
RECENT CASES

TITLE VII — SEX DISCRIMINATION — FOURTH CIRCUIT APPLIES “UNEQUAL BURDENS” ANALYSIS TO GENDER-NORMED FITNESS TEST. — *Bauer v. Lynch*, 812 F.3d 340 (4th Cir.), *reh’g en banc denied*, No. 14-2323 (4th Cir. Mar. 8, 2016).

In employment law, men and women are formally recognized as equally smart, equally capable, equally professional, but not equally strong. Courts often characterize physiological differences between men and women as “inherent”¹ or “immutable”² — biology dictated these differences and left them to the law to accommodate. Enacted in 1964, Title VII³ prohibits employment discrimination on the basis of race, color, religion, sex, or national origin.⁴ The Civil Rights Act of 1991⁵ amended Title VII by adding, among others,⁶ a provision forbidding the “discriminatory use of test scores” — including the use of “different cutoff scores” for different groups on “employment related tests.”⁷ Recently, in *Bauer v. Lynch*,⁸ the Fourth Circuit weighed in on the FBI’s gender-normed physical fitness test (PFT), which requires men and women to meet different raw cutoff scores. The court held that physical fitness tests that accommodate for sex-based⁹ physiological differences are not discriminatory as long as they impose equal burdens on men and women by requiring equal levels of fitness.¹⁰ Although the Fourth Circuit appropriately took into account the exceptional nature of physical fitness tests, its legal rule lacks clarity and fails to provide adequate guidance in the context of fitness-based employment testing.

¹ *E.g.*, *United States v. Virginia*, 518 U.S. 515, 533 (1996).

² *E.g.*, *Alspaugh v. Comm’n on Law Enf’t Standards*, 634 N.W.2d 161, 163 (Mich. Ct. App. 2001).

³ Civil Rights Act of 1964, Pub. L. No. 88-352, tit. VII, 78 Stat. 241, 253–66 (codified as amended at 42 U.S.C. §§ 2000e–2000e-17 (2012)).

⁴ 42 U.S.C. § 2000e-2(a) (2012).

⁵ Pub. L. No. 102-166, 105 Stat. 1071 (codified as amended in scattered sections of the U.S. Code).

⁶ Most notably, the Act codified disparate impact theory, which holds that facially neutral employment policies can be discriminatory if they have an adverse impact on protected groups. *The Civil Rights Act of 1991*, EQUAL EMP. OPPORTUNITY COMMISSION, <http://www.eeoc.gov/eeoc/history/35th/1990s/civilrights.html> [<http://perma.cc/7X7H-53Q5>].

⁷ The full text of the provision reads:

(l) Prohibition of discriminatory use of test scores

It shall be an unlawful employment practice for a respondent, in connection with the selection or referral of applicants or candidates for employment or promotion, to adjust the scores of, use different cutoff scores for, or otherwise alter the results of, employment related tests on the basis of race, color, religion, sex, or national origin.

⁴² U.S.C. § 2000e-2(l).

⁸ 812 F.3d 340 (4th Cir.), *reh’g en banc denied*, No. 14-2323 (4th Cir. Mar. 8, 2016).

⁹ Throughout this comment, “sex” and “gender” are used interchangeably.

¹⁰ *Bauer*, 812 F.3d at 351.

In July 2009, Jay Bauer failed out of the FBI Academy.¹¹ A former academic, Bauer applied to be an FBI Special Agent in 2008 and smoothly passed the screening process, which included written tests, interviews, and background checks.¹² The next step was a fitness test to gain admission to the training program. The PFT, consisting of sit-ups, a 300-meter sprint, a 1.5-mile run, and push-ups, serves to ensure a basic level of general fitness¹³ to (1) prevent injury during training and (2) support “effective training and application of the elements taught within the defensive tactics program.”¹⁴ All four PFT events were gender-normed according to the results of a 2003 pilot study — men were compared to men, women were compared to women, and different cutoff scores were determined for each group.¹⁵ For the push-up event, the passing threshold was set at approximately the fifteenth percentile,¹⁶ translating into thirty push-ups for men and fourteen for women.¹⁷ Bauer failed his first attempt, completing twenty-five pushups, but passed upon retake.¹⁸ In training, he “passed all academic tests, demonstrated proficiency in his firearms and defensive tactics training, and met all expectations for the practical applications and skills components of the Academy.”¹⁹ But a second PFT was required to graduate, and Bauer failed on five separate occasions, each time solely due to his inability to do thirty push-ups in a row.²⁰ On his fifth attempt, he fell one short with a completed total of twenty-nine.²¹ At the end of the twenty-two-week training period, Bauer resigned to preserve the opportunity of future employment with the FBI; two weeks later, he received an offer for an FBI Intelligence Analyst position.²² Bauer brought suit against the FBI, alleging that the PFT’s

¹¹ *Id.* at 342.

¹² *Id.* at 344.

¹³ *Id.* at 343. The PFT’s goal was to measure general fitness, not absolute strength or speed. Contrast the PFT with fitness tests used to screen for SWAT team members, which require applicants to simulate “specific job-related tasks, such as pulling oneself into an attic while wearing typical SWAT gear.” Reply Brief for Defendant-Appellant at 19, *Bauer*, 812 F.3d 340 (No. 14-2323) (internal citation omitted).

¹⁴ *Bauer*, 812 F.3d at 342.

¹⁵ See *id.* at 343–44. The 2003 pilot group comprised 258 male and 64 female trainees. *Id.* at 344. Results were verified in a follow-up study conducted in 2005. *Id.* at 344.

¹⁶ Accordingly, this threshold allowed around 85% of both men and women to pass. See *id.* at 344.

¹⁷ *Id.*

¹⁸ *Id.*

¹⁹ *Id.* at 344–45.

²⁰ *Id.* at 345.

²¹ *Id.*

²² *Id.* The Intelligence Analyst position, in contrast to the role of Special Agent, “is considered a ‘support’ position within the FBI.” *Bauer v. Holder*, 25 F. Supp. 3d 842, 849 (E.D. Va. 2014), vacated *sub nom.* *Bauer v. Lynch*, 812 F.3d 340 (4th Cir. 2016).

sex-based cutoff scores constituted Title VII discrimination. The parties subsequently filed cross-motions for summary judgment.²³

The district court granted summary judgment to Bauer, holding that the FBI's gender-normed PFT standards contravened the plain language of § 2000e-2(a)(1) — which prohibits sex-based employment discrimination generally — and § 2000e-2(l) — which prohibits using “different cutoff scores” for “employment related tests.”²⁴ Applying the “simple test” promulgated by *City of Los Angeles Department of Water & Power v. Manhart*,²⁵ the court concluded that the PFT was discriminatory because it treated Bauer “in a manner which but for his sex would have been different.”²⁶ Furthermore, the PFT's sex-based minimum standards plainly violated § 2000e-2(l)'s prohibition of “different cutoff scores.” Finally, the court rejected two defenses put forward by the FBI: one based on the bona fide occupational qualification (BFOQ)²⁷ and another based on the Supreme Court's holding in *Ricci v. DeStefano*.²⁸

The Fourth Circuit vacated and remanded.²⁹ Writing for the panel, Judge King³⁰ observed that the question of sex discrimination in gender-normed fitness tests is a “relatively novel issue” with limited precedent.³¹ The court held that the district court erred by applying *Manhart*'s “simple test” for sex discrimination;³² instead, the PFT should have been analyzed under the “unequal burdens” test espoused by the Ninth Circuit in *Gerdom v. Continental Airlines, Inc.*³³ Accordingly, the relevant question was whether PFT standards “impose[d] an equal burden of compliance on both men and women.”³⁴ Judge King

²³ *Bauer*, 812 F.3d at 345–46.

²⁴ *Bauer*, 25 F. Supp. 3d at 854–61. The Fourth Circuit noted that the district court incorrectly analyzed the claim under § 2000e-2(a), which applies to private sector employers, rather than § 2000e-16(a), which applies to federal employers. According to the court: “That is of no moment, however, as we have treated §§ 2000e-2(a) and 2000e-16(a) as comparable, with the liability standards governing the former being applicable to the latter.” *Bauer*, 812 F.3d at 345 n.3.

²⁵ 435 U.S. 702 (1978).

²⁶ *Bauer*, 25 F. Supp. 3d at 856 (citing *Manhart*, 435 U.S. at 711).

²⁷ *Id.* at 865. Disparate treatment may be permitted where sex “is a bona fide occupational qualification reasonably necessary to the normal operation of that particular business or enterprise.” 42 U.S.C. § 2000e-2(e) (2012).

²⁸ *Bauer*, 25 F. Supp. 3d at 860 & n.30 (citing *Ricci v. DeStefano*, 557 U.S. 557 (2009)). Disparate treatment may be used to avoid disparate impact where “there is a strong basis in evidence of disparate-impact liability.” *Ricci*, 557 U.S. at 583.

²⁹ *Bauer*, 812 F.3d at 352.

³⁰ Judge King was joined by Judge Harris and District Judge Hazel, sitting by designation from the District of Maryland.

³¹ *Bauer*, 812 F.3d at 347.

³² *Id.* at 350–51.

³³ 692 F.2d 602, 610 (9th Cir. 1982) (en banc) (holding that an airline's weight policy — mandating a strict weight requirement for female employees and no comparable weight requirement for males — was facially discriminatory because it imposed an unequal burden of compliance on women); *Bauer*, 812 F.3d at 348–49.

³⁴ *Bauer*, 812 F.3d at 351.

noted that the unequal burdens test, originally developed in the context of differing appearance standards for men and women,³⁵ has been adopted by the U.S. District Court for the District of Columbia and the Equal Employment Opportunity Commission (EEOC), specifically with regard to the FBI's PFT.³⁶ Additionally, the court pointed to the Supreme Court's language in *United States v. Virginia*,³⁷ which contemplated, in dicta, that women's admission to a military college "would require accommodations . . . [in] physical training programs for female cadets."³⁸ Since the district court had used an erroneous standard to determine whether sex discrimination had occurred, the Fourth Circuit did not reach the issues of the BFOQ and *Ricci* defenses.³⁹

Central to the court's holding was the idea that "[m]en and women simply are not physiologically the same for the purposes of physical fitness programs."⁴⁰ Due to innate physiological differences, "equally fit" men and women may not achieve the same raw scores on physical fitness tests.⁴¹ Under the Fourth Circuit's interpretation, the PFT's different raw score cutoffs for men and women were not necessarily discriminatory: "Whether physical fitness standards discriminate based on sex, therefore, depends on whether they require men and women to demonstrate different *levels* of fitness."⁴² The Fourth Circuit did not explicitly approve gender-norming — rather, it instructed the district court to determine on remand whether the gender-normed PFT "impose[d] an equal burden of compliance on both men and women, requiring the same level of physical fitness of each."⁴³

By holding that men and women must meet the same level of fitness, the Fourth Circuit appropriately recognized the exceptional nature of general fitness testing. This accommodation of sex-based physiological differences is necessary to ensure nondiscriminatory treatment of men and women. The court's articulation of the unequal burdens test, however, provides more questions than answers and does not clearly instruct how "burden" should relate to "fitness level."

³⁵ See *Gedrom*, 692 F.2d at 606.

³⁶ *Bauer*, 812 F.3d at 348–49 (citing *Powell v. Reno*, No. 96-2743, 1997 U.S. Dist. LEXIS 24169 (D.D.C. July 24, 1997); *Hale v. Holder*, No. 570-2007-00423x (E.E.O.C. Sept. 20, 2010)).

³⁷ 518 U.S. 515 (1996) (holding that the exclusion of women from the Virginia Military Institute violated the Equal Protection Clause).

³⁸ *Id.* at 540; see *Bauer*, 812 F.3d at 349.

³⁹ See *Bauer*, 812 F.3d at 352.

⁴⁰ *Id.* at 350. Though the court conceptualized physiological differences as innate, scholars have questioned this premise. See, e.g., Cass R. Sunstein, *Gender, Caste, and Law, in* WOMEN, CULTURE, AND DEVELOPMENT 332, 346 (Martha C. Nussbaum & Jonathan Glover eds., 1995) ("Differences in physical strength, for example, undoubtedly have a good deal to do with differences in expectations, nutrition, and training.").

⁴¹ *Bauer*, 812 F.3d at 351.

⁴² *Id.* (emphasis added).

⁴³ *Id.*

The Fourth Circuit correctly recognized that general fitness tests, reflecting sex-based physiological differences, should require men and women to achieve equal fitness levels rather than equal raw scores.⁴⁴ This holding takes into account both the purpose of the PFT to assess general fitness and the purpose of Title VII to prohibit employment discrimination. The goal of the PFT is not to measure how many push-ups a trainee can do. Rather, it is to support safety and success during training at the Academy.⁴⁵ Before implementing the PFT, the FBI found that an unacceptable number of trainees sustained injuries or failed training events.⁴⁶ In response, the PFT was designed to assess “the minimum fitness level that would give the FBI a reasonable amount of confidence that a [trainee] could complete the physically demanding [training program] without sustaining a training-related injury.”⁴⁷ For the FBI’s employment purposes, fitness is a proxy for an applicant’s level of preparedness for training,⁴⁸ and a well-designed PFT would demonstrate a strong correlation between measured fitness level and future performance in the Academy.

Under this understanding of fitness, equally fit men and women, even if they do unequal numbers of push-ups, are equally qualified for the Academy.⁴⁹ If “equally fit men and women demonstrate their fitness differently”⁵⁰ and men are physiologically equipped to do more push-ups, then identical raw score cutoffs would require female trainees to be *more* physically fit than their male counterparts.⁵¹ Such a result — holding women to higher physical fitness standards — would contravene § 2000e-16(a)’s prohibition of employment discrimination,

⁴⁴ Both the district court and Fourth Circuit accepted that the PFT is an employment-related test for the purposes of § 2000e-2(l). See *id.*; *Bauer v. Holder*, 25 F. Supp. 3d 842, 859 (E.D. Va. 2014).

⁴⁵ *Bauer*, 812 F.3d at 342.

⁴⁶ See Final Brief for Defendant-Appellant at 9, *Bauer*, 812 F.3d 340 (No. 14-2323).

⁴⁷ Declaration of Amy Grubb, Ph.D. at 4–5, *Bauer*, 25 F. Supp. 3d 842 (No. 1:13-cv-93).

⁴⁸ Importantly, this conception of fitness is not relative, as an applicant’s fitness level would not depend on the performance of others.

⁴⁹ Not surprisingly, the parties’ experts disagreed over whether the PFT truly screens for men and women who are equally prepared for training. Compare Expert Report of Kevin Murphy, Ph.D. at 6–7, *Bauer*, 25 F. Supp. 3d 842 (No. 1:13-cv-93) (stating that under the PFT, men and women with the same *percentile score* in push-ups had the same likelihood of injury during training), with Rebuttal in Support of Our Expert Report on the FBI’s Attempt to Validate Its Physical Fitness Test at 5, *Bauer*, 25 F. Supp. 3d 842 (No. 1:13-cv-93) (“One finding relating to this case was that low push up performance did not predict increased injury incidence.”). *Bauer* also noted that he had already completed training by the time he failed the second PFT, see Brief for Plaintiff-Appellee at 7, *Bauer*, 812 F.3d 340 (No. 14-2323), a point the Fourth Circuit did not address.

⁵⁰ *Bauer*, 812 F.3d at 351.

⁵¹ This raises the question of why push-ups were chosen to measure general fitness rather than, for example, exercises that measure flexibility (an area of fitness where women tend to outperform men). For a discussion on how facially gender-neutral employment standards can be biased against women, see Christine A. Littleton, *Reconstructing Sexual Equality*, 75 CALIF. L. REV. 1279 (1987).

as well as § 2000e-2(l)'s more specific prohibition of discriminatory use of test scores. Section 2000e-2(l) was drafted amid controversy surrounding the employment practice of "subgroup norming," especially with regard to cognitive aptitude tests,⁵² in which employees and applicants were scored based on their performance relative to their race and gender groups.⁵³ Congressional discourse criticized its perceived antimeritocratic nature: it was discriminatory to manipulate test scores to hire "less qualified, less productive applicants" on the basis of race or sex.⁵⁴ By contrast, general fitness tests are fundamentally different from cognitive aptitude tests; because physiological differences cause equally qualified men and women to score differently, sex must be taken into account to ensure nondiscriminatory treatment.

The question on remand should be, do PFT cutoff scores actually require men and women to meet equal fitness levels, by admitting men and women who are equally qualified to safely complete the training program? The Fourth Circuit's adoption of the unequal burdens test complicates this relatively straightforward inquiry. According to Judge King, the fundamental issue is whether PFT standards "impose an equal burden of compliance on both men and women, requiring the same level of physical fitness of each."⁵⁵ This language suggests that "fitness" and "burden" are synonymous, or at least compel the same results. What the court leaves open, however, is how "burden" should be independently defined, and how it should influence the analysis of "fitness."

One possible interpretation is "burden" as conceptualized by the appearance and grooming cases, birthplace of the unequal burdens test.⁵⁶ Employment practices that hold men and women to different appearance standards are considered discriminatory when they impose unequal economic, physical, or psychological burdens.⁵⁷ In *Frank v. United Airlines, Inc.*,⁵⁸ for example, the Ninth Circuit found that a

⁵² Linda S. Gottfredson, *The Science and Politics of Race-Norming*, 49 AM. PSYCHOLOGIST 955, 955 (1994).

⁵³ David W. Arnold & Alan J. Thiemann, *Test Scoring Under the Civil Rights Act of 1991*, INDUS.-ORGANIZATIONAL PSYCHOLOGIST, Jan. 1993, at 65, 65.

⁵⁴ 102 CONG. REC. S4923-24 (daily ed. Apr. 24, 1991) (remarks of Sen. Simpson) (quoting R. Gault Silberman, Vice Chairman, Equal Emp't Opportunity Comm'n, Remarks Before the Equal Employment Advisory Council (Feb. 28, 1991)). This supports the view that "Title VII seeks to ensure equal treatment of all similarly-qualified individuals regardless of gender." Note, "*Gradually Triumphant Over Ignorance*": Rhode Island's Treatment of Sexual Orientation Discrimination in the Workplace, 30 SUFFOLK U. L. REV. 439, 465 (1997).

⁵⁵ *Bauer*, 812 F.3d at 351.

⁵⁶ See *Jespersen v. Harrah's Operating Co.*, 444 F.3d 1104, 1109 (9th Cir. 2006) (en banc) (finding that grooming standards requiring female employees to wear makeup did not impose unequal burdens on women); *Carroll v. Talman Fed. Sav. & Loan Ass'n of Chi.*, 604 F.2d 1028, 1032 (7th Cir. 1979).

⁵⁷ Recent Case, 120 HARV. L. REV. 651, 654 (2006).

⁵⁸ 216 F.3d 845 (9th Cir. 2000).

sex-based weight policy was discriminatory, noting that female employees resorted to unacceptably burdensome measures, “including severely restricting their caloric intake, using diuretics, and purging,” to comply with stricter standards.⁵⁹ Similarly, when employment policies hold men and women to different grooming standards (such as makeup for women and short hair for men), the time and cost required for compliance are relevant considerations in evaluating burden.⁶⁰ Under an analogous unequal burdens test, the burden imposed by the PFT could be assessed according to a number of potential factors, including the time spent preparing for the test, the cost of joining a gym, and the physical effort needed to overcome constraints based on musculature. Notably, this test doesn’t directly evaluate applicants’ qualifications — men and women could theoretically invest the same amount of effort to prepare for the PFT, but still fail to be equally prepared or equally likely to avoid injury at the Academy.

Alternatively, “burden” could be understood to mean the likelihood of passing the PFT: if a member of one group is less likely to pass the test, then that group is more burdened by it. This view is implicit in the gender-norming scheme, which assumes that men and women are held to equal burdens as long as they, in the aggregate, can pass the test at approximately equal rates.⁶¹ The PFT used the same percentile cutoff for pilot groups of men and women, with the expectation that approximately equal percentages of men and women would continue to meet that threshold.⁶² Both Bauer and the FBI appear to have understood “burden” this way, but they disagreed on whether there was a statistical difference between the actual passage rates for men and women.⁶³ As in the former variation of the unequal burdens test, this

⁵⁹ *Id.* at 848.

⁶⁰ The *Jespersen* majority refused to take judicial notice of the time and cost associated with compliance, but implied that they would have been relevant factors had the plaintiff submitted evidence. See *Jespersen*, 444 F.3d at 1110. In dissent, Judge Kozinski argued that submitted evidence was not required to demonstrate the “incontrovertible facts” that makeup requirements impose additional time and cost burdens on women. *Id.* at 1117 (Kozinski, J., dissenting).

⁶¹ See *Bauer*, 812 F.3d at 344; see also *Hale v. Holder*, No. 570-2007-00423x, slip op. at 6 (E.E.O.C. Sept. 20, 2010) (“Moreover, following the establishment of the PFT at issue . . . the PFT pass rates for male [trainees] have equaled or exceeded that for their female counterparts. In light of these circumstances, it can hardly be said that the PFT more significantly burdens males . . .” (citation omitted)). Gender-norming further assumes that men and women with the same percentile score are by definition equally fit. Importantly, this conception of fitness is entirely relative.

⁶² See *Bauer*, 812 F.3d at 344 (“Within the push-up event, the FBI found that 84.3% of male Trainees and 84.1% of female Trainees in the Pilot Study achieved the minimum passing score or better. . . . [T]he FBI concluded that men and women of equal fitness levels were equally likely to pass the PFT.”).

⁶³ Compare Final Brief for Defendant-Appellant at 11, *Bauer*, 812 F.3d 340 (No. 14-2323) (“FBI experts . . . reviewed the standards to make sure that they were equally difficult for men and women. The experts expected 88.1% of minimally fit men to perform at least 30 push-ups, and a comparable 85.6% of minimally fit women to perform at least 14 push-ups.” (citation

interpretation of “burden” isn’t necessarily reflective of qualification for employment — it’s possible that in order to ensure equal passage rates for men and women, the PFT must require one gender to meet higher fitness levels. Both versions of the test ask the wrong question: whether the PFT is equally difficult for men and women, instead of whether the PFT selects for equally qualified men and women. This line of analysis could potentially require the sex-based differential treatment of equally qualified applicants, a discriminatory outcome under Title VII.⁶⁴

Ultimately, the only permissible reading of “burden” is the one that deprives it of any independent meaning in this context: burden as defined by fitness. This would mean that as long as men and women must meet the same threshold fitness level — even if it’s more difficult for one group to do so — they’re held to the same burden. This interpretation, of course, would render the unequal burdens test redundant because as long as equal fitness can be demonstrated, equal burdens will be logically deduced. It’s also fundamentally different from the other two variations of the test, in that its focus isn’t the difficulty of the test, but the qualification of the test taker. While this reading would comport with § 2000e-2(l)’s prohibition against discriminatory use of test scores and should be the one adopted on remand, it contravenes the understanding of the parties and finds little support in existing caselaw.⁶⁵

Physical fitness remains an important consideration in many areas of employment, including the military, law enforcement, and public safety. As the Supreme Court has said: “‘Inherent differences’ between men and women . . . [are not cause] for artificial constraints on an individual’s opportunity.”⁶⁶ In the context of employment testing, such “artificial constraints” can be eliminated only by recognizing that, while sex-based physiological differences may exist, the crucial question is whether an applicant is qualified for employment.

omitted)), *with* Brief for Plaintiff-Appellee at 6, *Bauer*, 812 F.3d 340 (No. 14-2323) (arguing that the PFT places an undue burden on males because out of the “22 that failed the new push-up minimum, 100% or all 22 were male”).

⁶⁴ Additionally, a requirement of equal passage rates raises the specter of the quota system, a much-feared outcome of subgroup norming. See Paul Oyer & Scott Schaefer, *Sorting, Quotas, and the Civil Rights Act of 1991: Who Hires When It’s Hard to Fire?*, 45 J.L. & ECON. 41, 41-43 (2002).

⁶⁵ In *Hale*, the administrative judge found that the fitness test was equally burdensome for men and women because, among other things, it “(1) screened out individuals of both genders who were not sufficiently fit to safely perform the duties of [a Special Agent]; and (2) did not screen out individuals of either gender who were sufficiently fit to safely perform as [a Special Agent].” *Hale*, No. 570-2007-00423x, slip op. at 6. However, these screening outcomes were considered alongside passage rates for men and women, as well as whether Special Agent jobs were eventually filled by a high proportion of one sex. *Id.*

⁶⁶ *United States v. Virginia*, 518 U.S. 515, 533 (1996).