopt less discriminatory alternatives if they required modification of departmental policies.” The Ninth Circuit disagreed, reasoning that “[s]o long as non-discriminatory alternatives serve the legitimate interests of the police in safe and efficient job performance, police departments cannot pursue policies that require the use of selection standards that are themselves prima facie violations of Title VII.”

---

188 *Id.* at 256.
189 *Id.* at 259.
190 *See, e.g.*, Harless v. Duck, 619 F.2d 611, 617 (6th Cir. 1980) (finding that alternative bases in the form of written intelligence exams existed for selecting police officers since these “tests did not have a statistically disparate impact between males and females”); EEOC v. Simpson Timber Co., No. 89-1455, 1992 WL 420897, at *4-5 (W.D. Wash. Jan. 7, 1992) (summarizing reasonable alternatives to the physical-abilities test used by the employer to select employees for logging and timber operations).
191 595 F.2d 1367, 1383 (9th Cir. 1979).
192 *Id.* One reasonable alternative that is currently popular in firefighter cases is the Candidate Physical-Ability Test (CPAT). For further information regarding the CPAT,
In contrast, a plaintiff’s failure to provide an alternative proposal can draw a rebuke from the court. Recently, in *United States v. Massachusetts*, a federal district court considered the legality of Massachusetts’s eleven-event Caritas Physical-Abilities Test (PAT), which the State used to select correctional officers. Roughly ninety-seven percent of men but only fifty-five percent of women passed the PAT in 2007. However, the plaintiff did not provide a reasonable alternative for screening and selecting correctional officer applicants. The court ultimately denied the Commonwealth’s motion for summary judgment, but it explicitly inveighed against the United States’ choice not to provide a reasonable alternative:

It is during trial of the third prong that the real heavy lifting would take place and the important Congressional purpose best could be achieved. The real question is not whether the Caritas PAT results in a disparate impact on women (it does), nor is it whether the test is job related and implemented to achieve public policy goals (we’ll see). The real question is whether we can do better. Can we achieve those same public policy goals and reduce or eliminate the disparate impact on women? Trial of the third prong would have explored those important issues in a fair and nuanced manner with the goal of truly achieving equal economic opportunity for both men and women while, at the same time, achieving the Commonwealth’s significant public policy goals.

Instead, the United States has reneged on its promise of evidence as to the third prong . . . .

Although Judge Young ultimately ruled against his original line of reasoning, it is clear from the opinion that his disdain for the government’s failure to proffer an alternative remained.

Additionally, plaintiffs have had success in deflating public safety concerns by pointing to irregularities in a city’s physical-fitness standard. In *Harless v. Duck*, for instance, the Sixth Circuit relied in part on the

---


3: Id. at 9.

4: Id. at 20-21.

5: See id. at 13. Indeed, Judge Young seemed to suggest that the third prong of the test was nearly the downfall of the litigation. He noted, “[A]s the United States simply gave up on the third prong, I [originally] thought myself entitled to infer that it had nothing better to offer than a reversion to the disparate treatment of the [Commonwealth’s previous] test . . . . On this record, I thought such an order [ordering the Commonwealth to develop a better alternative under federal court supervision] would contravene . . . congressional intent . . . .” Id.
fact that the city had eliminated the physical-ability test from a more recent selection procedure with “apparently . . . no detrimental effect on the police department.”

Similarly, in Blake, the Ninth Circuit rejected the business necessity defense: “The fact that the LAPD hired thousands of male police officers between 1968 and 1973 without using any pre-employment physical test suggests that the practice is not essential to safe and efficient job performance.”

Another argument that may be effective is to point out that in many instances applicants, once employed, are not required to pass physical examinations on a regular basis. In Lanning I, for example, the Third Circuit noted that SEPTA tried to discipline incumbent officers who failed fitness tests but that it discontinued this practice due to protests by the officers’ union. A city advocating that “more is better” with regard to the physical fitness of the group as a whole might struggle to substantiate that claim if it allows law enforcement officers to regress after they are hired.

D. The Doctrine Shall Set You Free

Finally, and perhaps most importantly, successful advocates consistently refer back to the disparate impact doctrine itself. The beauty of this doctrine, as laid out in Griggs and subsequent Supreme Court case law, lies in its fundamental simplicity: if the plaintiff proves that the defendant’s practice has a disparate impact on a protected class, that practice violates Title VII unless the defendant can prove that the practice is not based on stereotypical classifications but rather reflects an actual business necessity. If subsequent lower courts become mired in the details, it is helpful to remind them that such details may be inconsistent with the law as put forth by the Supreme Court. The Court has made no mention of adequate applicant preparation, nor

197 619 F.2d 611, 616 (6th Cir. 1980).
198 Blake v. City of Los Angeles, 595 F.2d 1367, 1382 (9th Cir. 1979).
199 181 F.3d 478, 483 (3d Cir. 1999). As the court explained, “SEPTA has promoted incumbent officers who have failed some or all of the components of the physical fitness test. SEPTA has also given special recognition, commendations, and satisfactory performance evaluations to incumbent officers who have failed the physical fitness test.” Id. In Lanning II, the Third Circuit likewise found that “[t]here has been no showing . . . that fitness level at the time of application is a reliable proxy for fitness level on the job over the ensuing years,” and “SEPTA has promoted officers who failed the running test” as well as “extended [offers] as much as two and one-half years after the aerobic running test is administered.” 308 F.3d 286, 296, 298 (3d Cir. 2002).
200 See supra notes 50-54 and accompanying text.
public safety doctrine, nor any idea that “more is better.” And it most certainly has not alluded to any doctrinal distinction between race-based and gender-based disparate impact claims.

Professor Michael Selmi has written, “The faith so many scholars and advocates have imbued in the disparate impact theory largely ignores much of what we have learned about the way in which the law works to preserve social norms rather than to upend them.” I disagree. History has shown time and again that the relationship between the law and the progression of social norms is symbiotic, complex, and dynamic. It is true that gender-based disparate impact case law as it currently stands is not as helpful as feminist advocates would like, but that does not mean that we should stop aiming for the ideal. Indeed, the success of one subset of gender-based disparate impact cases, physical-selection procedure cases, should give hope to advocates seeking to reclaim the ideals of Title VII.

CONCLUSION

Since the Supreme Court decided the landmark Dothard case, gender-based disparate impact causes of action have been limited in sphere and scope. But there remains room for growth. This Comment’s examination of the case law surrounding physical-selection procedures reflects the challenges of a theory highly dependent on the factfinder’s ability to free herself from pervasive assumptions. Female plaintiffs must combat unfounded assumptions and doctrinal nonstarters both in establishing prima facie cases and in rebutting affirmative defenses. They face failure-to-train claims and rallying cries of “more is better” and “strength is everything”—beliefs that are inconsistent with Griggs and other Supreme Court case law but which nevertheless persist.

However, the success of physical-selection procedure cases also suggests that the expansion of the disparate impact cause of action is possible, so long as advocates adhere to certain basic principles in framing their arguments. First, advocates must provide context for the problem at hand. Second, advocates must recognize any persistent stereotypes and make them palpable to the factfinder. Third, advocates must address valid concerns by providing reasonable alternatives or other similar analogies. Fourth, and most importantly, advocates must consistently tie arguments back to the Supreme Court doctrine and

201 Selmi, supra note 75, at 707-08.
the congressional intent that framed it. Physical-selection procedure advocates may confront a set of obstacles along the way, but advocates have and will continue to succeed. Maintaining this subset of successful litigation is pivotal not only to the gender-based disparate impact movement but also to the disparate impact movement overall.